International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:2, No:1, 2008

Novel adaptive channel equalization algorithms by statistical sampling

János Levendovszky, András Oláh

***Abstract*—** In this paper, novel statistical sampling based equalization techniques and CNN based detection are proposed to increase the spectral efficiency of multiuser communication systems over fading channels. Multiuser communication combined with selective fading can result in interferences which severely deteriorate the quality of service in wireless data transmission (e.g. CDMA in mobile communication). The paper introduces new equalization methods to combat interferences by minimizing the Bit Error Rate (BER) as a function of the equalizer coefficients. This provides higher performance than the traditional Minimum Mean Square Error equalization. Since the calculation of BER as a function of the equalizer coefficients is of exponential complexity, statistical sampling methods are proposed to approximate the gradient which yields fast equalization and superior performance to the traditional algorithms. Efficient estimation of the gradient is achieved by using stratified sampling and the Li-Silvester bounds. A simple mechanism is derived to identify the dominant samples in real-time, for the sake of efficient estimation. The equalizer weights are adapted recursively by minimizing the estimated BER. The near-optimal performance of the new algorithms is also demonstrated by extensive simulations. The paper has also developed a (Cellular Neural Network) CNN based approach to detection. In this case fast quadratic optimization has been carried out by t, whereas the task of equalizer is to ensure the required template structure (sparseness) for the CNN. The performance of the method has also been analyzed by simulations.

***Keywords*—** Cellular Neural Network, channel equalization, communication over fading channels, multiuser communication, spectral efficiency, statistical sampling.

I. INTRODUCTION

One of the major challenges of multiuser wireless communication is the effect of interferences which can cause severe degradation in performance and decrease the spectral efficiency. Therefore, techniques which can provide low BER communication over fading channels have received a great deal of interest (see [1],[2]). The two major sources of interferences are described as follows:

1) if non-orthogonal codes are assigned to the users it yields MultiUser Interference (MUI) [3]-[6];
2) multipath propagation can result in selective fading which entails the occurrence of InterSymbol Interference (ISI) [7]-[9].

Due to the interferences and to the resulting channel memory, the optimal decision comes down to the minimization of a quadratic objective function (see [10]) yielding exponential complexity, which prevents real-time detection. In order to reduce the detection complexity, a threshold detector followed by an adaptive channel equalizer is used. Traditionally the equalizer coefficients have been optimized subject to minimizing either the peak distortion (Zero Forcing (ZF) equalization), or the mean square error (Minimum Mean Square Error (MMSE) equalization) [11]-[13]. However, in the case of low-degree equalization (i.e. the equalizer has only a small number of free coefficients) these methods provide poor performance over channels affected with severe interferences [14],[15].

Thus, in this paper novel equalizer algorithms are introduced to reduce the BER in the presence of severe interferences. The new methods directly minimize BER in a computationally efficient manner and as a result, they yield superior performance in comparison to ZF or MMSE. To circumvent the exponential complexity of calculating the gradient of BER (which would prevent the standard gradient descent optimization), statistical sampling methods are used for BER estimation in each step of the algorithm. The proposed sampling methods rely on stratified sampling or on the Li-Silvester bounds which have already been known in the field of reliability analysis [17],[18].

II. THE MODEL

Multiuser communication can be modeled as follows:

- the number of users is denoted by K;
- N denotes the block-length;
- the overall number of transmitted bits is denoted by $M = K \cdot N$;
- the transmitted information vector is a binary vector $\mathbf{y} \in \{-1,1\}^M$, where $y_k^{(i)} \in \{-1,1\}$ represents the k^{th} information bit transmitted by user i (for the sake of the simplicity in the forthcoming discussion $y_k^{(i)}$ is simply

Manuscript received May 23, 2006. The research reported here was partly supported by the Katholieke Universiteit Leuven under the grant GOA/98/06 of Research Fund K.U.Leuven and by the Analogic Computing Laboratory MTA SZTAKI.

Janos Levendovszky is with the Department of Telecommunications Budapest University of Technology and Economics, Magyar tudósok körútja 2., H-1117, Budapest, Hungary (corresponding author to provide phone: + 36 1 463-2485; fax: +36 1 463-3263; e-mail: levendov@hit.bme.hu).

Andras Olah is with Peter Pazmany Catholic University Faculty of Information Technology, Práter utca 50/a, H-1083, Budapest, Hungary (olah@itk.ppke.hu).

denoted by y_k);

- each user is allocated a codeword of dimension *D*, selected from the codeset $C=\left\{\mathbf{c}_i; i=1,...,K : \mathbf{c}_i \in \{-1,1\}^D\right\}$ which is assumed to be the set of Gold codes [19];
- there is a signature signal $s_i(t)$ associated with each user, the signal pattern of which is based on the assigned codeword $\mathbf{c}_i$;
- the channel matrix is denoted by $\mathbf{R}$ and in the case of synchronous communication without multipath propagation it is described with the elements $R_{ij} := \frac{1}{D}\sum_{m=1}^{D} c_m^{(i)} c_m^{(j)};\ i,j=1,...,K$;
- in the case of asynchronous transmission corrupted by selective fading, the channel matrix becomes of type *M*x*M* and its elements can be calculated from the signature signal and from the channel impulse function, respectively (for further details see[20]);
- the communication is also corrupted by additive white Gaussian noise with average spectral energy N_0;
- the detected vector is denoted by $\hat{\mathbf{y}}$, where the k[th] detected bit is $\hat{y}_k$.

Based on the notations above, the communication model can be depicted by the following figure:

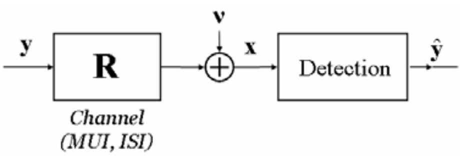

Fig. 1 the communication model

The received vector $\mathbf{x}$ can be expressed as $\mathbf{x} = \mathbf{R}\mathbf{y} + \mathbf{v}$, where $\mathbf{v}$ is multidimensional Gaussian random variable with zero mean and $N_0\mathbf{R}$ covariance matrix ($\mathbf{v} \sim \mathrm{N}\left(\mathbf{0}, N_0\mathbf{R}\right)$). Performing the Bayesian decision on the received vector, multiuser detection amounts to the global minimization of the following quadratic form [10]-[12],[16]:

$$\mathbf{y}_{\text{opt}} = \arg\min_{\mathbf{y}\in\{-1;1\}^M} \mathbf{y}^T\mathbf{R}\mathbf{y} - 2\mathbf{x}^T\mathbf{y} \tag{1}$$

Unfortunately, quadratic optimization over a discrete set in general cannot be carried out in polynomial complexity. Therefore, in the presence of interferences (which result in a non-diagonal channel matrix) the optimal detection is computationally out of reach.

