Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2008, Article ID 535269, 9 pages
doi:10.1155/2008/535269

*Research Article*

# Iterative Mean Removal Superimposed Training for SISO and MIMO Channel Estimation

**O. Longoria-Gandara,[1] R. Parra-Michel,[1] M. Bazdresch,[2] and A. G. Orozco-Lugo[3]**

[1] *Department of Electrical Engineering, CINVESTAV-IPN, Apartado Postal 31-438, Plaza La Luna, Guadalajara, 44550 JAL, Mexico*
[2] *Department of Electronics, Systems and Computer Science, ITESO, Per. Sur M. Gomez Morin 8585, Tlaquepaque, 45604 JAL, Mexico*
[3] *Department of Electrical Engineering, CINVESTAV-IPN, Apartado Postal 14-740, 07000 Mexico, DF, Mexico*

Correspondence should be addressed to O. Longoria-Gandara, olongo@gdl.cinvestav.mx

Received 2 April 2008; Revised 1 August 2008; Accepted 22 September 2008

Recommended by Fred Daneshgaran

This contribution describes a novel iterative radio channel estimation algorithm based on superimposed training (ST) estimation technique. The proposed algorithm draws an analogy with the data dependent ST (DDST) algorithm, that is, extracts the cycling mean of the data, but in this case at the receiver's end. We first demonstrate that this mean removal ST (MRST) applied to estimate a single-input single-output (SISO) wideband channel results in similar bit error rate (BER) performance in comparison with other iterative techniques, but with less complexity. Subsequently, we jointly use the MRST and Alamouti coding to obtain an estimate of the multiple-input multiple-output (MIMO) narrowband radio channel. The impact of imperfect channel on the BER performance is evidenced by a comparison between the MRST method and the best iterative techniques found in the literature. The proposed algorithm shows a good tradeoff performance between complexity, channel estimation error, and noise immunity.



## 1. INTRODUCTION

One of the most widely used approaches to channel estimation is to employ pilot assisted transmission (PAT), where a known training sequence, also referred as pilot, is inserted at each block of transmitted data [1]. Using the knowledge of the training symbols and the corresponding received signal, the channel estimator block at the receiver is able to make an estimate of the channel impulse response (CIR). However, these training pilots, using time division multiplexing (TDM) scheme, consume valuable bandwidth resulting in the reduction of the data rate.

There are two well-known channel estimation techniques to avoid the bandwidth waste of PAT schemes: superimposed training (ST) [2, 3] and data dependent ST (DDST) [4]. These techniques are based on the arithmetic addition (superimposed) of a training sequence to the information data. Both schemes provide a simple (unsophisticated) channel estimation processes; they differ only in that the cyclic mean of the transmitted data of the DDST scheme is superimposed into the transmitted sequence in similar way that the training signal.

Although DDST outperforms ST [4] in terms of channel estimation error, it is worth mentioning that the decoding of data under DDST is of iterative nature because it needs to extract the data-dependent distortion. Considering this, DDST with DDD removal (henceforth we will refer to this scheme as DDST-DDD removal) gives a similar performance as TDM-based channel estimation, but with fewer bandwidth losses. There are however some drawbacks in trade; DDST technique introduces a delay in the transmitted data in order to calculate the cyclic mean. It also assigns less power to the data signal, and hence the use of high-order symbol constellations has repercussions on the data decoding process [5].

The last constraints lead to research on iterative implementations of ST as in [6–10], starting from the SISO radio channel. These works are based on the use of the decoded data to eliminate the distortion introduced by the received data in the channel estimation process. The first approach uses ST in combination with a traditional least squares channel estimate (LSST) and was developed in [6]. In [6, 7] it is clearly shown that in terms of the channel estimation mean square error (MSE), LSST converges to

a fully trained system for high SNR in two iterations, thereby outperforming both conventional ST and DDST. The great disadvantage of LSST scheme is the computational burden of $\mathcal{O}(NM^2)$ for one iteration, where $N$ is the block length and $M$ is the order of the CIR. In [11], it is demonstrated that DDST-DDD achieves similar performance than LSST but with considerably less complexity.

Alternative iterative procedures of both ST and DDST methods were introduced in [7]. The first one, IST, uses the equalized symbols, obtained via ST, to improve the channel estimate in an iterative way but with less complexity than LSST [6]. The second one, LSDDST employs the LSST iterative approach but based on DDST instead of ST. LSDDST scheme, has the same computational burden as LSST, that is, $\mathcal{O}(NM^2)$, but it converges faster to the fully trained system. In terms of bit error rate (BER), LSDDST scheme yields almost the same BER than LSST.

From the previous works, we are now able to establish this contribution in the context of low-complexity iterative algorithms using ST that shows good BER performance. We introduce a new iterative mean removal ST (MRST) proposal and compare its performance with the previous and most relevant works.

This MRST yields similar performance to DDST-DDD removal and IST but with less complexity when they are compared with LSDDST and LSST. Because the iterative channel estimation methods depend on and work jointly with the equalization stage, we present the results using two equalizers widely used in communication systems: the minimum mean square error (MMSE) equalizer and the maximum likelihood sequence estimation (MLSE) equalizer. The inclusion of both techniques is helpful to accentuate some particularities of the channel estimation methods used.

Additionally, we extend the results of SISO to MIMO case, and study the performance of training-based flat block-fading MIMO channel estimation. Three training-based channel estimators are considered (TDM, DDST, and MRST), which offer different tradeoffs in terms of performance. We analyze the error performance of the MRST method based on the traditional least squares (LSs) method and obtain the corresponding MSE. The proposed MRST estimator for the MIMO case is illustrated using an orthogonal space-time block coder (OSTBC) with two transmit and two receive antennas, that is, Alamouti $2 \times 2$ space-time coding.

This paper is organized as follows. Section 2 deals with the new iterative ST approach in the framework for ST/DDST schemes for SISO systems. The performance analysis of the MRST for MIMO systems is obtained in Section 3. In Sections 4 and 5, the simulation results and performance comparison are given for SISO and MIMO, respectively. Finally, the conclusions are set down in Section 6.

## 2. MEAN REMOVAL ST FOR SISO SYSTEMS

### 2.1. SISO system model

Assuming a frequency-selective channel, Figure 1 depicts the discrete-time baseband block diagram of a digital communication system, where the channel input signal $\{s(k)\}$ represents a succession of information blocks of length $N$, and is given by

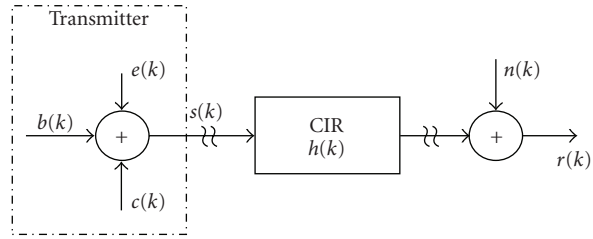


Figure 1: SISO digital communication system model.

$$s(k) = b(k) + c(k) + e(k), \tag{1}$$

taking into account that the block of interest is $k$ indexed and $k = 0, 1, \ldots, N-1$. The sequence $\{b(k)\}$ represents the data with zero mean and variance $E\{|b(k)|^2\} = \sigma_b^2$, $\{c(k)\}$ is a deterministic periodic training sequence with period $P$ and power $\sigma_c^2 = (1/P)\sum_{k=0}^{P-1}|c(k)|^2$, and $\{e(k)\}$ refers to a data-dependent sequence, with period $P$, obtained by periodically repeating $N_P$ times the signal

$$e(j) = -\left(\frac{1}{N_p}\right)\sum_{i=0}^{N_P-1} b(iP+j), \quad j = 0, \ldots, P-1, \tag{2}$$

where $N_p = N/P$.

