

Error Sign Feedback as an Alternative to Pilots for the Tracking of FEXT Transfer Functions in Downstream VDSL

J. Louveaux and A.-J. van der Veen

Delft University of Technology, 2600AA Delft, The Netherlands

Received 1 December 2004; Revised 11 August 2005; Accepted 22 August 2005

With increasing bandwidths and decreasing loop lengths, crosstalk becomes the main impairment in VDSL systems. For downstream communication, crosstalk precompensation techniques have been designed to cope with this issue by using the collocation of the transmitters. These techniques naturally need an accurate estimation of the crosstalk channel impulse responses. We investigate the issue of tracking these channels. Due to the lack of coordination between the receivers, and because the amplitude levels of the remaining interference from crosstalk after precompensation are very low, blind estimation schemes are inefficient in this case. So some part of the upstream or downstream bit rate needs to be used to help the estimation. In this paper, we design a new algorithm to try to limit the bandwidth used for the estimation purpose by exploiting the collocation at the transmitter side. The principle is to use feedback from the receiver to the transmitter instead of using pilots in the downstream signal. It is justified by computing the Cramer-Rao lower bound on the estimation error variance and showing that, for the levels of power in consideration, and for a given bit rate used to help the estimation, this bound is effectively lower for the proposed scheme. A simple algorithm based on the maximum likelihood is proposed. Its performance is analyzed in detail and is compared to a classical scheme using pilot symbols. Finally, an improved but more complex version is proposed to approach the performance bound.

Copyright © 2006 Hindawi Publishing Corporation. All rights reserved.

1. INTRODUCTION

Future DSL systems such as VDSL (very high-bit-rate digital subscriber line) evolve towards shorter loops thanks to the increasing development of optical fiber infrastructure. This allows the use of higher bandwidths, typically from 10 to as high as 30 MHz for very short loops. At these high frequencies and low attenuation channels, the FEXT (far-end crosstalk) becomes the main degradation in the system, higher than additive noise. In order to overcome this issue, multiuser detectors can be designed [1] when the receivers are coordinated, that is, when the receivers have access to the signals coming from all the different lines. However, in typical downstream VDSL systems, the receivers will not be coordinated. For this reason, a number of precancellation techniques have been designed to decrease the effect of FEXT [2–4] using the coordination at the CO (central office) and assuming no coordination at the receiver side. These systems are quite different than in the MIMO wireless case because each receiver can only use the signal from its own line. So each receiver essentially sees a MISO channel. In addition, the physical characteristics of the VDSL channel ensure that the useful signal, which is the one transmitted on the line, is of much higher amplitude than the crosstalk. This also has to be taken into account in the design of the precanceller. For

more information on the precancellation design, see previous references or [5–7].

All these precancellation schemes rely on a good estimation of the crosstalk channels between the various pairs of users (or equivalently pairs of lines). So the issue of crosstalk channel estimation has to be solved to be able to use those schemes. In this paper, we investigate the issue of tracking of these channel estimates. Copper wires generally have static channel impulse responses, but they can still vary slowly, for example, due to temperature changes. So in order to guarantee a constant behavior of the crosstalk mitigation technique, some kind of tracking of the channel estimates is necessary. Due to the lack of coordination between the CPEs (customer premise equipments, i.e., the users' receivers), the downstream channel estimation appears to be a much more complicate task than the upstream channel estimation. So we focus on downstream in this paper. There are basically two characteristics of the system that make the downstream crosstalk channel estimation difficult. First, because of the non-coordination, each receiver can only use the signal from its own line to perform its estimation and has no information on the symbols transmitted to the other users. Furthermore, due to the presence of the crosstalk mitigation techniques, the power of the signal corresponding to the other users becomes very low at the receiver of one user. In other terms,

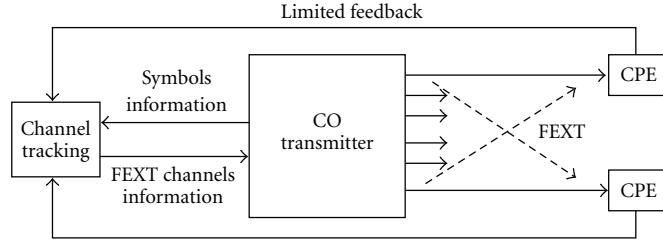


FIGURE 1: Principle of the proposed estimation structure.

the crosstalk impulse responses that need to be tracked are of very low amplitude with respect to the noise. So the downstream channel estimation appears as the joint estimation of multiple channels of very low amplitude corresponding to multiple independent sources (the different users' signal). This is a very difficult issue.

Blind techniques, such as the ones presented in [8, 9], are not practical in this context. They are useful for the estimation of the main transmission coefficient, that is, the direct transmission on the line itself. But concerning the crosstalk the low amplitude level with respect to the noise prevents from achieving reasonable performance. The easiest way to solve the problem would be to use a set of pilot symbols, sent periodically, to perform the tracking of the downstream channels at the CPEs. Many solutions exist in this framework [10, 11]. However, as the VDSL standards usually do not assume the use of preamble bits or periodically transmitted training sequences, it is necessary to use part of the useful bit rate as pilot symbols. In addition, the information about the estimates needs to be sent back to the CO periodically to perform an update of the crosstalk mitigating transmission scheme. So this may lead to a large amount of bit rate usage. In order to try to limit the quantity of bit rate needed for the tracking, we propose another method which takes advantage of the coordination that is present at the transmitter (CO).

The principle of the proposed algorithm (see Figure 1) is to send back to the CO some very limited amount of information about the signal received at the CPEs. Now thanks to the coordination at the CO, all symbols transmitted to all different lines are known, and that additional information can be used for the estimation. Furthermore, since the estimation is performed at the CO itself, feedback of the channel estimates is no longer needed. The algorithm is presented in this paper and it is compared through simulations to a simple solution using pilot symbols. It is shown that the proposed solution performs better for a given amount of bandwidth usage.

The issue of limiting the quantity of feedback for channel estimation has already been investigated in the MIMO wireless context in [12] and several other papers. However the problem considered here turns out to be very different. Indeed, in [12], the focus is on the feedback of the channel information to the transmitter. It is assumed that the estimation itself has been performed already. Here, the focus is on the estimation process and on limiting the total overhead

(both pilots and feedback) associated with the estimation process.

Note that we consider a DMT-based transmission and we focus on a simple algorithm that is working on a per tone basis. So we do not take into account the correlation between the tones, but it could be done in the same way as it is done with pilot schemes [10, 11], by performing the estimation on a limited number of tones and then interpolating between the estimated tones using the correlation across frequencies. Besides, we do not make use of the samples available in the cyclic extension [13].

The paper is organized as follows. First, the system model and the issue investigated are described. In Section 3, the proposed algorithm is derived. In Section 4, the Cramer-Rao bound for the proposed structure is investigated and compared to the use of pilot symbols, in order to show that the proposed scheme is indeed potentially superior. In Section 5, the performance of the proposed scheme is analyzed both theoretically and with simulations. Finally, an improved, but more complex, algorithm is proposed in Section 6. The basic algorithm has already been presented in [14] and a few simulations results have been shown. In this paper, we additionally provide a theoretical justification based on the Cramer-Rao bound, we provide a more detailed analysis of the performance both analytically and with extensive simulations. Finally, we also show how the algorithm can be improved to approach the performance bound.