In order to provide low-complexity detection, an equalizer is used to combat interferences followed by a threshold detector, as indicated by the Fig. 2., and the equalizer is a FIR filter depicted by Fig. 3. The equalizer carries out a linear transformation on the received sequence x_k yielding an output sequence $\tilde{y}_k = \sum_{j=1}^{J} w_j x_{k-j}$. This linear transformation can be rewritten into the form of $\tilde{\mathbf{y}} = \mathbf{W}\mathbf{x}$. Matrix $\mathbf{W}$ is fully determined by the coefficient vector $\mathbf{w}$ of the equalizer. Thus in the forthcoming discussion we refer to the equalizer either with matrix $\mathbf{W}$ or with vector $\mathbf{w}$.

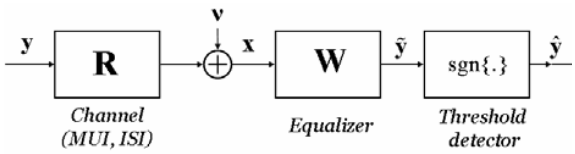

Fig. 2 channel equalization and threshold detector

The equalized signal (arriving at the input of the threshold detector) is given as

$$\tilde{\mathbf{y}} = \mathbf{W}\mathbf{R}\mathbf{y} + \mathbf{W}\mathbf{v} = \mathbf{G}\mathbf{y} + \boldsymbol{\eta}\,, \tag{2}$$

where $\mathbf{W}$ is the equalizer matrix and $\boldsymbol{\eta}$ is the transformed Gaussian noise with parameters $\boldsymbol{\eta} \sim \mathrm{N}\left(\mathbf{0}, N_0\mathbf{W}^T\mathbf{R}\mathbf{W}\right)$.

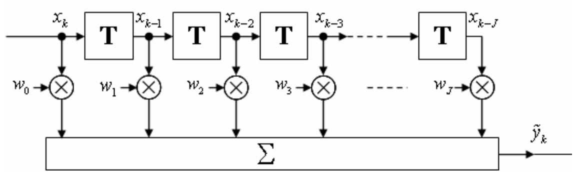

Fig. 3 the equalizer structure

III. Channel equalization by traditional algorithms

The main concern of equalization is to optimize the free parameters of the equalizer subject to a goal function $\mathbf{w}_{\text{opt}} := \min_{\mathbf{w}} J(\mathbf{w})$. In the case of the traditional MMSE equalization [13] the goal function is $J(\mathbf{w}) := E\left(y_k - \sum_{j=0}^{J} w_j x_{k-j}\right)^2$, while the ZF equalization minimizes the peak distortion [12] defined as $J(\mathbf{w}) := \sum_l \sum_{j=0}^{J} \left| w_j R_{lj} \right|$. Unfortunately, these two goal functions are not directly related to BER, thus the performance is rather poor in the presence of deep interference [14].

IV. Direct BER minimization

In order to achieve higher performance, our aim is to optimize the free parameters of the equalizer by minimizing BER.

Taking into account the Gaussian nature of the noise, the bit error probability denoted by $P_b^{(k)}$ can be expressed in close form as follows:

$$P_b^{(k)} = P\left(\hat{y}_k \neq y_k\right) =$$
$$= \frac{1}{2^M} \sum_{\mathbf{z}\in\{-1,1\}^M} \left[\Phi\left(\frac{-g_k + \sum_{j,j\neq k} g_j z_j}{\sigma_k} \right) + \Phi\left(\frac{-g_k - \sum_{j,j\neq k} g_j z_j}{\sigma_k} \right) \right],$$

where $\Phi(.)$ denotes the standard Gaussian distribution function and $g_i = \sum_{j=1}^{M} R_{ij} w_j;\ \ i = 1,...,M$.

The bit error probability can be rewritten into

$$P_b^{(k)} = \frac{1}{2^{M-1}} \sum_{\substack{\mathbf{z}\in\{-1,1\}^M \\ z_k=-1}} \Phi\left(\frac{\mathbf{z}^T\mathbf{g}}{\sqrt{N_0 \mathbf{w}^T \mathbf{g}}} \right) = \\ = \frac{1}{2^{M-1}} \sum_{\substack{\mathbf{z}\in\{-1,1\}^M \\ z_k=-1}} \Psi\left(\mathbf{w},\mathbf{z}\right), \tag{3}$$

where $\Psi\left(\mathbf{w},\mathbf{z}\right) := \Phi\left(\frac{\mathbf{z}^T\mathbf{g}}{\sqrt{N_0 \mathbf{w}^T \mathbf{g}}} \right)$.

The direct optimization of the equalizer amounts to searching for the optimal weights, as follows:

$$\mathbf{w}_{\text{opt}} := \min_{\mathbf{w}} P_b^{(k)}\left(\mathbf{w}\right) \tag{4}$$

Since $P_b^{(k)}$ is a differentiable function of w_i, $i = 1,...,J$, this minimization can be achieved by gradient search:

$$\mathbf{w}\left(n+1\right) = \mathbf{w}\left(n\right) - \Delta \operatorname*{grad}_{\mathbf{w}} \left\{ P_b^{(k)}\left(\mathbf{w}\left(n\right)\right) \right\} = \\ = \mathbf{w}\left(n\right) - \Delta \frac{1}{2^{M-1}} \sum_{\substack{\mathbf{z}\in\{-1,1\}^M \\ z_k=-1}} \mathbf{h}\left(\mathbf{z},\mathbf{w}\left(n\right)\right), \tag{5}$$

where

$$\mathbf{h}\left(\mathbf{z},\mathbf{w}\right) = \varphi\left(\frac{\mathbf{z}^T\mathbf{g}}{\sqrt{N_0 \mathbf{w}^T \mathbf{g}}} \right) \frac{\frac{1}{2} N_0}{\left(N_0 \mathbf{w}^T \mathbf{g}\right)^{3/2}} \cdot \\ \cdot \left(2\mathbf{R}\mathbf{z}\mathbf{w}^T\mathbf{g} - \mathbf{z}^T\mathbf{g}\mathbf{g}\right). \tag{6}$$

However, due to the exponential number of terms in the summation of the gradient, each step of algorithm (5) is of exponential complexity, which makes its implementation impossible. Thus, the challenge of combating interferences by equalization rests with the question of how to get rid of the large summation in the expression of the gradient. To achieve this end, statistical sampling techniques are introduced to estimate BER.