For the ST case, $e(k) = 0$ and the sequences $\{c(k)\}$, $\{e(k)\}$ are arithmetically added in a superimposed way according to [4], before transmission. Then, the received signal, assuming that exact synchronization and DC-offset are provided, can be observed at the receiver as

$$r(k) = \sum_{l=0}^{M-1} h(l)s(k-l) + n(k), \tag{3}$$

where $h(k)$ is the impulse response of order $M-1$ (i.e., $h(0) \neq 0$ and $h(M-1) \neq 0$), and $\{n(k)\}$ is the complex Gaussian random noise with zero-mean, white, uncorrelated to $\{b(k)\}$ and independent real and imaginary parts with variance $\sigma_n^2 = N_0/2$ per dimension.

Assuming that the channel is quasistatic, that is, the channel is time-invariant during the information block received, and the exact channel order $(M-1)$ is known in advance, then the strong constraint $(P \geq 2M+1)$ showed in [11], with exact synchronization and DC-offset provided, can be relaxed to $P = M$. We used this assumption for the mathematical analysis of the system. Despite this relaxation, all simulations were carried out considering the strong constraint.

### 2.2. Performance analysis

It is clear that DDST method must compute the sequence $\{e(k)\}$ from the data block that will be transmitted; consequently this data processing has an impact on the total delay

of every transmitted block. On the other hand, the sequence $\{s(k)\}$ has a wider dynamic range in comparison with the sequence obtained with ST method. This results in a higher peak-to-average power ratio (PAPR) of the preamplifier communication building block. Another implication using DDST is the fact that the sequence $\{b(k)\}$ will have less power in comparison again with ST and then less noise immunity. For these reasons, iterative ST schemes are very attractive.

We developed the MRST scheme starting from the hypothesis that if we could obtain an estimate of the signal $\{e(k)\}$ (i.e., a cyclic mean of sequence $\{b(k)\}$) at the receiver side, then we would achieve the performance of DDST in terms of channel error estimate MSE but with more power assigned to the data sequence $\{b(k)\}$, having an impact on a better performance in terms of BER.

Estimating the cycling mean of period $P$ as in [3], we can write

$$\tilde{y}(j) = \frac{1}{N_P}\sum_{i=0}^{N_P-1} r(iP+j), \tag{4}$$

with $j = 0, 1, \ldots, P-1$.

Combining (3) and (4) yields

$$\tilde{y}(j) = \sum_{m=0}^{M-1} h(m)\left\{\tilde{b}(j-m) + e(j-m) + c(j-m)\right\} + \tilde{n}(j), \tag{5}$$

where

$$\tilde{b}(k) = \frac{1}{N_P}\sum_{i=0}^{N_P-1} b(iP+j), \qquad \tilde{n}(k) = \frac{1}{N_P}\sum_{i=0}^{N_P-1} \tilde{n}(iP+j). \tag{6}$$

From (5), it follows that in matrix form

$$\left(\mathbf{C} + \tilde{\mathbf{B}} + \mathbf{E}\right)\mathbf{h} = \tilde{\mathbf{y}} - \tilde{\mathbf{n}}, \tag{7}$$

where $\mathbf{C}$ and $\mathbf{E}$ are $P\times P$ circulant matrices with first columns $[c(0)\, c(1)\, \cdots\, c(P-1)]^T$ and $[e(0)\, e(1)\, \cdots\, e(P-1)]^T$, respectively, and $\mathbf{h} = [h(0)\, h(1)\, \cdots\, h(P-1)]$. The $P\times 1$ column vectors $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{n}}$ have similar expression to $\mathbf{h}$. Assuming a cyclic prefix of length $P$ (as was done in [4, 7]), $\tilde{\mathbf{B}}$ is circulant matrix with first column $[\tilde{b}(0)\, \tilde{b}(1)\, \cdots\, \tilde{b}(P-1)]^T$.

For the ST case (i.e., when $\mathbf{E} = \mathbf{0}$ in (7)) we have $\tilde{\mathbf{y}} = (\mathbf{C} + \tilde{\mathbf{B}})\mathbf{h} + \tilde{\mathbf{n}}$, and using the channel estimate $\mathbf{C}^{-1}\tilde{\mathbf{y}}$ from [3] then

$$\hat{\mathbf{h}}_{\mathrm{ST}} = \mathbf{C}^{-1}\tilde{\mathbf{y}} = \mathbf{h} + \mathbf{C}^{-1}\tilde{\mathbf{B}}\mathbf{h} + \mathbf{C}^{-1}\tilde{\mathbf{n}}. \tag{8}$$

For the DDST case $\mathbf{E} = -\tilde{\mathbf{B}}$ in (7), so we have the channel estimate [4] as

$$\hat{\mathbf{h}}_{\mathrm{DDST}} = \mathbf{C}^{-1}\tilde{\mathbf{y}} = \mathbf{h} + \mathbf{C}^{-1}\tilde{\mathbf{n}}. \tag{9}$$

Clearly, the difference between these two channel estimation schemes is the factor $\mathbf{C}^{-1}\tilde{\mathbf{B}}\mathbf{h}$.

Now using the hypothesis early mentioned (i.e., $\mathbf{E} = -\hat{\tilde{\mathbf{B}}} \simeq -\tilde{\mathbf{B}}$ with ST scheme), we proceed to make the channel estimation for the MRST scheme, then (7) becomes

$$\tilde{\mathbf{y}} = \left(\mathbf{C} + \tilde{\mathbf{B}} - \hat{\tilde{\mathbf{B}}}\right)\mathbf{h} + \tilde{\mathbf{n}}, \tag{10}$$

and multiplying (10) by $\mathbf{C}^{-1}$ we obtain

$$\hat{\mathbf{h}}_{\mathrm{MRST}} = \mathbf{C}^{-1}\tilde{\mathbf{y}} = \mathbf{C}^{-1}\left\{\left(\mathbf{C} + \tilde{\mathbf{B}} - \hat{\tilde{\mathbf{B}}}\right)\mathbf{h} + \tilde{\mathbf{n}}\right\}, \tag{11}$$

and finally

$$\hat{\mathbf{h}}_{\mathrm{MRST}} = \mathbf{h} + \mathbf{C}^{-1}\left(\tilde{\mathbf{B}} - \hat{\tilde{\mathbf{B}}}\right)\mathbf{h} + \mathbf{C}^{-1}\tilde{\mathbf{n}}. \tag{12}$$

From (12) it follows that if $\hat{\tilde{\mathbf{B}}} = \tilde{\mathbf{B}}$ then $\hat{\mathbf{h}}_{\mathrm{MRST}} = \hat{\mathbf{h}}_{\mathrm{DDST}}$. In order to make a good estimate $(\hat{\tilde{\mathbf{B}}})$ of $\tilde{\mathbf{B}}$, the following steps are made.