2. SYSTEM MODEL

We consider the estimation of the downstream crosstalk channels in a DSL environment. DMT modulation is assumed. It is also assumed that the cyclic prefix is long enough and the different users are transmitted synchronously from the CO so that the channel (including crosstalk) is free of intersymbol interference and intercarrier interference. Hence, for a given tone, the channel model is written as

$$\mathbf{y}' = \mathbf{Cx} + \mathbf{n}, \quad (1)$$

where \mathbf{x} , \mathbf{y}' are the vectors of transmitted and received samples, respectively,¹ for the different users (or equivalently, on

¹ The notation \mathbf{y}' is used here because the actual observations used will be a slightly modified version of this (see later).

the different lines), \mathbf{C} is the channel matrix, and \mathbf{n} is the vector of noise samples at the different receivers (CPEs). In this paper, we focus on one fixed tone. The same developments can be done independently for each tone (or a subset of the tones if the frequency-domain correlation is used). The additive noise is assumed to be Gaussian and white with independent elements. The noise variance for user (receiver) i is denoted by $\sigma_{n,i}^2$. In the model (1), the diagonal elements of \mathbf{C} correspond to the line transmission (also called direct channel later in this paper), the off-diagonal elements correspond to crosstalk. We assume N users, the channel matrix \mathbf{C} is thus $N \times N$. It must be noted that the channel model considered here is supposed to take into account all the operations from the DMT modulation, through the channel, and until the input of the decision device. This thus includes the channel shortening, the cyclic extension operations, possible equalization and may even incorporate, for instance, some alien crosstalk suppression schemes at the receiver. The precoder (or precanceller) can be viewed as an additional layer working on top of all these operations.

2.1. Precoder

Because the receivers (CPEs) are not collocated, each one of them can only use one received signal y_i for detection and/or estimation purposes. In order to mitigate the effect of FEXT, it is assumed that the CO uses some kind of precoder. We assume a linear precoder as presented in [2] and later improved in [4].

The CO designs a matrix \mathbf{F} such that \mathbf{CF} is diagonal,

$$\mathbf{F} = \mathbf{C}^{-1} \mathbf{C}_d, \quad (2)$$

where \mathbf{C}_d represents the diagonal matrix formed by keeping only the diagonal elements of \mathbf{C} , and sends

$$\mathbf{x} = \mathbf{Fu} \quad (3)$$

on the different lines, where \mathbf{u} are the transmitted information symbols for the different users. Thanks to the precoder design, the received samples for one user suffer from little interference from other users. Regarding the transmitted symbols, it is assumed that all the users have the same transmitted power, and we therefore normalize the symbol variance to $\sigma_u^2 = 1$ for all users without loss of generality. The sizes of the user constellations are different however. They are adapted to the SNR (signal-to-noise ratio) available on the given tone by the various users, in such a way that the bit error rate is maintained below 10^{-7} for each user. In order to simplify the notations, the symbols are assumed to be real throughout the paper, but the extension to complex symbols is straightforward.

2.2. Initialization procedure and tracking issue

In this paper, we focus on the issue of tracking the crosstalk channel coefficients. Hence it is assumed that some initial estimate of the crosstalk channel has been obtained during the initialization phase. Here is a little description of a possible

way of handling this initialization. First, the DMT initialization is performed. Then transmission can start at a lower rate, without any crosstalk cancellation, considering crosstalk as noise. During this first part, some coarse estimation of the crosstalk channel can be performed, for instance using pilot symbols. The method proposed here would also be able to perform this coarse estimation. However, for reasons explained later, it might not be as efficient *in the initialization phase*. The precoder can then be computed and transmission can start at the highest rate. Then, the channel is changing slowly, for example due to changes in temperature, or possibly due to changes in the alien crosstalk environment if such a cancellation scheme is used. Equivalently, the initial estimate might just be inaccurate. Therefore, the precoder might not diagonalize the channel perfectly and the remaining interference due to crosstalk might increase around the same power level as the additive noise, thereby decreasing the performance. Mathematically, this means that the matrix \mathbf{CF} in the received signal expression

$$\mathbf{y}' = \mathbf{Cx} + \mathbf{n} = \mathbf{CFu} + \mathbf{n} \quad (4)$$

is not perfectly diagonal. In order to update the precoder and recover a low level of interference, some estimation (or tracking) of the nondiagonal elements of this matrix is necessary. In the remainder of this paper, we call these values the *interference coefficients*. They correspond to the interference between lines that remains due to a mismatch between the precoder and the actual channel and are thus generally of low amplitude. We will refer to *channel coefficients* to denote the crosstalk coefficients of the channel (matrix \mathbf{C}) before the precoder is applied.

3. PROPOSED ALGORITHM

3.1. Algorithm derivation

In this section, the proposed estimation algorithm is derived in detail. The solution (Figure 1) investigated here is to allow a limited feedback from the various users about their received samples. This information is collected at the CO and the channel estimation is performed there. It is important to limit drastically the information that is sent back in order to keep an acceptable usage of the upstream bit rate. Even with a limited amount of feedback, and since the CO knows perfectly what was sent on the different lines (the samples \mathbf{x} and the symbols \mathbf{u}), the channel estimation is possible.

It is first assumed that the direct channel coefficients (diagonal ones) are estimated perfectly at the receivers (this can be done easily with a decision-directed scheme since the power of the useful signal is high). After detection, the contribution of the corresponding user's symbol is subtracted at the receiver, only remaining with the crosstalk interference and the noise. We call this quantity (crosstalk + noise) the *symbol error*. The receivers send back the *sign* of this symbol error, so that the smallest possible amount of the upstream bit rate is used: 1 bit. We focus on real-valued symbols here. The extension to complex symbols can easily be

done by splitting the complex values in real and imaginary parts, feeding back the sign of both quantities.

Mathematically, K DMT blocks are stacked up (still focusing on one tone only) in the following way:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}^0 & \cdots & \mathbf{x}^{K-1} \end{bmatrix}, \quad (5)$$

where \mathbf{x}^k denotes the vector of transmitted samples for block k . The matrices \mathbf{U} , \mathbf{Y}' , and \mathbf{N} are built similarly. K is the number of observations used by the algorithm. Since VDSL channels are varying slowly, this number can be quite large in practice. The channel model and precoding operations are rewritten as

$$\begin{aligned} \mathbf{Y}' &= \mathbf{C}\mathbf{X} + \mathbf{N}, \\ \mathbf{X} &= \mathbf{F}\mathbf{U}. \end{aligned} \quad (6)$$

At the receivers, the diagonal elements of $\mathbf{C}\mathbf{F}$ are assumed to be estimated perfectly, and the symbols transmitted to the corresponding users are also assumed to be detected perfectly. Their contribution is then subtracted to obtain the so-called symbol errors

$$\mathbf{Y} = \mathbf{Y}' - \{\mathbf{C}\mathbf{F}\}_d\mathbf{U} \quad (7)$$

$$= (\mathbf{C}\mathbf{F} - \{\mathbf{C}\mathbf{F}\}_d)\mathbf{U} + \mathbf{N} \quad (8)$$

$$= \mathbf{H}\mathbf{U} + \mathbf{N}, \quad (9)$$

where the last line defines a new matrix \mathbf{H} with zeros on the diagonal. We call it the interference matrix. It represents the residual interference at the output of the receiver in presence of the precoding scheme, and its elements are thus of low amplitude. The nondiagonal elements are the so-called interference coefficients. This is the matrix that will be estimated at the CO by the algorithm.