A. Minimizing BER by using statistical sampling techniques

In this section we develop novel equalization algorithms by which BER minimization becomes tractable in real-time. The underlying idea is to view the expression

$$\frac{1}{2^{M-1}} \sum_{\substack{\mathbf{z}\in\{-1;1\}^M \\ z_k=-1}} \mathbf{h}\left(\mathbf{z},\mathbf{w}\right) = E_{\mathbf{y}}\left\{\mathbf{h}\left(\mathbf{y},\mathbf{w}\right)\right\} \tag{7}$$

as an expected value taken over random variable y which is subject to uniform distribution $p\left(\mathbf{y}=\mathbf{z}\right) = 1/2^{M-1}$ (the realizations of random vector $\mathbf{y}$ are denoted by vectors $\mathbf{z} \in \{-1,1\}^M$). As was mentioned before, the calculation of this expected value involves a summation of exponential number of terms. In order to circumvent this complexity, statistical estimate is used to replace the expected value in (5). The estimate of the gradient of BER generally denoted by $\mathbf{f}\left(\mathbf{z}_{nL}, \mathbf{z}_{nL+1}..., \mathbf{z}_{(n+1)L-1}, \mathbf{w}\left(n\right)\right)$ and is obtained over a sample sequence $\mathbf{z}_{nL}, \mathbf{z}_{nL+1}..., \mathbf{z}_{(n+1)L-1}$. The choice of the sample sequence can either be stochastic or deterministic. By using this estimate algorithm instead of the true gradient in (5), one can obtain a General Sampling Based Equalization algorithm (GSBE) for minimizing BER. The GSBE algorithm can be given as follows:

TABLE I
GSBE ALGORITHM

Step 0		Set iteration number n=0 and initialize $\mathbf{w}$(0) with an arbitrary vector; perform the following iteration;
n^{th} step of the algorithm	I	select samples $\mathbf{z}_{nL}, \mathbf{z}_{nL+1}..., \mathbf{z}_{(n+1)L-1}$ of size L according to a sampling rule (specified later);
	II	adjust the equalizer coefficients: $\mathbf{w}\left(n+1\right) = \mathbf{w}\left(n\right) - \Delta \cdot \mathbf{f}\left(\mathbf{z}_{nL},...,\mathbf{z}_{(n+1)L-1}, \mathbf{w}\left(n\right)\right)$;
	III	check the stopping criterion $\left\|\mathbf{w}\left(n+1\right) - \mathbf{w}\left(n\right)\right\| < \varepsilon$ with a given ε and if the criterion has been fulfilled then the algorithm stops, otherwise increment iteration number n and go back to Step I.

The possible choices of selecting $\mathbf{f}\left(\mathbf{z}_{nL}, \mathbf{z}_{nL+1}..., \mathbf{z}_{(n+1)L-1}, \mathbf{w}\left(n\right)\right)$ to provide efficient estimates of the gradient are given in the next sections.

It must also be noted that *BER* can also be expressed as an expected value $BER = E_{\mathbf{y}}\left\{\Psi\left(\mathbf{y},\mathbf{w}\right)\right\}$, which is deduced from formula (3). It is important, because sometimes it is easier to approximate the expected value of $\Psi\left(\mathbf{y},\mathbf{w}\right)$ from the samples $\mathbf{z}_{nL}, \mathbf{z}_{nL+1}..., \mathbf{z}_{(n+1)L-1}$, than the expected value of the gradient $\mathbf{h}\left(\mathbf{y},\mathbf{w}\right)$. In the former case, the function $F\left(\mathbf{z}_{nL}, \mathbf{z}_{nL+1}..., \mathbf{z}_{(n+1)L-1}, \mathbf{w}\left(n\right)\right)$ will be used as an estimate of *BER* and we calculate the gradient of $F\left(\mathbf{z}_{nL}, \mathbf{z}_{nL+1}..., \mathbf{z}_{(n+1)L-1}, \mathbf{w}\left(n\right)\right)$ to estimate $E_{\mathbf{y}}\left(\mathbf{h}\left(\mathbf{y},\mathbf{w}\right)\right)$. As a result, Step II of the GSBE is modified as

$$\tilde{\mathbf{f}}\left(\mathbf{z}_{nL},...,\mathbf{z}_{(n+1)L-1}, \mathbf{w}\left(n\right)\right) = \\ = \operatorname*{grad}_{\mathbf{w}} F\left(\mathbf{z}_{nL},...,\mathbf{z}_{(n+1)L-1}, \mathbf{w}\left(n\right)\right). \tag{8}$$

It is important to note that $\tilde{\mathbf{f}}$ refers to the gradient of the estimation of *BER*, while **f** denotes the direct estimation of the gradient itself. In this paper both approaches are dealt with.