(1) Use (8) to have an initial channel estimate as plain ST and make $\tilde{\mathbf{y}}_{\mathrm{old}} = \tilde{\mathbf{y}}$.

(2) Use the channel estimated to obtain the equalized symbols and employ a hard decision detector.

(3) Use the hard-decision symbols detected to calculate $\hat{\tilde{\mathbf{B}}}$.

(4) Remove $\hat{\tilde{\mathbf{B}}}$ from the received signal to obtain a new $\tilde{\mathbf{y}}$ according to

$$\tilde{\mathbf{y}}_{\mathrm{new}} = \tilde{\mathbf{y}}_{\mathrm{old}} - \hat{\tilde{\mathbf{B}}}\hat{\mathbf{h}}, \tag{13}$$

where $\tilde{\mathbf{y}}_{\mathrm{new}}$ is a column vector $P\times 1$.

(5) Use (11) with $\tilde{\mathbf{y}} = \tilde{\mathbf{y}}_{\mathrm{new}}$ and update the channel estimate $\hat{\mathbf{h}}_{\mathrm{MRST}}$.

(6) Go to step 2 and repeat as need it.

Defining the $\mathrm{MSE}(\hat{\mathbf{h}}) := \mathrm{E}\{\sum_{k=0}^{M-1}|\hat{h}(k) - h(k)|^2\}$, then from [3, 4], and having a better estimation $(\hat{\tilde{\mathbf{B}}})$ of $\tilde{\mathbf{B}}$, it follows that $\hat{\mathbf{h}}_{\mathrm{MRST}} \simeq \hat{\mathbf{h}}_{\mathrm{DDST}}$, then

$$\mathrm{MSE}\left(\hat{\mathbf{h}}_{\mathrm{MRST}}\right) \simeq \frac{\sigma_n^2}{N_p\sigma_c^2}. \tag{14}$$

From (9) and (11), DDST and MRST, respectively, we note the fact that $\mathbf{C}^{-1}$ is calculated only once beforehand. The IST scheme, however, uses the estimate $(\hat{\tilde{\mathbf{B}}})$ of $\tilde{\mathbf{B}}$ to calculate $(\mathbf{C} + \hat{\tilde{\mathbf{B}}})^{-1}$ (see [7, (12)]) for every iteration and information block received. This action implies an additional complexity to the computational burden.

On the other hand, as we will see in Section 4, the estimate $\hat{\tilde{\mathbf{B}}}$ used to estimate the channel in IST and MRST methods depends on what type of equalizer is used. Conversely, the scheme DDST-DDD removal cannot take advantage of MLSE equalizer, because this equalizer delivers hard-decision symbols, and DDST-DDD removal uses the equalized symbols previous to hard-decision procedure.

Table 1 summarizes the computational burden, the computation of inverse matrix $(\mathbf{C}^{-1})$, and the MSE performance approach. The complexity of an algorithm is a quite important metric when the algorithm is to be implemented in HW, as this is the case we are very concern of this value. As for computational burden metric, we choose the number

Table 1: Synoptic Table of the Computational burden, $\mathbf{C}^{-1}$ Computation and the Channel Estimate MSE.

| Procedure | Computational burden | $\mathbf{C}^{-1}$ | MSE($\hat{\mathbf{h}}$) approach |
|---|---|---|---|
| DDST-DDD removal [11] | $\mathcal{O}(\mathcal{Q}N)$ 1 iteration | Pre-calculated | DDST ([4]) |
| IST scheme in [7] | $\mathcal{O}(3\mathcal{Q}N)$ 2 iterations | $(\mathbf{C}+\hat{\tilde{\mathbf{B}}})^{-1}$ every iteration | DDST ([4]) |
| MRST | $\mathcal{O}(3\mathcal{Q}N)$ 2 iterations | Precalculated | DDST ([4]) |

$\mathcal{O}(\cdot)$ notation expresses complexity, $\mathcal{Q}$, and $N$ denote the equalizer length and the block length, respectively.

of iterations as function of the transmitted block length $N$ and the equalizer length $\mathcal{Q}$, to be precise, and the number of coefficients in the equalizer.

Although DDST has the least computational complexity at the receiver side, it needs to calculate the data-dependent sequence $\{e(k)\}$, that implies additional complexity and time delay. Furthermore, it assigns less transmission power to the data signal.

## 3. MEAN REMOVAL ST FOR MIMO SYSTEMS

### 3.1. MIMO system model

We consider a wireless MIMO communication link with $n_T$ transmit and $n_R$ receive antennas, operating in a Rayleigh flat-fading environment. The fading coefficient $h_{i,j}$ is the complex path gain from transmit antenna $i$ to receive antenna $j$. We assume that the coefficients are independently complex circular symmetric Gaussian with unit variance, and then $\mathbf{H} = [h_{i,j}] \in \mathbb{C}^{n_T \times n_R}$. The expression for the received symbols can be expressed as

$$\mathbf{r} = \mathbf{x}\mathbf{H} + \mathbf{n}, \tag{15}$$

where $\mathbf{r}$ is the $1 \times n_R$ received signal vector, $\mathbf{x}$ is the $1 \times n_T$ transmitted signal vector, and $\mathbf{n}$ is an $1 \times n_R$ vector of additive noise terms, assuming that noise is spatially and temporally white Gaussian with zero mean and independent real and imaginary parts with variance $\sigma_n^2 = N_0/2$ per dimension.

Let us assume the block transmission scheme with the block length $N$ at times $k = 1, \ldots, N$, and we also assume that the channel matrix $\mathbf{H}$ remains constant within a block of $N$ symbols, that is, the block length is much small than the channel coherence time. Under these assumptions, the channel, within one block, can be written as

$$\mathbf{R} = \mathbf{X}\mathbf{H} + \mathbf{N}, \tag{16}$$

where

$$\mathbf{R} \triangleq \begin{bmatrix} \mathbf{r}(1) \\ \mathbf{r}(2) \\ \vdots \\ \mathbf{r}(N) \end{bmatrix}, \qquad \mathbf{X} \triangleq \begin{bmatrix} \mathbf{x}(1) \\ \mathbf{x}(2) \\ \vdots \\ \mathbf{x}(N) \end{bmatrix}, \qquad \mathbf{N} \triangleq \begin{bmatrix} \mathbf{n}(1) \\ \mathbf{n}(2) \\ \vdots \\ \mathbf{n}(N) \end{bmatrix} \tag{17}$$

are the matrices of the received signal, transmitted signals, and noise, respectively [12].