The algorithm is based on the ML (maximum likelihood) principle. We denote by $\mathbf{Z} = \text{sign}(\mathbf{Y})$, the set of received signs of the symbol errors coming from the different lines. They are the observations on which the estimation will be based. The error sign sample received from user i for block k is denoted by z_i^k (similarly for y_i^k , u_i^k , and n_i^k). It is assumed that the noise variance of each receiver is known at the CO. This will be necessary in the computation of the algorithm as shown later. The noise variance at receiver i is denoted by $\sigma_{n,i}^2$. The likelihood of a set of interference coefficients can be written as

$$\Lambda(\mathbf{H}) = \prod_{k=0}^{K-1} \prod_{i=0}^{N-1} P(\text{sign}(y_i^k) = z_i^k | \mathbf{H}, \mathbf{U}), \quad (10)$$

where $P(\text{sign}(y_i^k) = z_i^k | \mathbf{H}, \mathbf{U})$ denotes the conditional probability on the value of some error sign sample, given the transmitted symbols and given the set of interference coefficients. Note that the estimation can be performed independently for each line as the interference coefficients related to one line only impact the received samples from the corresponding line. However, for generality, the matrix formalism is kept here. For one specific error sign sample, the

probability is

$$\begin{aligned} P(\text{sign}(y_i^k) = z_i^k | \mathbf{H}, \mathbf{U}) &= Q\left(-z_i^k \frac{\mathbf{h}_i \mathbf{u}^k}{\sqrt{\sigma_{n,i}^2}}\right), \\ \Lambda(\mathbf{H}) &= \prod_{k=0}^{K-1} \prod_{i=0}^{N-1} Q\left(\frac{-z_i^k \mathbf{h}_i \mathbf{u}^k}{\sigma_{n,i}}\right), \end{aligned} \quad (11)$$

where \mathbf{h}_i is the i th row of \mathbf{H} , \mathbf{u}^k is the k th column of \mathbf{U} , and where

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_v^\infty e^{-t^2/2} dt. \quad (12)$$

The tracking algorithm is obtained by taking the derivative of the likelihood function, and performing a simplified steepest descent procedure. The gradient of the likelihood function is given by

$$\frac{\partial \Lambda(\mathbf{H})}{\partial \mathbf{h}_i} = \frac{\Lambda(\mathbf{H})}{\sqrt{2\pi\sigma_{n,i}^2}} \sum_{k=0}^{K-1} z_i^k (\mathbf{u}^k)^T \frac{e^{-(\mathbf{h}_i \mathbf{u}^k)^2/2\sigma_{n,i}^2}}{Q(-z_i^k \mathbf{h}_i \mathbf{u}^k / \sigma_{n,i})}. \quad (13)$$

The proposed basic tracking algorithm computes the corresponding term of the gradient for each new received sample (each block k) and adapts the coefficients estimates in the direction of the gradient. In other words, it realizes the sum over k in (13) by adapting progressively for each new coming sample (except that the interference coefficient estimates $\hat{\mathbf{h}}_i$ are changing slowly). It is important to keep the weightings that depend on the sample k (i.e., the big fraction) because it contains the information on the relative importance of each term of the gradient. The common factor can be removed of course, and incorporated in the stepsize. Finally, the following algorithm is provided:

$$\hat{\mathbf{h}}_i^{k+1} = \hat{\mathbf{h}}_i^k + \mu z_i^k D \left(\frac{-z_i^k \hat{\mathbf{h}}_i^k \mathbf{u}^k}{\sqrt{\sigma_{n,i}^2}} \right) \cdot (\mathbf{u}^k)^T, \quad (14)$$

where $\hat{\mathbf{h}}_i^k$ denotes the current estimate at block k of row i of the interference matrix \mathbf{H} , μ is the stepsize which can be chosen to tune the properties of the algorithm, and where

$$D(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}Q(x)}. \quad (15)$$

The tracking algorithm (14) appears to be similar to an LMS algorithm, or more precisely to the sign LMS [15]. However it is very different because, in the sign-LMS algorithm, the sign operation is taken on the “prediction error” computed between the observation y_i^k and the predicted version $\hat{\mathbf{h}}_i^k \mathbf{u}^k$, based on the estimation. In our case, the sign is directly applied on the symbol error y_i^k and the “prediction error” is not available. As can be seen in (14), it is replaced here by some more complicated expression. Consequently, the behavior and performance of this algorithm can be expected to be very different.

Finally, the ultimate goal is to adapt the precoder to the changes in the channel. To achieve this, the diagonal coefficients of the matrix $\mathbf{C}\mathbf{F}$ (direct channel coefficients), which

are easy to estimate at the CPEs, have to be sent back periodically as well. This allows the CO to reconstruct CF and hence C, and then to compute the new precoder with (2).

3.2. Comparison with pilot symbols

In order to verify the behavior of the proposed algorithm, it will be compared to an estimation method based on pilot symbols. We assume the use of an LMS algorithm at each receiver, using the different pilots to estimate the interference coefficients. Hence it is also an iterative algorithm but it is performed at the receivers instead of the CO. The adaptation can be written as

$$\hat{\mathbf{h}}_i^{k+1} = \hat{\mathbf{h}}_i^k + \mu_{\text{LMS}} (y_i^k - \hat{\mathbf{h}}_i^k \mathbf{u}^k) (\mathbf{u}^k)^T. \quad (16)$$

The symbols \mathbf{u}^k are the pilots. Now, the purpose of the comparison is to evaluate which algorithm consumes the smallest amount of bit rate for the estimation purpose, or equivalently, which has the best performance for a given bit rate usage. Hence the bit rate usage of the two different methods is computed in this section. The proposed algorithm uses one bit of the upstream for each feedback of a symbol error. So, for K transmitted symbols and N users, the bit rate usage of the proposed method for the estimation of all the coefficients is KN bits. The LMS solution using pilots consumes the downstream bit rate of the pilots, as well as some additional upstream bit rate needed to feedback the value of the estimated channel coefficients. Here we neglect this feedback, but this is of course an additional overhead with respect to the proposed method. The downstream bit rate used by the pilots actually depends on the constellation size of the symbols they replace, and thus on the SNR of the corresponding tone for the different users. If we denote by b_i the number of bits that could be transmitted on the tone of interest for user i , and by K_{LMS} the number of pilot symbols transmitted, the total amount of downstream bit rate used by the pilots is $K_{\text{LMS}} \sum_{i=0}^{N-1} b_i$. It is assumed that the consumed bit rates on upstream and on downstream are treated equally. Then, a fair comparison between the two methods can be done when the same number of bits is consumed in both cases (for one precoder update), that is, when $KN = K_{\text{LMS}} \sum_{i=0}^{N-1} b_i$. So in practice, the number of symbols K will be higher in the proposed method (constellation sizes can go up to 1024 depending on the available SNR on the corresponding tone). The actual bit rate usage for the estimation purpose is of course dependent on the update rate of the crosstalk model, which will be the same for the two methods and has no further influence on the performance.