B. BER minimization by Monte Carlo methods

The most straightforward choice is to set $\mathbf{f}\left(\mathbf{z}_{nL},\mathbf{z}_{nL+1}\ldots,\mathbf{z}_{(n+1)L-1},\mathbf{w}(n)\right):=\frac{1}{L}\sum_{i=nL}^{(n+1)L-1}\mathbf{h}(\mathbf{z}_i,\mathbf{w})$ according to the Monte Carlo (MC) sampling. This yields the following equalization algorithm:

TABLE II
MONTE CARLO SAMPLING EQUALIZATION

Step 0		Set iteration number n=0 and initialize $\mathbf{w}(0)$ with an arbitrary vector; perform the following iteration;
n^{th} step of the algorithm	I	generate L sample vectors $\mathbf{z}_{nL},\mathbf{z}_{nL+1}\ldots,\mathbf{z}_{(n+1)L-1}$ subject to uniform distribution;
	II	set $\mathbf{f}\left(\mathbf{z}_{nL},\ldots,\mathbf{z}_{(n+1)L-1},\mathbf{w}(n)\right):=\frac{1}{L}\sum_{i=nL}^{(n+1)L-1}\mathbf{h}(\mathbf{z}_i,\mathbf{w})$;
	III	adjust the equalizer coefficients: $\mathbf{w}(n+1)=\mathbf{w}(n)-\Delta\cdot\mathbf{f}\left(\mathbf{z}_{nL},\ldots,\mathbf{z}_{(n+1)L-1},\mathbf{w}(n)\right)$;
	IV	check the stopping criterion $\left\|\mathbf{w}(n+1)-\mathbf{w}(n)\right\|<\varepsilon$ with a given ε and if the criterion has been fulfilled then the algorithm stops, otherwise increment iteration number n and go back to Step I.

Unfortunately, MC method does not provide an accurate estimate of the true gradient in the case of a small sample size. Therefore, the performance can further be improved by introducing stratified sampling.

C. BER minimization by stratified sampling

In this case a partition is given over the state space $\{-1,1\}^M$ denoted by $\Omega=\{Y_i,P_i;i=1,\ldots,V\}$, with the properties $\bigcup_{i=1}^{V}Y_i=\{-1,1\}^M$, $Y_i\bigcap Y_j=\varnothing\ \ \forall i,j,i\neq j$ and $P_i=\sum_{\mathbf{y}\in Y_i}p(\mathbf{y})$.

The gradient of the BER can be expressed as

$$E_{\mathbf{y}}\left(\mathbf{h}(\mathbf{y},\mathbf{w})\right)=\sum_{i=1}^{V}P_i\mathbf{m}_i\ , \tag{9}$$

where $\mathbf{m}_i:=E_{\mathbf{y}}\left(\mathbf{h}(\mathbf{y},\mathbf{w})\middle|\mathbf{y}\in Y_i\right)$.

Let us assume that from each set Y_i we are allowed to take L_i number of samples. In this way, we have a sample allocation scheme $(L_1,\ldots,L_V)$ for which $\sum_{i=1}^{V}L_i=L$. (This last equation results from the fact that the overall number of samples is restricted to be not larger than a previously fixed L). Stratified sampling (StS) is then defined as estimating the gradient of BER as follows:

$$\mathbf{f}\left(\mathbf{z}_{nL},\ldots,\mathbf{z}_{(n+1)L-1},\mathbf{w}(n)\right):=$$
$$:=\sum_{i=1}^{V}P_i\frac{1}{L_i}\sum_{k=nL_i}^{(n+1)L_i-1}\mathbf{h}\left(\mathbf{z}_k^{(i)},\mathbf{w}(n)\right), \tag{10}$$

where $\mathbf{z}_{nL_i}^{(i)},\ldots,\mathbf{z}_{(n+1)L_i-1}^{(i)}\in Y_i$.

One can see that in StS, the conditioned expected value of a specific class is estimated based on a given number of samples taken from this class.

It can be proven (see [17]) that the optimal sample allocation scheme is

$$L_{i,\text{opt}}=L\frac{P_i\sigma_i}{\sum_{j=1}^{V}P_j\sigma_j} \tag{11}$$

where σ_j denotes the conditional variance of class *j. W*ith this sample allocation scheme the mean square error is much smaller than the one achieved by the simple MC method.

TABLE III
STRATIFIED SAMPLING EQUALIZATION

Step 0		Set iteration number n=0 and initialize $\mathbf{w}(0)$ with an arbitrary vector; $L=V\cdot L_{\text{pre}}+L_{\text{post}}$, $L_{i,\text{pre}}:=L_{\text{pre}}$; perform the following iteration;
n^{th} step of the algorithm	I	generate $\left\{\mathbf{z}_k^{(i)},\ k=1,\ldots,L_{\text{pre}}\right\}$ samples for pre-estimation $i=1,\ldots,V$;
	II	calculate the empirical variances as $\gamma_i:=\frac{1}{L_{\text{pre}}}\sum_{k=1}^{L_{\text{pre}}}\left(\Psi\left(\mathbf{z}_k^{(i)},\mathbf{w}(n)\right)-\frac{1}{L_{\text{pre}}}\sum_{l=1}^{L_{\text{pre}}}\Psi\left(\mathbf{z}_l^{(i)},\mathbf{w}(n)\right)\right)^2$;
	III	set $L_{i,\text{post}}:=L_{\text{post}}\frac{\gamma_iP_i}{\sum_{j=1}^{V}\gamma_jP_j}$;
	IV	generate $\left\{\mathbf{z}_k^{(i)},\ k=1,\ldots,L_{i,\text{post}}\right\}$ samples for post-processing $i=1,\ldots,V$;
	V	generate L sample vectors $\mathbf{z}_{nL},\mathbf{z}_{nL+1}\ldots,\mathbf{z}_{(n+1)L-1}$ subject to uniform distribution;
	VI	perform the estimation of the expected value according to the StS principle: $\mathbf{f}\left(\mathbf{z}_{nL},\ldots,\mathbf{z}_{(n+1)L-1},\mathbf{w}(n)\right):=\sum_{i=1}^{V}P_i\frac{1}{L_i}\sum_{k=nL_i}^{(n+1)L_i-1}\mathbf{h}\left(\mathbf{z}_k^{(i)},\mathbf{w}(n)\right)$;
	VII	adjust the equalizer coefficients: $\mathbf{w}(n+1)=\mathbf{w}(n)-\Delta\cdot\mathbf{f}\left(\mathbf{z}_{nL},\ldots,\mathbf{z}_{(n+1)L-1},\mathbf{w}(n)\right)$;
	VIII	check the stopping criterion: $\left\|\mathbf{w}(n+1)-\mathbf{w}(n)\right\|<\varepsilon$ with a given ε and if the criterion has been fulfilled then the algorithm stops, otherwise increment iteration number n and go back to Step I.

The problem of StS lies in the fact that the conditional variances must be known to determine the optimal sample allocation scheme. In the lack of this knowledge, a portion of the overall sample size must be used to estimate the conditional variances before estimating the conditional means.