Let us denote the set of complex information symbols prior to space-time encoding as $\{s_\lambda \mid s_\lambda \in \mathcal{U}_\lambda, \lambda = 1, 2, \ldots, \Lambda\}$, where each $\mathcal{U}_\lambda$ denotes a set of signal constellation points. The symbols $s_1, s_2, \ldots, s_\Lambda$ are zero-mean mutually uncorrelated random variables. Let us introduce the vector

$$\mathbf{s} \triangleq [s_1 \; s_2 \; \cdots \; s_\Lambda]^T, \tag{18}$$

where $(\cdot)^T$ denotes transpose. Note that $\mathbf{s} \in \mathcal{S}$, where $\mathcal{S} = \{\mathbf{s}^{(1)} \; \mathbf{s}^{(2)} \; \cdots \; \mathbf{s}^{(L)}\}$ is the set of all possible symbol vectors and $L$ is the cardinality of this set. The $N \times n_T$ complex matrix-valued function $\mathbf{X}(\mathbf{s})$ is called OSTBC [13] if it satisfies that

(1) all the entries of $\mathbf{X}(\mathbf{s})$ are linear functions of the $\Lambda$ complex variables $s_1, s_2, \ldots, s_\Lambda$ and their complex conjugates;

(2) for any arbitrary $\mathbf{s}$ in $\mathbb{C}^{\Lambda \times 1}$,

$$\mathbf{X}^H(\mathbf{s})\mathbf{X}(\mathbf{s}) = \|\mathbf{s}\|^2 \mathbf{I}_{n_T}, \tag{19}$$

where $\mathbf{I}_{n_T}$ is the identity $n_T \times n_T$ matrix, $\|\cdot\|$ is the Euclidean norm, and $(\cdot)^H$ denotes Hermitian transpose.

### 3.2. Performance analysis

In order to estimate the channel matrix $\mathbf{H}$, it should be emphasized that in any statistical expectation below, the matrix $\mathbf{H}$ is treated as random; at the same time, any estimator of $\mathbf{H}$ is supposed to obtain an estimate of a particular realization of this random matrix that corresponds to the current block of the received data.

In the conventional ST estimation technique for MIMO systems [9], a known training $N \times n_T$ matrix $\mathbf{C}$, is added arithmetically to the data $N \times n_T$ matrix $\mathbf{B}$ during every block transmitted. In this way, the transmitted signal matrix $\mathbf{X}$ can be expressed as $\mathbf{X} = \mathbf{B} + \mathbf{C}$. It is clear that the total transmitted power is distributed between the data and the training signals, that is, $P_x = P_b + P_c$.

The ST system is depicted in Figure 2, where a signal matrix $\mathbf{X}$ is transmitted over the radio MIMO channel with the channel matrix $\mathbf{H}$, and distorted with the noise matrix $\mathbf{N}$.

Based on the received signal $\mathbf{R}$, the MRST delivers an estimate of the MIMO channel matrix, denoted as $\hat{\mathbf{H}}$; subsequently the decoder obtains an estimate of $\mathbf{X}$, denoted as $\hat{\mathbf{X}}$, and an estimate of the mean of the data, $\hat{\tilde{\mathbf{b}}}$, which can be used by the channel estimator block, in an iterative way, to provide the decoder with a better estimate of $\mathbf{H}$.

Now, the task of this channel estimation algorithm is to recover the channel matrix $\mathbf{H}$ based on the knowledge of $\mathbf{R}$ and $\mathbf{C}$. Assuming flat-frequency channel conditions then all the row vectors of the training matrix $\mathbf{C}$ can be equal. Hence, a time-domain estimator based on the synchronized

averaging of the received signal can be implemented. The time-average of the signal $\mathbf{R}$ is given by

$$\begin{aligned}\mathbf{v} = E\{\mathbf{R}\} &= \left(\frac{1}{N}\right)\mathbf{uR} \\ &= \left[\left(\frac{1}{N}\right)(\mathbf{uB}+\mathbf{uC})\right]\mathbf{H} + \left(\frac{1}{N}\right)\mathbf{uN} \\ &= \left[\tilde{\mathbf{b}}+\mathbf{c}\right]\mathbf{H} + \tilde{\mathbf{n}},\end{aligned} \tag{20}$$

where $\mathbf{v}$ is a $1 \times n_R$ column vector, $\mathbf{u}$ is the unit $1 \times N$ vector, $\mathbf{c}$ is the $1 \times n_T$ training vector repeated $N$ times and $\tilde{\mathbf{b}}$, $\tilde{\mathbf{n}}$ represent the time-average of the data matrix $\mathbf{B}$ and noise matrix $\mathbf{N}$, respectively, for every block transmitted.

Using the LS approach [14, 15], an estimate of $\mathbf{H}$ can be obtained as

$$\hat{\mathbf{H}}_{\mathrm{LS}} = \left(\mathbf{c}^H\mathbf{c}\right)^{-1}\mathbf{c}^H\mathbf{v} = \mathbf{c}^{\dagger}\mathbf{v}, \tag{21}$$

where $(\mathbf{c}^H\mathbf{c})^{-1}\mathbf{c}^H$ is the pseudoinverse of $\mathbf{c}$.

We will use the following transmitted training power constraint

$$\mathbf{c}^H\mathbf{c} = P_c\mathbf{I}, \tag{22}$$

where $\mathbf{I}$ is the $n_T \times n_T$ identity matrix; now the channel estimate is

$$\hat{\mathbf{H}}_{\mathrm{LS}} = \mathbf{H} + \left(\frac{1}{P_c}\right)\mathbf{c}^H\left(\tilde{\mathbf{b}}\mathbf{H}+\tilde{\mathbf{n}}\right). \tag{23}$$

It is clear that the average $\tilde{\mathbf{b}}$ over the data signal $\mathbf{B}$ represents an extra term for the channel estimate. This average is exploited by the DDST method at the transmitter or by the MRST method at the receiver.

In order to obtain the MSE achieved by MRST method, we use the performance analysis of the DDST method, explicitly we use $\tilde{\mathbf{b}} = (1/N)\mathbf{uB}$ and define the perturbation matrix $\mathbf{E} = -\mathbf{u}^T\tilde{\mathbf{b}}$ that will be arithmetically added to the data signal $\mathbf{B}$ every transmitted block. Hence, the transmitted signal is $\mathbf{X} = \mathbf{B}+\mathbf{E}+\mathbf{C}$ and the corresponding received signal is $\mathbf{R} = (\mathbf{B}+\mathbf{E}+\mathbf{C})\mathbf{H}+\mathbf{N}$, then from (21) we obtain

$$\hat{\mathbf{H}}_{\mathrm{LS}} = \mathbf{H} + \left(\frac{1}{P_c}\right)\mathbf{c}^H\tilde{\mathbf{n}}. \tag{24}$$

Considering (22) under optimal training and $\mathrm{E}\{\mathrm{tr}(\mathbf{N}^{\mathrm{H}}\mathbf{N})\} = \sigma_n^2 n_R$, the MSE for DDST [9] is given by

$$\mathrm{MSE}\left(\hat{\mathbf{H}}_{\mathrm{DDST}}\right) = \frac{\sigma_n^2 n_T n_R}{NP_c}. \tag{25}$$