As an additional comment, it can be pointed out that a system where the symbol errors y_i^k are fed back in full precision to the transmitter would actually have access to the same information as the system using pilot symbols (except that the information is available at the transmission side instead of the receiving side). Such a system would therefore be able to provide equal performance than estimation methods based on pilots. However, the feedback in full precision is much more demanding in terms of consumed bit rate than

the same amount of pilots (unless the constellation sizes are very high) and such a system is thus not worthwhile in practice.

4. CRAMER-RAO BOUND

In this section, the CRB (Cramer-Rao lower bound) associated with the proposed estimation structure (i.e., using the sign feedback) is investigated. Then, it is compared to the CRB of the estimation performed using pilot symbols. The objective is to show that, for a fixed number of bits used (as either feedback or pilot symbols), and in presence of high noise, the proposed structure has a higher potential than the pilot-based estimation (the CRB is lower). This thus provides a theoretical justification for the proposed approach.

Regarding the CRB computation, the two basic differences between the two schemes are the following.

- (i) The “sign” scheme only uses the sign of the observation while the “pilot” scheme can use the full observations y to make the estimations.
- (ii) In counterpart, the “pilot” scheme needs to transmit pilots instead of full symbols, corresponding to multiple bits, while the “sign” scheme only uses one bit per symbol in feedback (see previous section).

As in our case the observation interval is long,² the so-called modified CRB (MCRB) can be used [16] and provides a very good approximation to the true CRB. For the estimation of some set of parameters Θ , using observations \mathbf{y} and with a set of nuisance parameters \mathbf{U} , the modified Cramer-Rao lower bound on the variance of any unbiased estimator for one parameter θ_m is given by

$$\sigma_{\theta_m}^2 \geq -\frac{1}{\mathcal{E}_{\mathbf{U}}[\mathcal{E}_n[\partial^2 \ln P(\mathbf{y} | \Theta, \mathbf{U}) / \partial \theta_m^2]]}, \quad (17)$$

where $\mathcal{E}_n[\cdot]$ denotes the expectation with respect to the noise. This is a lower bound looser than the true CRB. But when the number of observations is very large as it is the case here, it gets tight thanks to the fact that the Fisher information matrix is almost diagonal and tightly distributed.

4.1. Modified CRB for pilot symbols

We first compute the MCRB for the simple pilot scheme. This corresponds to a classical DA (data aided) scheme. The model (9) applies, but we focus on one row of \mathbf{H} only:

$$\mathbf{y}_i = \mathbf{h}_i \mathbf{U} + \mathbf{n}_i, \quad (18)$$

where \mathbf{y}_i denotes the row vector obtained from \mathbf{Y} by taking only the received samples for user i . Assuming the noise is

² This is required due to the high level of noise with respect to the interference coefficients to estimate, and this is possible since the channel variations are slow.

white and Gaussian, it follows that

$$\begin{aligned} P(\mathbf{y}_i \mid \mathbf{h}_i, \mathbf{U}) &= \prod_{k=0}^{K_{\text{LMS}}-1} \frac{1}{\sqrt{2\pi\sigma_{n,i}^2}} e^{-(y_i^k - \mathbf{h}_i \mathbf{u}^k)^2 / 2\sigma_{n,i}^2}, \\ \frac{\partial^2 \ln P(\mathbf{y}_i \mid \mathbf{h}_i, \mathbf{U})}{\partial h_{i,m}^2} &= -\frac{1}{\sigma_{n,i}^2} \sum_{k=0}^{K_{\text{LMS}}-1} (u_m^k)^2. \end{aligned} \quad (19)$$

Finally, the lower bound is obtained as

$$\sigma_{h_{i,m}}^2 \geq \frac{\sigma_{n,i}^2}{K_{\text{LMS}}\sigma_u^2} \triangleq \sigma_{h_{i,m},\text{min,pilot}}^2, \quad (20)$$

where $h_{i,m}$ denotes the element of \mathbf{H} on the i th row and the m th column, and where σ_u^2 denotes the symbol variance, normalized to $\sigma_u^2 = 1$ in this paper.

4.2. Modified CRB for the proposed scheme

For the proposed scheme, the channel model is again given by (9) and the observations used at the CO for the estimation are $\mathbf{Z} = \text{sign}(\mathbf{Y})$. We focus on one row of the interference matrix (i.e., on one user i only). For simplification of the equations, we define the normalized interference coefficients for row i as

$$\bar{\mathbf{h}}_i \triangleq \frac{\mathbf{h}_i}{\sqrt{\sigma_{n,i}^2}}. \quad (21)$$

Note that they are just used for notation, we are of course still interested in the variance on the estimation of the true interference coefficients.

The probability distribution of the observations is written as

$$\ln P(\mathbf{z}_i \mid \mathbf{h}_i, \mathbf{U}) = \sum_{k=0}^{K-1} \ln Q(-z_i^k \bar{\mathbf{h}}_i \mathbf{u}^k). \quad (22)$$

Then

$$\begin{aligned} \frac{\partial^2 \ln P(\mathbf{z}_i \mid \mathbf{h}_i, \mathbf{U})}{\partial h_{i,m}^2} &= \sum_{k=0}^{K-1} -\frac{(u_m^k)^2}{\sigma_{n,i}^2} D(-z_i^k \bar{\mathbf{h}}_i \mathbf{u}^k) \cdot \{z_i^k \bar{\mathbf{h}}_i \mathbf{u}^k + D(-z_i^k \bar{\mathbf{h}}_i \mathbf{u}^k)\}, \\ \mathcal{E}_n \left[\frac{\partial^2 \ln P(\mathbf{z}_i \mid \mathbf{h}_i, \mathbf{U})}{\partial h_{i,m}^2} \right] &= \sum_{k=0}^{K-1} -\frac{(u_m^k)^2}{\sigma_{n,i}^2} \bar{\mathbf{h}}_i \mathbf{u}^k \frac{e^{-(\bar{\mathbf{h}}_i \mathbf{u}^k)^2/2}}{\sqrt{2\pi}} \mathcal{E}_n \left[\frac{z_i^k}{Q(-z_i^k \bar{\mathbf{h}}_i \mathbf{u}^k)} \right] \\ &\quad + \sum_{k=0}^{K-1} -\frac{(u_m^k)^2}{\sigma_{n,i}^2} \left(\frac{e^{-(\bar{\mathbf{h}}_i \mathbf{u}^k)^2/2}}{\sqrt{2\pi}} \right)^2 \mathcal{E}_n \left[\frac{1}{Q^2(-z_i^k \bar{\mathbf{h}}_i \mathbf{u}^k)} \right]. \end{aligned} \quad (23)$$