This increment in complexity still pays off by having a more accurate estimation on the BER than by using the MC method.

As a result BER estimation by StS is done according to the following two-phase algorithm:

- the first phase is the so-called deviation estimation (estimating the stratum deviation values in order to calculate the sample allocation scheme L_i, $i=1,...,V$),
- while the second phase is dedicated to the expected value estimation.

It is noteworthy that the number of classes V must be determined in advance. The algorithm is described in Table III.

D. BER minimization by Li-Silvester bounds

In order to develop a deterministic estimation of the BER, one can select the first L most dominant samples over which the function $\Psi(\mathbf{w},\mathbf{z})$ has the first L largest values arranged in a descending order:

$$\tilde{Z} := \left\{\mathbf{z}_1,...,\mathbf{z}_L : \Psi(\mathbf{z}_1,\mathbf{w}) \geq ... \geq \Psi(\mathbf{z}_L,\mathbf{w})\right\}. \quad (12)$$

In this way, BER can be upper and lower bounded as follows:

$$\frac{1}{2^{M-1}}\sum_{\mathbf{z}\in\tilde{Z}}\Psi(\mathbf{z},\mathbf{w}) < BER < \frac{1}{2^{M-1}}\sum_{\mathbf{z}\in\tilde{Z}}\Psi(\mathbf{z},\mathbf{w}) + \frac{2^{M-1}-L}{2^{M-1}}$$

due to the fact of $0 \leq \Psi(\mathbf{z},\mathbf{w}) \leq 1$.

The corresponding equalization algorithm is constructed as follows:

TABLE IV
EQUALIZATION BY THE DOMINANT SAMPLES

<table>
<tr><td colspan="2">Step 0</td><td>Set iteration number n=0 and initialize $\mathbf{w}(0)$ with an arbitrary vector;
perform the following iteration;</td></tr>
<tr><td rowspan="4">n^{th} step of the algorithm</td><td>I</td><td>select the first L dominant samples for which (12) holds;</td></tr>
<tr><td>II</td><td>set
$F\left(\mathbf{z}_{nL},...,\mathbf{z}_{(n+1)L-1},\mathbf{w}(n)\right) := \frac{1}{2^{M-1}}\sum_{i=nL}^{(n+1)L-1}\Psi(\mathbf{z}_i,\mathbf{w})$;</td></tr>
<tr><td>III</td><td>adjust the equalizer coefficients:
$\mathbf{w}(n+1) = \mathbf{w}(n) - \Delta\cdot\tilde{\mathbf{f}}\left(\mathbf{z}_{nL},...,\mathbf{z}_{(n+1)L-1},\mathbf{w}(n)\right)$,
where $\tilde{\mathbf{f}}(\mathbf{z},\mathbf{w}) = \underset{\mathbf{w}}{\operatorname{grad}} F(\mathbf{z},\mathbf{w})$;</td></tr>
<tr><td>IV</td><td>check the stopping criterion $\left\|\mathbf{w}(n+1)-\mathbf{w}(n)\right\| < \varepsilon$ with a given ε and if the criterion has been fulfilled then the algorithm stops, otherwise increment iteration number n and go back to Step I.</td></tr>
</table>

One must note two issues in the recursion above: (i) we use the samples to approximate the BER itself instead of its gradient (for the reason of numerical convenience); and (ii) the BER is estimated by the most dominant samples (unlike in the MC method). In this way a much more accurate estimation of the true BER can be obtained. In order to perform this algorithm, one needs to select the most dominant samples in each step of the gradient descent algorithm. First we choose the most dominant sample defined as follows:

$$\mathbf{z}_1 = \underset{\substack{\mathbf{z}\in\{-1;1\}^M \\ z_k=-1}}{\arg\max}\left\{\Phi\left(\frac{\mathbf{z}^T\mathbf{g}}{\sqrt{N_0\mathbf{w}^T\mathbf{g}}}\right)\right\} = \underset{\substack{\mathbf{z}\in\{-1;1\}^M \\ z_k=-1}}{\arg\max}\left\{\mathbf{z}^T\mathbf{g}\right\}. \quad (13)$$

Then the rest of the dominant samples can be selected by the calculating the absolute values of g_i, $i=1,...,M-1; i\neq k$ and arranging them into a monotone decreasing sequence:

$$\left|g_{i_1}\right| \geq \left|g_{i_2}\right| \geq ... \geq \left|g_{i_j}\right| \geq ... \geq \left|g_{i_{M-1}}\right|, \quad (14)$$

where the indices $i_1,...,i_{M-1}$ refer to the ordered sequence.

Based on this sequence, the dominant samples $\left\{\mathbf{z}^{(l)} : l=1,...,L\right\}$ can be calculated as follows:

$$z_{i_j}^{(l)} = (-1)^{u_{i_j}^{(l)}}\operatorname{sgn}\left(g_{i_j}\right), \quad (15)$$

where

$$u_{i_j}^{(l)} = \begin{cases} 0 \text{ if } (j \leq M-l) \cup \left(\operatorname{ind}\min_j\left\{\left|g_{i_{M-l+1}}\right|, C\left(\left|g_{i_{M-l+2}}\right|,...,\left|g_{i_{M-1}}\right|\right)\right\}\right) \\ 1 \text{ otherwise} \end{cases}$$

Having calculated the first L dominant samples by the procedure indicated above, BER can be estimated as follows:

$$F(\mathbf{z},\mathbf{w}) = \frac{1}{2^{M-1}}\sum_{\ell=1}^{L}\Phi\left(\frac{-g_k + \sum_{i=1,i\neq k}^{M} g_i z_i^{(l)}}{\sqrt{N_0\mathbf{w}^T\mathbf{g}}}\right). \quad (16)$$