Instead to take away the contribution of $\tilde{\mathbf{b}}$ at the transmitter, the MRST uses the plain ST and removes $\tilde{\mathbf{b}}$ in an iterative manner at the receiver. From (21), the channel estimation error can be expressed as

$$\begin{aligned}\hat{\mathbf{H}}_{\mathrm{MRST}} &= \mathbf{c}^{\dagger}\left(\left[\tilde{\mathbf{b}}+\mathbf{c}\right]\mathbf{H} - \hat{\tilde{\mathbf{b}}}\hat{\mathbf{H}} + \tilde{\mathbf{n}}\right) \\ &= \mathbf{H} + \left(\frac{1}{P_c}\right)\mathbf{c}^H\left(\tilde{\mathbf{b}}\mathbf{H} - \hat{\tilde{\mathbf{b}}}\hat{\mathbf{H}}\right) + \left(\frac{1}{P_c}\right)\mathbf{c}^H\tilde{\mathbf{n}}.\end{aligned} \tag{26}$$

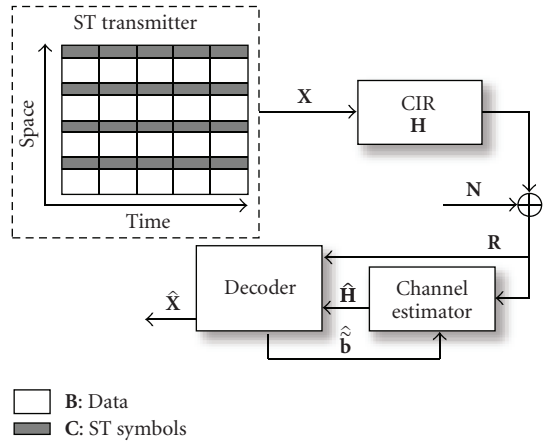


Figure 2: MIMO digital communication system model.

It follows that if $\hat{\mathbf{H}}\hat{\tilde{\mathbf{b}}} \simeq \mathbf{H}\tilde{\mathbf{b}}$, then $\hat{\mathbf{H}}_{\mathrm{MRST}} \simeq \hat{\mathbf{H}}_{\mathrm{DDST}}$, therefore the $\mathrm{MSE}(\hat{\mathbf{H}}_{\mathrm{MRST}}) \simeq \mathrm{MSE}(\hat{\mathbf{H}}_{\mathrm{DDST}})$ as it will be corroborated by the simulation results in Section 5. With respect to the iterative procedure to obtain a better estimate of $\tilde{\mathbf{b}}$, we follow the steps explained for the SISO case.

### 3.3. MRST with Alamouti space-time coding

Let us use MRST estimation method with the $\mathcal{G}2$-OSTBC $n_T = 2$, $n_R = 2$ system (Alamouti coding scheme). Let us assume $N$ to be even, and

$$\mathbf{X} = \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \\ s_3 & s_4 \\ -s_4^* & s_3^* \\ \vdots & \vdots \\ s_{N-1} & s_N \\ -s_N^* & s_{N-1}^* \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} b_1 & b_2 \\ -b_2^* & b_1^* \\ b_3 & b_4 \\ -b_4^* & b_3^* \\ \vdots & \vdots \\ b_{N-1} & b_N \\ -b_N^* & b_{N-1}^* \end{bmatrix},$$
$$\mathbf{C} = \begin{bmatrix} c_1 & c_2 \\ -c_2^* & c_1^* \\ c_1 & c_2 \\ -c_2^* & c_1^* \\ \vdots & \vdots \\ c_1 & c_2 \\ -c_2^* & c_1^* \end{bmatrix} \tag{27}$$

are the matrices of the transmitted signal, data signal, and training signal, respectively, with $(\cdot)^*$ denoting complex conjugation. The matrix's rows and columns indicate the transmission time and transmit antenna, respectively. Assuming flat-fading scenario, no cyclic prefix is required, and the signal training selection can be done choosing two symbols $c_1$, $c_2$ and exploiting the property of orthogonally achieved with the OSTBC system, that is, the two column vectors of the matrix $\mathbf{C}$ are orthogonal.

The estimate of the cycling mean in flat-fading scenario can be done obtaining the mean of the receive matrix $\mathbf{R}$, denoted as $\widetilde{\mathbf{R}}$. Each element of $\widetilde{\mathbf{R}}$ is given by

$$\widetilde{\mathbf{R}}_{i,j} = \frac{1}{N/2}\sum_{k=1}^{N/2}\mathbf{R}_{i+2(k-1),j}, \tag{28}$$

where the indexes $i, j$ correspond to Alamouti's emission time ($i = 1, 2$) and transmit antenna ($j = 1, 2$) respectively for this particular block coding transmission.

Let us estimate the channel matrix $\mathbf{H}$ using (28) and assuming a noisy free scenario (without loss of generality), hence, we have

$$\mathbf{A}\,\mathrm{vec}\{\mathbf{H}\} = \mathrm{vec}\left\{\widetilde{\mathbf{R}}\right\},$$

$$\text{with } \mathbf{A} = \begin{bmatrix} \left(c_1 - \widehat{\widetilde{b}}_1\right) & \left(c_2 - \widehat{\widetilde{b}}_2\right) & 0 & 0 \\ 0 & 0 & \left(c_1 - \widehat{\widetilde{b}}_1\right) & \left(c_2 - \widehat{\widetilde{b}}_2\right) \\ -\left(c_2 - \widehat{\widetilde{b}}_2\right)^* & \left(c_1 - \widehat{\widetilde{b}}_1\right)^* & 0 & 0 \\ 0 & 0 & -\left(c_2 - \widehat{\widetilde{b}}_2\right)^* & \left(c_1 - \widehat{\widetilde{b}}_1\right)^* \end{bmatrix}, \tag{29}$$

and $\mathrm{vec}\{\cdot\}$ is the vectorization operator stacking all rows of a matrix on top of each other. Therefore, the channel estimate is given by

$$\mathrm{vec}\left\{\widehat{\mathbf{H}}\right\} = \mathbf{A}^{-1}\mathrm{vec}\left\{\widetilde{\mathbf{R}}\right\}. \tag{30}$$

Because matrix $\mathbf{A}$ is unitary, then (30) can be rewritten as

$$\mathrm{vec}\left\{\widehat{\mathbf{H}}\right\} = \mathbf{A}^{H}\mathrm{vec}\left\{\widetilde{\mathbf{R}}\right\}, \tag{31}$$

and finally,

$$\widehat{\mathbf{H}} = \left[\mathrm{vec}\{\mathbf{I}_{n_R}\}^T \otimes \mathbf{I}_{n_T}\right]\left[\mathbf{I}_{n_R} \otimes \mathrm{vec}\left\{\widehat{\mathbf{H}}\right\}\right], \tag{32}$$

where $\mathbf{I}_{n_T}$ and $\mathbf{I}_{n_R}$ are the $n_T \times n_T$ and $n_R \times n_R$ identity matrix, respectively, and $\otimes$ denotes the Kronecker product. The matrix $\widehat{\mathbf{H}}$ represents the new MIMO channel estimate used in the iterative procedure to get a new estimate of the mean of the data-bearing ($\widehat{\widetilde{\mathbf{b}}}$).