Computing the expectations

$$\begin{aligned} \mathcal{E}_n \left[\frac{z_i^k}{Q(-z_i^k \bar{\mathbf{h}}_i \mathbf{u}^k)} \right] &= \frac{P(z_i^k = 1 \mid \mathbf{h}_i, \mathbf{u}^k)}{Q(-\bar{\mathbf{h}}_i \mathbf{u}^k)} + (-1) \frac{P(z_i^k = -1 \mid \mathbf{h}_i, \mathbf{u}^k)}{Q(\bar{\mathbf{h}}_i \mathbf{u}^k)} \\ &= 1 - 1 = 0, \end{aligned} \quad (24)$$

$$\mathcal{E}_n \left[\frac{1}{Q^2(-z_i^k \bar{\mathbf{h}}_i \mathbf{u}^k)} \right] = \frac{1}{Q(\bar{\mathbf{h}}_i \mathbf{u}^k)} + \frac{1}{Q(-\bar{\mathbf{h}}_i \mathbf{u}^k)},$$

it becomes

$$\begin{aligned} -\mathcal{E}_n \left[\frac{\partial^2 \ln P(\mathbf{z}_i \mid \mathbf{h}_i, \mathbf{U})}{\partial h_{i,m}^2} \right] &= \sum_{k=0}^{K-1} \frac{(u_m^k)^2}{\sigma_{n,i}^2} \left[\frac{e^{-(\bar{\mathbf{h}}_i \mathbf{u}^k)^2/2}}{\sqrt{2\pi}} \right]^2 \left[\frac{1}{Q(\bar{\mathbf{h}}_i \mathbf{u}^k)} + \frac{1}{Q(-\bar{\mathbf{h}}_i \mathbf{u}^k)} \right]. \end{aligned} \quad (25)$$

The modified CRB is thus

$$\begin{aligned} \sigma_{h_{i,m},\text{min,sign}}^2 &= \frac{\sigma_{n,i}^2}{K \mathcal{E}_{\mathbf{u}} [u_m^2 (e^{-(\bar{\mathbf{h}}_i \mathbf{u})^2/2} / \sqrt{2\pi}) \{D(\bar{\mathbf{h}}_i \mathbf{u}) + D(-\bar{\mathbf{h}}_i \mathbf{u})\}]} \end{aligned} \quad (26)$$

where \mathbf{u} is a random vector of transmitted symbols for one block. The expectation in (26) is not tractable analytically so it is computed numerically. It must be noted that it is clearly dependent on the various parameters: the constellation sizes of the different users, the interference coefficients themselves, and of course the noise variance. Now, another interesting value to compute is the gain (or loss) of our method with respect to the use of pilot symbols. It can be done by computing the ratio between the two CRBs. Since the symbol variance can be assumed equal to 1 without loss of generality, it follows that

$$\begin{aligned} G_{i,m} &\triangleq \frac{\sigma_{h_{i,m},\text{min,pilot}}^2}{\sigma_{h_{i,m},\text{min,sign}}^2} \\ &= \mathcal{E}_{\mathbf{u}} \left[u_m^2 \left(\frac{e^{-(\bar{\mathbf{h}}_i \mathbf{u})^2/2}}{\sqrt{2\pi}} \right) \{D(\bar{\mathbf{h}}_i \mathbf{u}) + D(-\bar{\mathbf{h}}_i \mathbf{u})\} \right]. \end{aligned} \quad (27)$$

This represents the “gain” of the proposed method (using sign feedback) with respect to the use of pilot symbols for the estimation of interference coefficient $h_{i,m}$ for an identical number of symbols sent (i.e., for fixed $K = K_{\text{LMS}}$). The gain is dependent on the interference coefficients and may be different for all coefficients $h_{i,m}$. As defined here, the gain should be always smaller than 1 since the pilot scheme has always more information available. However, as mentioned earlier a fair comparison should be done for an identical number of bits used. In that case, the gain becomes

$$G_{i,m, \text{fixed \# bits}} = G_{i,m} \frac{\sum_{i=0}^{N-1} b_i}{N}, \quad (28)$$

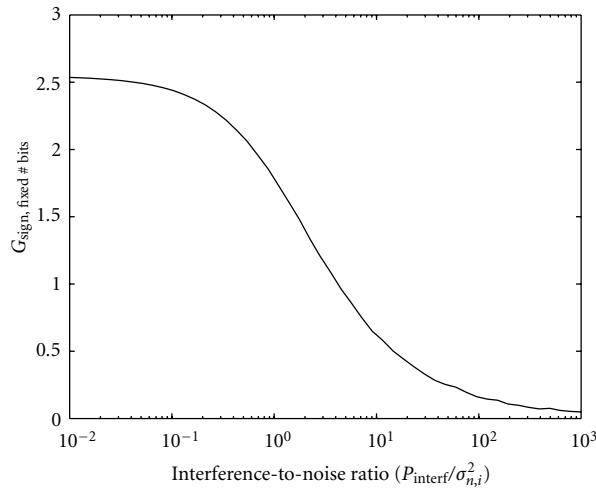


FIGURE 2: Average gain of the sign method as a function of the ratio between the power of the interference coefficients to estimate P_{interf} and the noise variance. The constellation sizes are 16.

where b_i denotes the number of bits transmitted per symbol for user i . So if $G_{i,m}$ is not too small, the gain (28) can become much larger than 1. It can be observed that this gain is only dependent on the constellation sizes of the different users and on the *normalized* interference coefficients.

4.3. Comparison

The gain (28) and the MCRB are evaluated in this section. Both are however dependent on the true interference coefficients (the vector \mathbf{h}_i). So, in order to get some valuable result, the MCRB and the gain are averaged over several realizations of the channel with a fixed interference power. Mathematically, it is assumed that the interference coefficients $h_{i,m}$ are Gaussian distributed, but are then proportionally corrected to satisfy $\sum_m h_{i,m}^2 = P_{\text{interf}}$ for some constant power of interference P_{interf} . Figure 2 shows the average gain (28) of the proposed (sign) scheme as a function of ratio between the interference level (P_{interf}) and the noise variance $\sigma_{n,i}^2$, and for constellation sizes of 16. Each result is averaged over 3000 realizations³ of the channel as described above. It can be seen that the gain is always decreasing for increasing interference coefficients (or decreasing noise variance). It can also be seen that the gain is indeed higher than 1 for reasonable cases: it does not seem reasonable to allow the interference, which is due to changes in the channel, to go significantly above the noise as it would unacceptably decrease the performance. So this shows that *for a given bit rate usage, a well-*

designed estimator is likely to perform better in the proposed scheme than with pilot symbols. This confirms the results obtained previously. The figure also shows that the interest of the proposed structure is limited to situations were the interference⁴ is about the same level as the noise or lower. For high interference-to-noise ratio, the traditional pilot schemes are likely to perform better.

For illustration, Figure 3 shows the MCRB (variance) of the proposed (sign) scheme as a function of the noise variance for a given set of interference coefficients.

5. EVALUATION OF PERFORMANCE

5.1. Relation between estimation variance and transmission performance

One drawback of the Cramer-Rao bound is that it provides a performance evaluation of the channel estimation in terms of *error variance*. But, in practice, the purpose of our estimation is to be able to compute a refined precoder and finally get better SNIRs for transmission on the different lines. So, in this section, we show how to relate the estimation performance, in terms of variance, to the achievable SNIRs on the different lines after the refined precoding. This is done using a few assumptions, and it is later shown by simulations that the obtained relation is closely followed.