However, the dominant samples of BER may not necessarily coincide with dominant samples of the gradient. In the case of $\left|R_{ii}\right| >> \sum_{j=1,j\neq i}^{M}\left|R_{ij}\right|$ (light interference), the dominant samples of BER are the same as the ones which maximize the norm of its gradient. In this way selecting the dominant samples of BER, we obtain the dominant samples of the gradient as well. To show this, let us analyze the first dominant sample of the gradient:

$$\mathbf{z}_1 = \underset{\substack{\mathbf{z}\in\{-1;1\}^M \\ z_k=-1}}{\arg\max}\left\{\left\|\mathbf{h}(\mathbf{z},\mathbf{w})\right\|\right\}, \quad (17)$$

where

$$\left\|\mathbf{h}(\mathbf{z},\mathbf{w})\right\| = \varphi\left(\frac{\mathbf{z}^T\mathbf{g}}{\sqrt{N_0\mathbf{w}^T\mathbf{g}}}\right)\left|\frac{\frac{1}{2}N_0}{\left(N_0\mathbf{w}^T\mathbf{g}\right)^{3/2}}\right|\cdot\left\|2\mathbf{R}\mathbf{z}\mathbf{w}^T\mathbf{g} - \mathbf{z}^T\mathbf{g}\mathbf{g}\right\| \geq$$

$$\geq \varphi\left(\frac{\mathbf{z}^T\mathbf{g}}{\sqrt{N_0\mathbf{w}^T\mathbf{g}}}\right)\left|\frac{\frac{1}{2}N_0}{\left(N_0\mathbf{w}^T\mathbf{g}\right)^{3/2}}\right|\cdot\left(2\left|\mathbf{w}^T\mathbf{g}\right|\left\|\mathbf{R}\mathbf{z}\right\| - \left|\mathbf{z}^T\mathbf{g}\right|\left\|\mathbf{g}\right\|\right).$$

Deriving this expression, we took advantage of the fact that when if $\left|R_{ii}\right| >> \sum_{j\neq i}\left|R_{ij}\right|$ then the approximation $\left\|\mathbf{R}\mathbf{z}\right\| \approx M$ holds. This yields that the calculation of the dominant samples are given by minimizing the expression $\left|\mathbf{z}^T\mathbf{g}\right|$. However, this is equivalent with calculating the first L dominant samples of BER, according to the formula (15).

V. Novel CNN based detection combined with equalization

As expression (1) indicated, in the presence of interference optimal detection boils down to quadratic minimization.

Therefore, Cellular Neural Networks (CNNs) can also be applied as detectors in communication systems since their dynamics are governed by a set of nonlinear differential equations defined in [22],**Hiba! A hivatkozási forrás nem található.** which optimizes a quadratic energy function.

The CNN dynamics are expressed by the following differential equations:

$$\dot{x}_{ij} = -x_{ij} + \sum_{C(k,l)\in S_r(i,j)} A(i,j;k,l)\, y_{kl} + \\ + \sum_{C(k,l)\in S_r(i,j)} B(i,j;k,l)\, u_{kl} + z_{ij}, \tag{18}$$

where $x_{ij} \in R$, $y_{kl} \in R$, $u_{kl} \in R$, and $z_{ij} \in R$ are called state, output, input, and threshold of cell $C(i,j)$, respectively. $A(i,j;k,l)$ and $B(i,j;k,l)$ are called the feedback and the input synaptic operator, and the output equation is

$$y_{ij} = f(x_{ij}) = \tfrac{1}{2}\left|x_{ij}+1\right| - \tfrac{1}{2}\left|x_{ij}-1\right| \tag{19}$$

called standard nonlinearity.

The behavior of the corresponding solution is determined by a template matrix

$$\mathbf{A} = \left[a_{kl}\right]_{\substack{k=-T_r,\dots,-1,0,1,\dots,T_r \\ l=-T_r,\dots,-1,0,1,\dots,T_r}} \tag{20}$$

which expresses the local connection pattern among the processing elements. We use CNN as a pure feedback system with zero input and zero bias. The corresponding Lyapunov function is given as $L(\mathbf{y}) = \mathbf{y}^T\mathbf{V}\mathbf{y} - 2\mathbf{r}^T\mathbf{y}$, which prompts the use of CNN as a quadratic optimizer. Here matrix **V** is a sparse matrix determined by the template matrix **A**. As a result, one can theoretically perform optimal detection by assigning $\mathbf{V} := \mathbf{R}$ and $\mathbf{r} = \mathbf{x}$.

The only problem which arises is due to the local connection pattern of CNN which only supports the minimization of quadratic forms generated by sparse matrices. On the other hand, the quadratic form emerging from the detection problem may not necessarily be generated by a sparse matrix (Template Oriented Matrix, TOM).

In order to circumvent the problem, novel equalization algorithms are introduced to modify the channel impulse response in order to enforce the sparse matrix structure. In this approach channel equalization is put to the task of providing optimal characteristics for CNN based detection as opposed to minimizing the main square error for the traditional threshold detection. This new equalizer algorithm is developed under the assumption that the effect of additive Gaussian noise is negligible in comparison with the effect of ISI or MAI. The overall channel impulse response is denoted by q_k and given as $q_n := \sum_{j=0}^{J} w_j h_{n-j}$, where w_j; $j = 0,1,\dots,J$ are the free coefficients of the equalizer subject to optimization. In the noise-less approximation the quadratic form to be minimized for optimal detection is also given by (1) however in this case $\mathbf{W} = \mathbf{Q}^T\mathbf{Q}$ and $\mathbf{b} = \mathbf{Hx}$. Since **W** must fulfill the 1DCM condition, we want to manipulate matrix **Q** by optimizing the equalizer coefficients according to the goal function

$$\mathbf{w}_{\text{opt}} : \min_{\mathbf{w}} \sum_{n,n\geq 2} \left(\sum_{j=0}^{J} w_j h_{n-j} \right)^2 \tag{21}$$

In this way, those elements of **W** which are supposed to be zero in order to satisfy the 1DCM condition will at least be minimal. The corresponding recursive algorithm for the weights of the equalizer (based on the gradient) is given as follows:

$$w_l(k+1) = w_l(k) + \Delta \sum_{n,n\geq 2} \sum_{j=0}^{J} w_j(k)\, h_{n-j} h_{n-l} \tag{22}$$

where Δ is learning parameter used to optimize the convergence speed of the algorithm. This algorithm can be extended to the case of unknown channel by using a corresponding stochastic identification scheme.

VI. Numerical Results

We have analyzed the performance of both the statistical sampling based equalizer and the CNN detector.

The new equalization algorithms have been tested in the case of user numbers M=7 and M=4 with block length 2. The channel was a typical urban model (for specific details see [7]) and the codes have been generated as Gold codes.