The key idea of this implementation is the way to remove the mean at the receiver. Instead of subtracts the mean of the data, from the received matrix $\mathbf{R}$, we incorporate the mean estimate in the matrix $\mathbf{A}$ and use the Alamouti decoding procedure to estimate the channel matrix.

## 4. SIMULATION AND RESULTS FOR SISO SYSTEMS

Equalization is a well-known technique used to combat intersymbol interference (ISI) whereby the receiver attempts to compensate for the effects of the channel of the transmitted symbols. An equalizer attempts to determine the transmitted data from the received distorted symbols using an estimate of the channel that caused the distortions. In this contribution, we consider two types of equalizers widely used in communications systems: the MMSE and MLSE equalizers. The MMSE has lesser complexity and performance in terms of BER than the MLSE equalizer that is optimal for ISI.

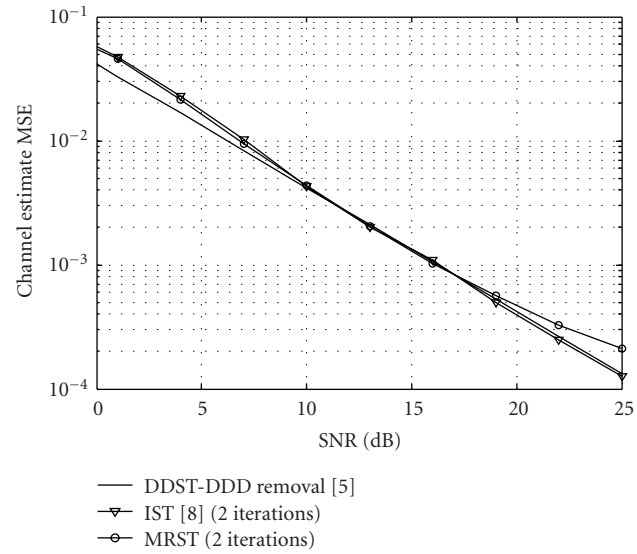


Figure 3: Channel estimate MSE with MMSE equalizer.

### 4.1. SISO system using MMSE equalizer

We considered a time-invariant random three-tap frequency-selective Rayleigh fading channel $h(k)$ with $h(0) \neq 0$ and $h(2) \neq 0$. The channel coefficients were complex Gaussian, i.i.d. with unit variance, rescaled to achieve unitary mean energy. The sequence $\{b(k)\} \in \{-1, +1\}$ is an R.V with uniform p.d.f. and variance $\sigma_b^2$. The parameters $\sigma_b^2$ and $\sigma_c^2$ are chosen for ST and DDST independently such that $\sigma_{b+e}^2 + \sigma_c^2 = 1$. The training to information power ratio $\sigma_c^2/\sigma_{b+e}^2$ was set to $-6.9798$ dB with $P = 7$. The block length is fixed to $N = 420$ and a cyclic prefix of length $P$ added at the beginning of each block in both ST and DDST methods. The channel estimated is used to design an MMSE equalizer of length 11 and equalization delay $d = 7$. All simulations that were run until 1000 blocks with errors were found.

Figure 3 shows the channel estimate MSE obtained from the three iterative procedures considered (a) DDST with DDD removal exposed in [11], (b) IST scheme introduced in [7], and (c) MRST method presented here. It can be observed that IST and MRST have a significant approach to DDST (from 10 dB to 20 dB) just after two iterations.

Figure 4 shows the BER performance comparison where for 2 iterations, all schemes show similar behavior for low and medium SNR levels; DDST exhibits a slight advantage for high SNRs because it achieves a better channel estimate at these levels as is shown in Figure 3.

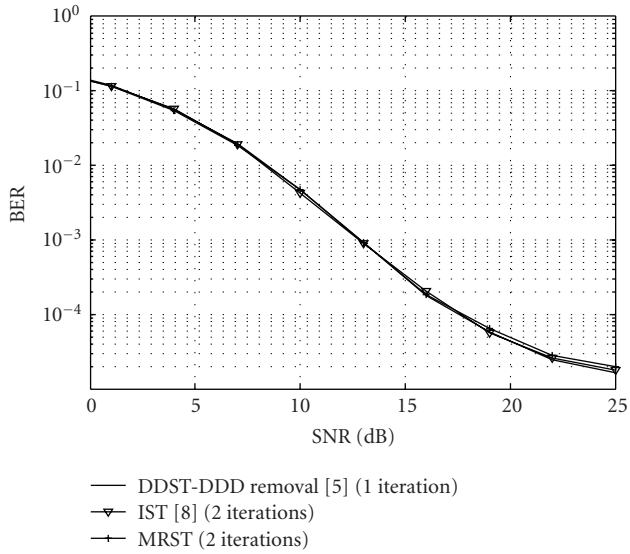


Figure 4: BER performance of BPSK signals with MMSE equalizer.

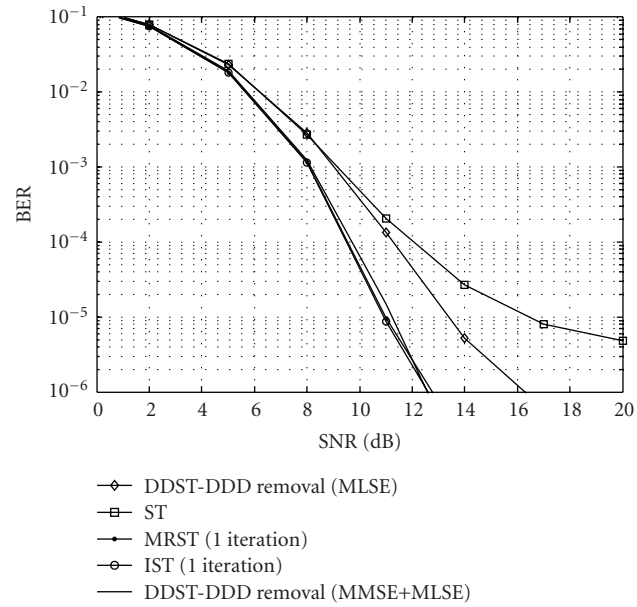


Figure 5: BER performance of BPSK signals with MLSE equalizer.

### 4.2. SISO system using MLSE equalizer

In order to get a better estimate ($\hat{\tilde{\mathbf{B}}}$) of $\tilde{\mathbf{B}}$ for both IST and MRST schemes, we use an MLSE equalizer with traceback length of 11. To be more specific, this equalizer comprises a Viterbi algorithm, which finds the most likely data sequence transmitted. Although, to perform close to ideal MLSE, the equalizer requires traceback lengths of the order of 5-6 times the ISI span, we chose the traceback length of 11, that is, the same number of taps of the MMSE equalizer used in Section 4.1.