The precoder may be written as

$$\mathbf{F} = \hat{\mathbf{C}}^{-1} \hat{\mathbf{C}}_d, \quad (29)$$

where $\hat{\mathbf{C}}$ is the estimation of the channel matrix \mathbf{C} available at the transmitter. We write

$$\hat{\mathbf{C}} = \mathbf{C} + \mathbf{E}\mathbf{F}_{\text{old}}^{-1}, \quad (30)$$

where \mathbf{E} is the estimation error matrix on the interference matrix \mathbf{H} , and \mathbf{F}_{old} is the old precoder, needed to compute the estimate of the channel matrix \mathbf{C} from the estimate of the interference matrix. It is assumed that the error matrix \mathbf{E} is a zero mean Gaussian random matrix with i.i.d. elements having variance σ_e^2 . Although the proposed estimation scheme may result in correlation between the errors, it is reasonable to assume that, using a large number of samples, this correlation may vanish. The estimation error variances may also *not* be the same for the different interference coefficients, but in practice, it appears that the differences are not large, so this approximation is acceptable. This is confirmed by the simulation results and partly by the performance analysis in the next section. The inverse of $\hat{\mathbf{C}}$ is approximated as

$$\hat{\mathbf{C}}^{-1} \approx \mathbf{C}^{-1} - \mathbf{C}^{-1} \mathbf{E} \mathbf{F}_{\text{old}}^{-1} \mathbf{C}^{-1}. \quad (31)$$

So the vector of received samples is

$$\mathbf{y}' = \mathbf{C}\mathbf{F}\mathbf{u} + \mathbf{n} = \hat{\mathbf{C}}_d \mathbf{u} - \mathbf{E} \mathbf{F}_{\text{old}}^{-1} \mathbf{C}^{-1} \hat{\mathbf{C}}_d \mathbf{u} + \mathbf{n} \quad (32)$$

³ Note that because the MCRB is inversely proportional to the gain (28), we actually compute the inverse of the average of the inverse of the gains—that is, the so-called harmonic mean. It provides a slightly lower value than a direct average of the gain. Also note that, for a given ratio, the gains corresponding to the various channel realizations usually differ only by 1-2 dB from the mean value.

⁴ Or, in a more general context, the power of the signal for which the channel needs to be estimated.

and the vector of symbol estimates at the receivers is

$$\hat{\mathbf{u}} = \hat{\mathbf{C}}_d^{-1} \mathbf{y}' = (\mathbf{I} - \hat{\mathbf{C}}_d^{-1} \mathbf{E} \mathbf{F}_{\text{old}}^{-1} \mathbf{C}^{-1} \hat{\mathbf{C}}_d) \mathbf{u} + \hat{\mathbf{C}}_d^{-1} \mathbf{n}. \quad (33)$$

There is an additional ISI term

$$\mathbf{u}_{\text{ISI}} = \hat{\mathbf{C}}_d^{-1} \mathbf{E} \mathbf{F}_{\text{old}}^{-1} \mathbf{C}^{-1} \hat{\mathbf{C}}_d \mathbf{u}. \quad (34)$$

Thanks to the independence of the estimation errors on the different interference coefficients, it can be shown that the ISI covariance matrix $\mathbf{R}_{\text{ISI}} = \mathcal{E}[\mathbf{u}_{\text{ISI}} \mathbf{u}_{\text{ISI}}^T]$ is diagonal (i.e., the ISI terms are not correlated). Indeed, using the i.i.d assumption on the elements of \mathbf{E} , it can easily be shown that, for any matrix \mathbf{A} ,

$$\mathcal{E}[\mathbf{E} \mathbf{A} \mathbf{E}^T] = \sigma_e^2 \text{Tr}\{\mathbf{A}\} \mathbf{I}, \quad (35)$$

where \mathbf{I} is the identity matrix. Since the symbols from the different users are also assumed independent, with fixed symbol variance σ_u^2 , the covariance matrix of the ISI is

$$\mathbf{R}_{\text{ISI}} = \sigma_u^2 \sigma_e^2 \text{Tr}\{\mathbf{F}_{\text{old}}^{-1} \mathbf{C}^{-1} \hat{\mathbf{C}}_d \hat{\mathbf{C}}_d^T \mathbf{C}^{-T} \mathbf{F}_{\text{old}}^{-T}\} \hat{\mathbf{C}}_d^{-1} \hat{\mathbf{C}}_d^{-T}. \quad (36)$$

It is a diagonal matrix. Now, in order to compute (36), the estimations are replaced by the true value, and furthermore, due to the diagonal dominance of the channel matrix \mathbf{C} , the trace in (36) is well approximated by N . So, finally,

$$\mathbf{R}_{\text{ISI}} \approx N \sigma_u^2 \sigma_e^2 \mathbf{C}_d^{-1} \mathbf{C}_d^{-1, T}. \quad (37)$$

This provides the power of interference present after the update of the precoding on the different lines when the interference coefficients (before the update) are estimated with a variance σ_e^2 . The value of the power provided by (37) is normalized for a useful signal of power σ_u^2 . It can thus be directly translated in terms of SIR or SNIR.

5.2. Steady-state performance analysis

In this section, we investigate the performance of the algorithm itself. Thanks to the relation given in the previous section, it is now sufficient to investigate the performance of the proposed adaptive algorithm in terms of the error variance σ_e^2 . The steady-state error variance is computed in this section, using a method similar to [15]. Let us consider only one line here, so the subscript i (user index) is temporarily dropped for legibility. First, the following definition of the estimation error vector is used

$$\tilde{\mathbf{h}}^k = \hat{\mathbf{h}}^k - \mathbf{h}. \quad (38)$$

The adaptation rule (14) is obviously unchanged when it is written for $\tilde{\mathbf{h}}^k$ instead of $\hat{\mathbf{h}}^k$. The square norm of the adaptation rule (in $\tilde{\mathbf{h}}^k$) is written:

$$\begin{aligned} |\tilde{\mathbf{h}}^{k+1}|^2 &= |\tilde{\mathbf{h}}^k|^2 + 2\mu z^k D\left(-\frac{z^k \hat{\mathbf{h}}^k \mathbf{u}^k}{\sigma_n}\right) \tilde{\mathbf{h}}^k \mathbf{u}^k \\ &\quad + \mu^2 D^2\left(-\frac{z^k \hat{\mathbf{h}}^k \mathbf{u}^k}{\sigma_n}\right) |\mathbf{u}^k|^2. \end{aligned} \quad (39)$$

Then, the expectation is taken. In steady state, it is assumed that $\mathcal{E}[|\tilde{\mathbf{h}}^k|^2] = \mathcal{E}[|\tilde{\mathbf{h}}^{k+1}|^2]$, so it follows that

$$\begin{aligned} \mathcal{E}\left[z^k D\left(-\frac{z^k \hat{\mathbf{h}}^k \mathbf{u}^k}{\sigma_n}\right) \tilde{\mathbf{h}}^k \mathbf{u}^k\right] \\ = -\frac{\mu}{2} \mathcal{E}\left[D^2\left(-\frac{z^k \hat{\mathbf{h}}^k \mathbf{u}^k}{\sigma_n}\right) |\mathbf{u}^k|^2\right]. \end{aligned} \quad (40)$$