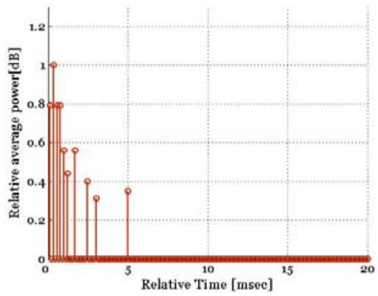

Fig. 4 Impulse response of the urban model

We used the theoretical BPSK BER-SNR curve defined by $BER_{\text{BPSK}} = \int_{-\infty}^{-\sqrt{SNR}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$, as the reference curve (see [20]) to analyze the performance.

A. Performance analysis of statistical sampling based channel equalization

The performance is of the different methods are measured by plotting the Bit Error Rate against the Signal to Noise Ratio (the BER-SNR curve) shown by the Fig. 5.

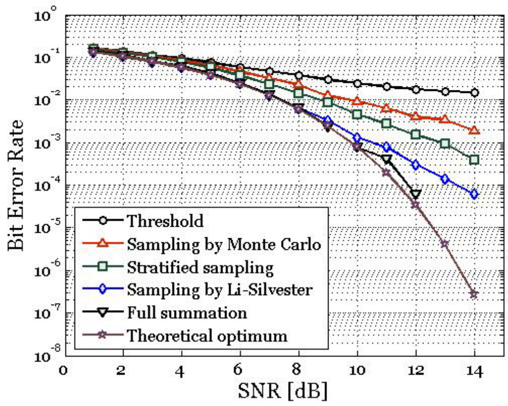

Fig. 5 The performance of different equalizer algorithms

The sample size for each method was selected to be L=20.

In the case of StS, the partition is given as follows:

$$Y_i := \left\{ \mathbf{z} : \mathrm{d}\left(\mathbf{z}_1, \mathbf{z}\right) = i \right\} \; i = 1,...,V \, , \tag{23}$$

where $\mathbf{z}_1 : \max_{\mathbf{z}} \left\{ \mathbf{z}^T \mathbf{g} \right\}$ and $\mathrm{d}(.)$ is the Hamming distance.

As one can see, LS equalization performs better than StS and MC. Furthermore, the performance of LS method may be close to the theoretical optimum which implies that they can indeed increase the spectral efficiency of wireless systems.

The next figure depicts the BER-SNR curves for using an equalizer with degree J=8 (the user number is 7 and the block-length is 2).

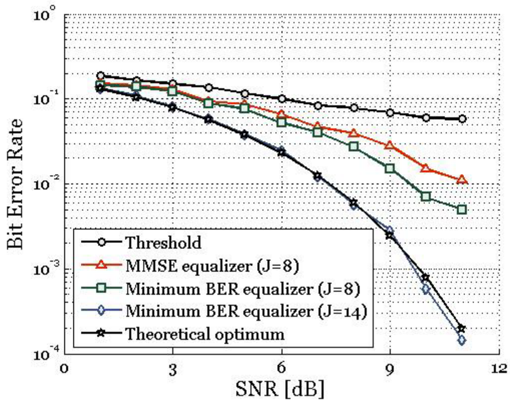

Fig. 6 Equalization performances with low degree equalizers

Here one can also see that MMSE equalizer performs worse than the minimum BER equalizer.

B. Performance analysis of CNN based detection

In the next figure the BER-SNR curves are shown for CNN detector. One can see that CNN based detection both by truncation and by equalization yield superior performance to the original threshold detection. Furthermore, in the case of high SNR (under light noise assumption) the performance of CNN and equalizer indeed yields much better performance than any other method.

(a)

(b)

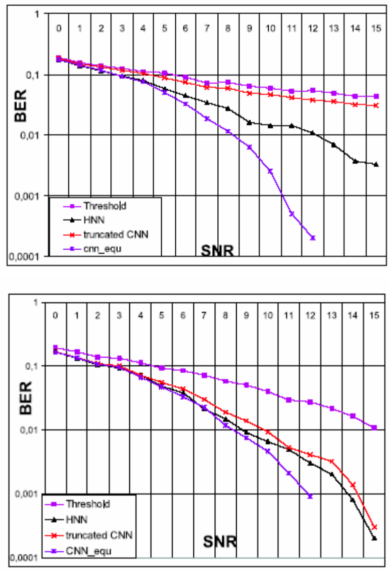

Fig. 7 The simulation of the different CNN detection techniques with (a) urban and (b) rural channel

C. Complexity analysis of statistical sampling based channel equalization

The proposed methods yield real-time equalization algorithms, as the gradient is only estimated by a relatively small number of samples (no exponential summation is needed). In the next bar chart the complexities of the proposed equalizer algorithms are compared. The complexity of a method is expressed as the number of iteration required to fulfill the stopping criterion multiplied by the terms in the summation needed to estimate the gradient. The complexities of the algorithms with respect to SNR are indicated by Fig. 8.

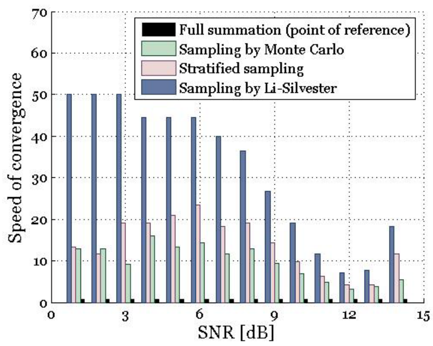

Fig. 8 The convergence of different equalizer algorithms

One can see that the LS based equalization has the smallest complexity.

On Fig. 9, the "dominancy" of the samples are described by plotting the value of each term in the summation of BER. The curve plotted with small circles was obtained by calculating the BER with the initial equalizer vector $\mathbf{w}(0)$. The other curve plotted with small squares was obtained by calculating the BER with the optimal equalizer vector $\mathbf{w}_{\text{opt}}$.