Due to the fact that DDST-DDD removal [11] works with the equalizer output before proceeding with a new hard-decision process, that is, equalization and detection stages are carried out separately, it cannot exploit the benefits of the MLSE equalizer. In order to have similar conditions for the DDST method in the decoding procedure, firstly we used the MMSE equalizer, secondly we removed the data distortion, and finally we used the MLSE equalizer to obtain the data symbols. Figure 5 shows the BER performance of these three methods and the conventional ST method.

Clearly, we observe the benefits of use IST or MRST coupling with an MLSE equalizer, where both schemes outperform DDST-DDD removal when it only uses the MLSE equalizer. However, there is no noticeable performance difference when DDST uses the concatenation of the MMSE and MLSE equalizers.

Figure 6 shows the block error rate (BLER) performance. BLER is the statistical measurement of the ratio of the number of blocks with error received to the total number of blocks transmitted and it is part of the performance requirements of 3GPP test. We observe that there is no noticeable performance difference between the MRST and IST with one iteration, and DDST-DDD removal using the equalizer concatenation and removing the data distortion.

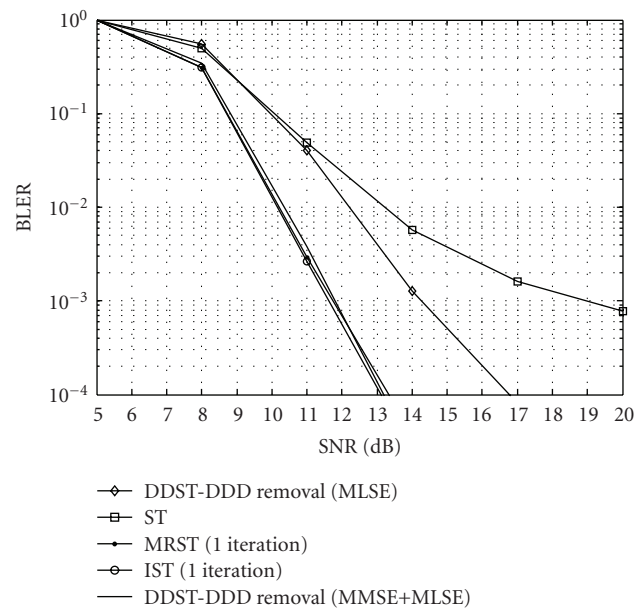


Figure 6: BLER performance of BPSK signals with MLSE equalizer using the three iterative methods and ST.

## 5. SIMULATION AND RESULTS FOR MIMO SYSTEMS

In what follows, we illustrate the performance of the proposed scheme on MIMO systems working in a flat-fading scenario. We use the $\mathcal{G}2$-OSTBC system, $n_T = n_R = 2$ transmit and receive antennas, respectively, with ideally uncorrelated elements. The block length is fixed to $N = 256$ symbols, and all simulations were run until 1000 blocks were in error. The BER is represented as a function of the average SNR, where $N_0 = E_s n_T/\text{SNR}$, and $E_s$ is the average energy per symbol.

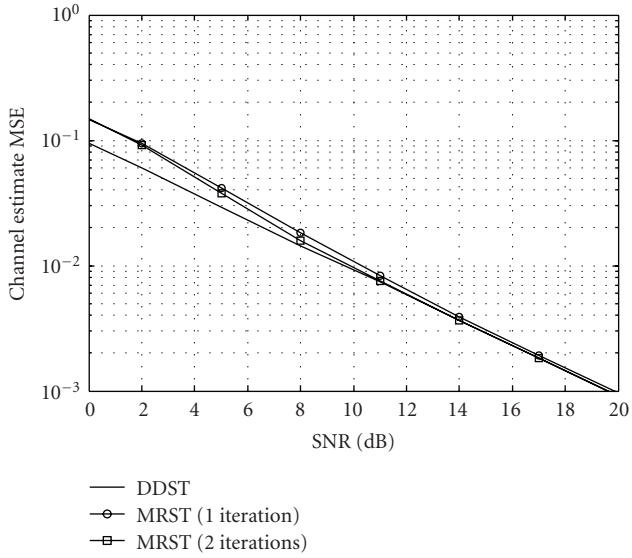


Figure 7: Channel estimate MSE of $\mathcal{G}2$-OSTBC $2 \times 2$.

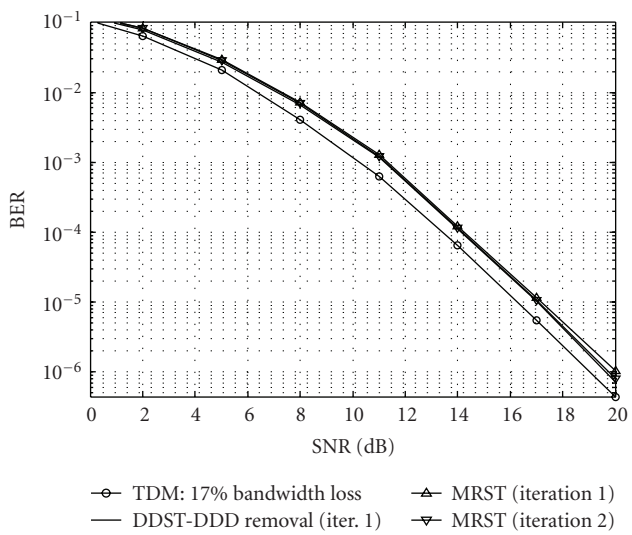


Figure 8: BER comparisons for various channel estimation techniques.

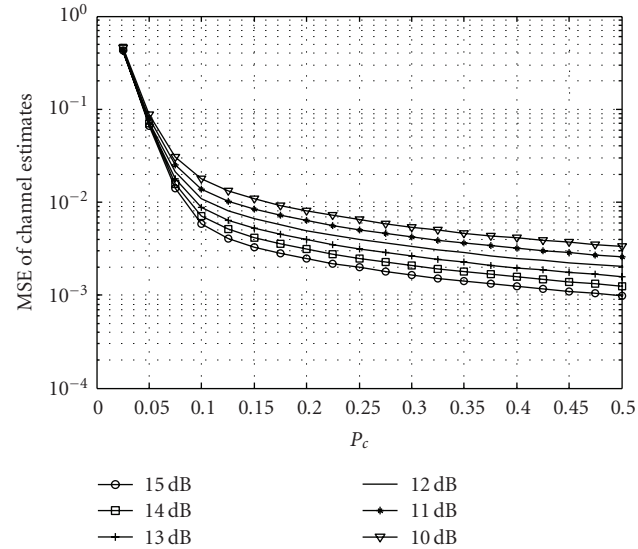


Figure 9: MSE of channel estimates versus power fraction for training signal.