This expectation is taken over all noise samples and all symbols. Clearly, $\hat{\mathbf{h}}^k$ is influenced by all past noise samples and past symbols. But only z^k is dependent on the noise at the current time n^k . So the expectation can be first carried out with respect to the n^k with fixed $\hat{\mathbf{h}}^k$ and \mathbf{u}^k :

$$\begin{aligned} \mathcal{E}_{n^k}\left[z^k D\left(-\frac{z^k \hat{\mathbf{h}}^k \mathbf{u}^k}{\sigma_n}\right)\right] &= Q\left(-\frac{\mathbf{h} \mathbf{u}^k}{\sigma_n}\right) D\left(-\frac{\hat{\mathbf{h}}^k \mathbf{u}^k}{\sigma_n}\right) \\ &\quad - Q\left(\frac{\mathbf{h} \mathbf{u}^k}{\sigma_n}\right) D\left(\frac{\hat{\mathbf{h}}^k \mathbf{u}^k}{\sigma_n}\right). \end{aligned} \quad (41)$$

Now, it is assumed that $\hat{\mathbf{h}}^k = \tilde{\mathbf{h}}^k + \mathbf{h}$ is close to \mathbf{h} and a Taylor approximation is applied around the true interference coefficients such that

$$D\left(\frac{\hat{\mathbf{h}}^k \mathbf{u}^k}{\sigma_n}\right) \approx D\left(\frac{\mathbf{h} \mathbf{u}^k}{\sigma_n}\right) + \frac{\tilde{\mathbf{h}}^k \mathbf{u}^k}{\sigma_n} \dot{D}\left(\frac{\mathbf{h} \mathbf{u}^k}{\sigma_n}\right), \quad (42)$$

where $\dot{D}(x)$ denotes the derivative of $D(x)$. It follows, after some simple computations, that

$$\begin{aligned} \mathcal{E}_{n^k}\left[z^k D\left(-\frac{z^k \hat{\mathbf{h}}^k \mathbf{u}^k}{\sigma_n}\right)\right] \\ = -\frac{\tilde{\mathbf{h}}^k \mathbf{u}^k}{\sigma_n} \frac{e^{-(\mathbf{h} \mathbf{u}^k)^2/2\sigma_n^2}}{\sqrt{2\pi}} \left[D\left(-\frac{\mathbf{h} \mathbf{u}^k}{\sigma_n}\right) + D\left(\frac{\mathbf{h} \mathbf{u}^k}{\sigma_n}\right) \right]. \end{aligned} \quad (43)$$

On the other hand,

$$\begin{aligned} \mathcal{E}_{n^k}\left[d^2\left(-\frac{z^k \hat{\mathbf{h}}^k \mathbf{u}^k}{\sigma_n}\right)\right] &\approx Q\left(-\frac{\mathbf{h} \mathbf{u}^k}{\sigma_n}\right) D^2\left(-\frac{\mathbf{h} \mathbf{u}^k}{\sigma_n}\right) \\ &\quad + Q\left(\frac{\mathbf{h} \mathbf{u}^k}{\sigma_n}\right) D^2\left(\frac{\mathbf{h} \mathbf{u}^k}{\sigma_n}\right) \end{aligned} \quad (44)$$

by assuming⁵ $\hat{\mathbf{h}}^k \approx \mathbf{h}$. Finally, by inserting (43) and (44) into (40), the following is obtained:

$$\begin{aligned} \mathcal{E}\left[\frac{(\tilde{\mathbf{h}}^k \mathbf{u}^k)^2 e^{-(\mathbf{h} \mathbf{u}^k)^2/2\sigma_n^2}}{\sigma_n \sqrt{2\pi}} \left[D\left(-\frac{\mathbf{h} \mathbf{u}^k}{\sigma_n}\right) + D\left(\frac{\mathbf{h} \mathbf{u}^k}{\sigma_n}\right) \right]\right] \\ = \frac{\mu}{2} \mathcal{E}\left[\frac{e^{-(\mathbf{h} \mathbf{u}^k)^2/2\sigma_n^2}}{\sqrt{2\pi}} \left[D\left(-\frac{\mathbf{h} \mathbf{u}^k}{\sigma_n}\right) + D\left(\frac{\mathbf{h} \mathbf{u}^k}{\sigma_n}\right) \right] |\mathbf{u}^k|^2\right]. \end{aligned} \quad (45)$$

⁵ It is not necessary this time to use a Taylor approximation because the Taylor correction is much smaller than the 0-order value.

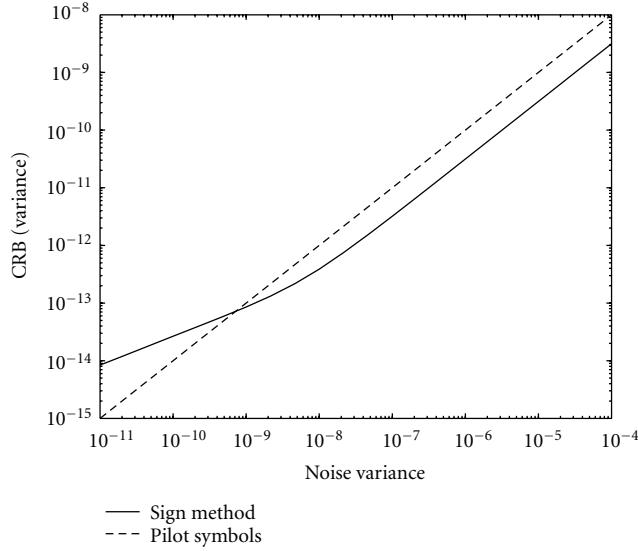


FIGURE 3: Modified CRB for a fixed set of interference coefficients, as a function of the noise variance. Interference coefficients are $1e-6 \cdot [-1.4 \ 2.1 \ -40.9 \ 69.2]$ ($P_{\text{interf}} = 6.5 \cdot 10^{-9}$), $K = 50\,000$, the constellation sizes are 32.