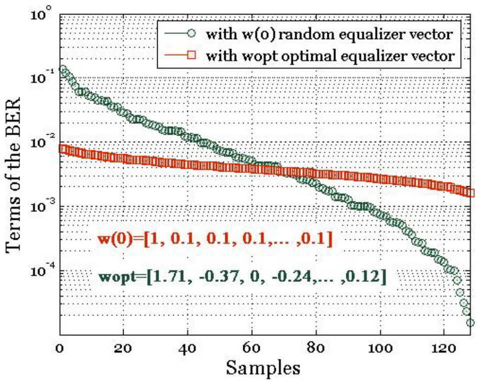

Fig. 9 Distribution of the terms in the summation belonging to BER

One can see, that the samples used to estimate the BER are really dominant at the beginning of the algorithm. Near the end of the algorithm (approaching $\mathbf{w}_{opt}$) the samples are more equally distributed. Therefore, a large computational gain can be obtained by the LS method at the beginning of the algorithm. Nearing to the steady state, the LS method can be replaced by the simple MC method (as the samples are more or less equal), therefore selecting the dominant samples will not considerably speed up the algorithm any longer. Combining the LS sampling with the MC methods (starting with LS sampling and when the weight vector is close to stabilization switching to MC) may further decrease the complexity, because the dominant samples need not be evaluated in each step at the end of the algorithm.

VII. Conclusions

In this paper novel equalization algorithms have been proposed based on different sampling methods. The new methods are capable of direct BER minimization, thus they can achieve better performance than traditional equalization strategies.

The best approach is the one which adopts the LS sampling in each cycle of adaptation of the equalization algorithm. The complexity of the new algorithms is low and they yield fast equalization. As a result, they can contribute to improving the spectral efficiency of multiuser mobile systems.

Acknowledgment

The research carried out here was supported by the Mobile Innovation Centre (MIK) Project 2.1.3 in Hungary.

References

[1] A. Kapur, M.K. Varanasi, Multiuser Detection for Overloaded CDMA Systems, *IEEE Trans. Information Theory*, vol. 49, 2003, pp. 1728-1742.

[2] F. Vanhaverbeke, M. Moeneclaey, Overloaded CDMA systems with displaced binary signatures, *EURASIP Journal on W. Comm. and Net.*, vol. 1, pp. 161-171, 2004;

[3] A.J. Viterbi, *CDMA, Principles of spread spectrum communication*, Addison-Wesley, 1992.

[4] M.K. Varanasi, B. Aazhang, Near-optimum detection in synchronous code division multiple access system, *IEEE Trans. Commun.*, vol. 39, 1991, pp. 725-736.

[5] R. Lupas and S. Verdu, Linear multiuser detectors for synchronous code-division multiple-access channels, *IEEE Trans. Inform. Theory*, vol. 35, 1989, pp. 123-136.

[6] Verdu, S., Shamai, S., Spectral Efficiency of CDMA with Random Spreading, *IEEE Transactions on Information Theory*, vol. 45, no. 2, 1999, pp 622-640.

[7] J.G. Proakis, *Digital Communications*, McGrawHill, 2001.

[8] T. S. Rappaport, *Wireless Communications*, Prentice-Hall, 1996.

[9] A. Sayeed, A. Sendonaris, and B. Aazhang, Multiuser Detection in Fast Fading Multipath Environments, *IEEE Journal on Selected Areas in Comm.*, 1998, pp. 1691-1701.

[10] S. Verdu, *Multiuser detection*, Cambridge University Pres, 1999.

[11] S. Haykin, *Adaptive filter theory*, Prentice Hall, 1996.

[12] R. W. Lucky, Automatic equalization for digital communication, *Bell Syst. Tech. Journal,* vol. 44, pp. 547-588, 1965.

[13] J.G. Proakis, J.H. Miller, An adaptive receiver for digital signaling through channels with intersymbol interference, *IEEE Trans. Information Theory*, vol. IT-45, pp. 484-497, 1969.

[14] H.V. Poor, S. Verdu, Probability of error in MMSE multiuser detection, *IEEE Trans. Commun*., vol. 38., 1997, pp. 858-871.

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

[16] J. Levendovszky, A. Olah, D. Varga, Novel CNN based detection for increased spectral efficiency. *8th IEEE International Biannual Workshop on Cellular Neural Networks and their Applications*, 2004, pp. 483-486.

[17] J. Levendovszky, L. Jereb, Adaptive statistical methods in network reliability analysis, *2nd Symposium on Rare Events simulations,* 2000.

[18] O.K. Li, J.A. Silvester, Performance Analysis of Networks with Unreliable Components, *IEEE Trans. Commun*., vol. COM-32, 1984, pp. 1105-1110.

[19] R. Gold, Optimal Binary Sequences for Spread Spectrum Multiplexing, *IEEE Transactions on Information Theory*, vol. 14. pp 154-156, 1967;

[20] W. G. Teich, M. Seidl, G. Jeney, S. Imre, L. Pap, Code division multiple access communications: multiuser detection based on a recurrent neural network structure, *IEEE Trans. Veh. Technol.*, vol. 46, 1996, pp. 979-984.

[21] Chua, L.O., Roska T. and Venetianer, P.L., The CNN is as Universal as the Turing Machine. *IEEE Trans. on Circuits and Systems*, Vol. 40., March, 1993.

[22] L.O. Chua and T. Roska.: The CNN paradigm. *IEEE Trans. on Circuits and Systems*, Vol. 40, 1993.

Janos Levendovszky received his M.Sc. degree from the Budapest University of Technology and Economics in 1986. He obtained his Ph.D. from the Hungarian Academy of Sciences in 1989 in the filed of adaptive signal processing. He was visiting scholar at Oxford University, UK, and conducted research on neural network theory at the Department of Mathematics, Katholieke Universiteit Leuven, Belgium. Presently, he is professor at the Department of Telecommunications, Budapest University of Technology and Economics and Deputy Dean of the Faculty of Electrical Engineering and Informatics. He teaches and researches in the area of information and communication theory, networking and soft computing. He published numerous papers in the above mentioned fields and is involved in international projects related to ATM networks and neural modeling.

Andras Olah received the M. Sc. Degree in electrical engineering, and the Ph.D degree from the Technical University of Budapest, Hungary, in 2002 and 2006, respectively. His Ph.D dissertation was on the spectral efficiency in wireless systems.

He works as a postdoctor in the Peter Pazmany Catholic University Faculty of Information Technology. His scientific interests include networking and wireless communications, and wireless sensor networks.