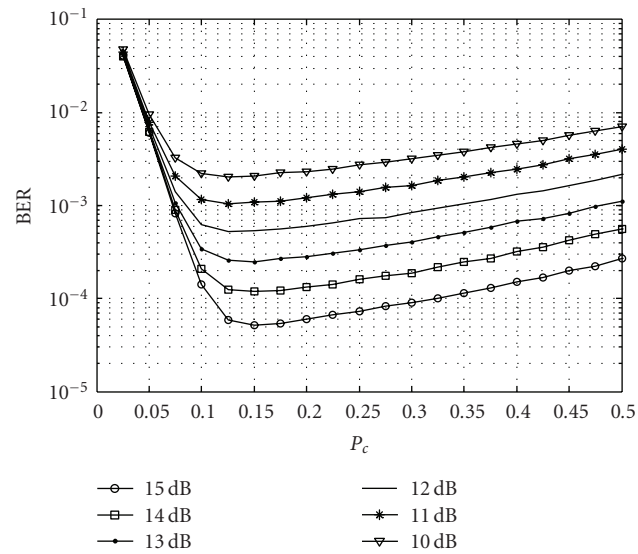


Figure 10: BER versus power fraction for training signal.

A QPSK symbol constellation is used, and the power transmitted for each antenna is normalized to one, that is, $P_x = 1$ [Watt].

Figure 7 depicts the channel estimate MSE comparisons for DDST and the MRST techniques with one and two iterations; both of them use $P_c = 17\%$ of the power transmitted $P_x$, that correspond to an approximate upper bound of the region with the best BER shown in Figure 10. Making an analysis of this performance, we can realize that the proposed scheme starts to approach the DDST performance from 10 [dB]. The MRST plot for the second iteration reaches the best approach. Note that this benefit comes at the price of decoding complexity.

Figure 8 shows the BER performance comparison between TDM with 17% bandwidth loss, DDST with DDD removal and the MRST algorithm. We observe that DDST and MRST plots with one or two iterations practically achieved the same performance. TDM-based channel estimation attained the best BER performance at the expenses of 17% bandwidth loss.

Figure 9 depicts the MSE of MRST versus $P_c$ when SNR = 15 dB. For example, with $P_c = 0.125$, we have that $\mathrm{MSE}(P_c) = 4[0.0316/(0.125*256)] = 0.004031$. A better MSE performance is directly proportional to the training power used.

Figure 10 plots the BER of MRST versus $P_c$ from SNR = 10 dB to 15 dB. Note that the minimum is achieved in the range where $0.1 < P_c \leq 0.17$ approximately. Because the BER achievement is the key and critical factor in digital communication systems, we worked in the region with the

best BER in the selection of the training power assigned. This selection is only a guide or suggestion for this type of radio propagation conditions at these particular SNR levels. Similar training power range was chosen for DDST in [9], but using a spatial multiplexing system. A more detailed analysis of the training power allocation problem is shown in [16]. Particularly, the simulations showed in Figures 9 and 10 were run until 10 000 blocks were in error.

## 6. CONCLUSION

low-complexity iterative superimposed training schemes that work jointly with equalization stage can offer similar or better performance than the iterative symbol-by-symbol detection DDST with DDD removal. The MRST introduced in this work effectively compensates the data dependent distortion that ST is unable to deal with, but at the receiver side. This leads to communication systems that perform similar to DDST but without suffering its drawbacks. Furthermore, MRST can be successfully applied to both SISO and MIMO systems, and in cooperation with the most widely used equalizers. Particularly, for the SISO case, MRST shows similar performance like the previous proposed method IST, but it avoids the inverse matrix computation $(\mathbf{C}+\hat{\tilde{\mathbf{B}}})^{-1}$ that every iteration IST does. Although the results of both iterative show insignificant gaps in the performance of BER and MSE, MRST is preferred by its less complexity hardware implementation.

For the MIMO case, the performance results are closely similar, that is, the performance of DDST is attained with the proposed method. Additionally, iterative ST methods can still be applied to time-varying channels, while DDST based system does not. This is true because the concept of cyclic mean in DDST is meaningless due to the fact that every symbol of the block transmitted is distorted in different way by the channel.

## ACKNOWLEDGMENTS

The authors thank the reviewers for their careful reading of the manuscript and their constructive comments. This work was supported by INTEL research grants DCIT2006 and CERMIMO2008, and CONACYT research grants 47909-Y and 84559-Y.

## REFERENCES

[1] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: general model, design criteria, and signal processing," *IEEE Signal Processing Magazine*, vol. 21, no. 6, pp. 12–25, 2004.

[2] B. Farhang-Boroujeny, "Pilot-based channel identification: proposal for semi-blind identification of communication channels," *Electronics Letters*, vol. 31, no. 13, pp. 1044–1046, 1995.

[3] A. G. Orozco-Lugo, M. M. Lara, and D. C. McLernon, "Channel estimation using implicit training," *IEEE Transactions on Signal Processing*, vol. 52, no. 1, pp. 240–254, 2004.

[4] M. Ghogho, D. C. McLernon, E. Alameda-Hernandez, and A. Swami, "Channel estimation and symbol detection for block transmission using data-dependent superimposed training," *IEEE Signal Processing Letters*, vol. 12, no. 3, pp. 226–229, 2005.

[5] T. Whitworth, M. Ghogho, and D. C. McLernon, "Data identifiability for data-dependent superimposed training," in *Proceedings of the IEEE International Conference on Communications (ICC '07)*, pp. 2545–2550, Glasgow, Scotland, June 2007.

[6] X. Meng and J. K. Tugnait, "Semi-blind channel estimation and detection using superimposed training," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04)*, vol. 4, pp. 417–420, Montreal, Canada, May 2004.

[7] S. M. Moosvi, D. C. McLernon, E. Alameda-Hernandez, A. G. Orozco-Lugo, and M. M. Lara, "A low complexity iterative channel estimation and equalisation scheme for (data dependent) superimposed training," in *Proceedings of the 14th European Signal Processing Conference (EUSIPCO '06)*, Florence, Italy, September 2006.

[8] M. Ghogho, D. C. McLernon, E. Alameda-Hernandez, and A. Swami, "SISO and MIMO channel estimation and symbol detection using data-dependent superimposed training," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, vol. 3, pp. 461–464, Philadelphia, Pa, USA, March 2005.

[9] M. Ghogho and A. Swami, "Channel estimation for MIMO systems using data-dependent superimposed training," in *Proceedings of the 42nd Annual Allerton Conference on Communication, Control and Computing*, Monticello, Ill, USA, September-October 2004.

[10] S. He, J. K. Tugnait, and X. Meng, "On superimposed training for MIMO channel estimation and symbol detection," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, part 2, pp. 3007–3021, 2007.

[11] E. Alameda-Hernandez, D. C. McLernon, A. G. Orozco-Lugo, M. M. Lara, and M. Ghogho, "Frame/training sequence synchronization and DC-offset removal for (data-dependent) superimposed training based channel estimation," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, part 1, pp. 2557–2569, 2007.

[12] A. B. Gershman and N. D. Sidiropoulos, *Space-Time Processing for MIMO Communications*, John Wiley & Sons, New York, NY, USA, 2005.

[13] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456–1467, 1999.

[14] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, vol. 1 of *Prentice Hall Signal Processing Series*, Prentice Hall, Upper Saddle River, NJ, USA, 1993.

[15] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 884–893, 2006.

[16] J. K. Tugnait and X. Meng, "On superimposed training for channel estimation: performance analysis, training power allocation, and frame synchronization," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 752–765, 2006.