Now since $\tilde{\mathbf{h}}^k$ only depends on *past* noise samples and, symbols

it is independent of \mathbf{u}^k , and the relation can be rewritten as

$$\begin{aligned} & \mathbb{E} \left[\frac{1}{\sqrt{\sigma_n^2}} \tilde{\mathbf{h}}^k \mathcal{E}_{\mathbf{u}^k} \left[\frac{e^{-(\mathbf{h}\mathbf{u}^k)^2/2\sigma_n^2}}{\sqrt{2\pi}} \left[D\left(-\frac{\mathbf{h}\mathbf{u}^k}{\sigma_n}\right) + D\left(\frac{\mathbf{h}\mathbf{u}^k}{\sigma_n}\right) \right] \mathbf{u}^k (\mathbf{u}^k)^T \right] (\tilde{\mathbf{h}}^k)^T \right] \\ &= \frac{\mu}{2} \mathcal{E}_{\mathbf{u}^k} \left[\frac{e^{-(\mathbf{h}\mathbf{u}^k)^2/2\sigma_n^2}}{\sqrt{2\pi}} \left[D\left(-\frac{\mathbf{h}\mathbf{u}^k}{\sigma_n}\right) + D\left(\frac{\mathbf{h}\mathbf{u}^k}{\sigma_n}\right) \right] |\mathbf{u}^k|^2 \right]. \end{aligned} \quad (46)$$

Thanks to the approximations, most of the expectations remaining are taken on \mathbf{u}^k only (\mathbf{h} is the true interference vector and is fixed), which is much more tractable. Note that the inner expectation on the left-hand term is a matrix while the expectation on the right-hand term is a scalar. This matrix equation, for a fixed interference vector, characterizes the behavior of the various estimates in steady state. Defining the covariance matrix of the estimation error (the superscript k is dropped because it corresponds to the steady-state behavior, but we reintroduce the subscript i corresponding to the line of interest),

$$\mathbf{R}_{e,i} = \mathbb{E} \left[(\tilde{\mathbf{h}}_i^\infty)^T \tilde{\mathbf{h}}_i^\infty \right], \quad (47)$$

and defining

$$\mathbf{A}_{1,i} = \mathcal{E}_{\mathbf{u}} \left[\frac{e^{-(\tilde{\mathbf{h}}_i \mathbf{u})^2/2}}{\sqrt{2\pi}} [D(-\tilde{\mathbf{h}}_i \mathbf{u}) + D(\tilde{\mathbf{h}}_i \mathbf{u})] \mathbf{u} (\mathbf{u})^T \right], \quad (48)$$

$$a_{2,i} = \mathcal{E}_{\mathbf{u}} \left[\frac{e^{-(\tilde{\mathbf{h}}_i \mathbf{u})^2/2}}{\sqrt{2\pi}} [D(-\tilde{\mathbf{h}}_i \mathbf{u}) + D(\tilde{\mathbf{h}}_i \mathbf{u})] |\mathbf{u}|^2 \right], \quad (49)$$

the matrix equation (46) can be rewritten in the simpler form

$$\text{Tr} \{ \mathbf{R}_{e,i} \mathbf{A}_{1,i} \} = \frac{\mu \sqrt{\sigma_{n,i}^2}}{2} a_{2,i}. \quad (50)$$

It is readily seen that the definitions (48) and (49) are very similar to the gain definition (27). We have

$$\mathbf{A}_{1,i} \approx \text{diag}([G_{i0} \ \cdots \ G_{iN-1}]), \quad (51)$$

where $\text{diag}(\cdot)$ denotes the diagonal matrix formed with the given elements. The matrix $\mathbf{A}_{1,i}$ can be shown to be approximately diagonal, although the nondiagonal elements are not exactly zero. The diagonal elements are the gains defined in (27). Furthermore,

$$a_{2,i} = \sum_{m=0}^{N-1} G_{i,m}. \quad (52)$$

In practice, both $\mathbf{R}_{e,i}$ and $\mathbf{A}_{1,i}$ are approximately diagonal, so the nondiagonal elements can be neglected in (50), and the

performance can finally be described by

$$\sum_{m=0}^{N-1} G_{i,m} \sigma_{e,i,m}^2 = \frac{\mu \sqrt{\sigma_{n,i}^2}}{2} \sum_{m=0}^{N-1} G_{i,m}, \quad (53)$$

where $\sigma_{e,i,m}^2$ is the estimation error variance for interference coefficient $h_{i,m}$. If we further assume that all the estimates corresponding to one line i have the same error variance, it follows that

$$\sigma_{e,i}^2 = \frac{\mu \sqrt{\sigma_{n,i}^2}}{2}. \quad (54)$$

So, finally, we obtain a very simple expression of the estimation error variance that can be achieved by the algorithm. As we can see, the assumption made in the previous section that the estimation error variances of all interference coefficients are equal between the different lines ($\sigma_{e,i}^2$ equal for all i) is coherent with this result if the noise variances at the various CPEs are the same. The analysis does not provide any justification for the assumption that the estimation error variances are equal within one line i as (53) only provides information about the sum (or a weighted sum) of the variances for that line. However, this assumption was verified to be acceptable by simulations.

5.3. Simulation results

The simulations are performed for $N = 5$ lines, and hence 4 interfering users. The insertion loss and FEXT transfer functions used here come from a set of measurements conducted by France Telecom R&D, which include both the amplitude and phase. A detailed analysis of the measurements is given in [17]. The values used here correspond to a cable of length 300 m, and a tone at frequency around 10 MHz. The other parameters are set according to the standards [18, 19]: the transmitted PSD is limited at -60 dBm/Hz and the noise PSD is -140 dBm/Hz. However, in order to consider different SNR situations, various values around -140 dBm/Hz will be considered. For the computation of the constellation sizes, a target error probability of 10^{-7} is considered with a coding gain of 3 dB and a noise margin of 6 dB.

The first set of simulations aims at comparing the average performance of the proposed method with a classical LMS method. The stepsizes for the proposed algorithm and for the LMS are adjusted so as to provide similar convergence speeds. Several noise variances are investigated. For each one, a set of 1000 simulations is run. Each simulation uses a block of $K = 60\,000$ symbols. The output of the algorithm is taken at the end of the K blocks and the performance (in terms of the estimation error, SIR and SNIR) is averaged over the 1000 simulations. Note that the constellation sizes are always adjusted according to the available SNR on the line. Figure 4 provides the estimation error variance, averaged over all coefficients, for various noise variances (solid line). It is compared to the performance of the LMS using pilots for the same simulation setup (dashed line) but, for a fair comparison, with a lower K_{LMS} (see Section 3.2). The results are presented as a function of the ratio between $P_{\text{interf}} = \sum_m h_{i,m}^2$ and

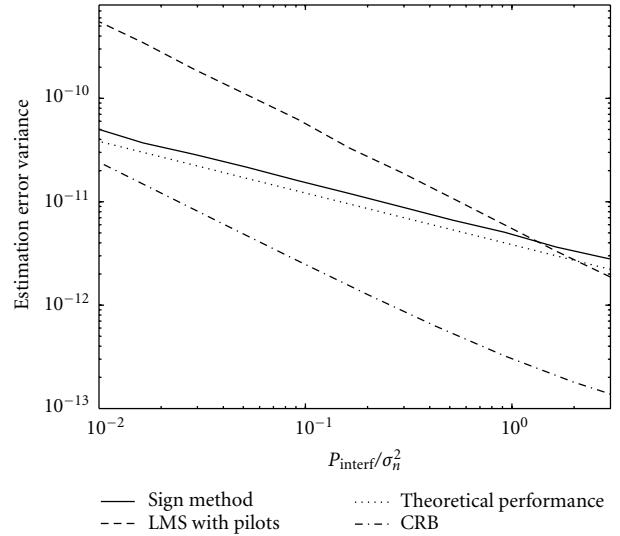


FIGURE 4: Estimation error variance σ_e^2 averaged over 1000 simulations, and averaged over all interference coefficients. The stepsizes are kept fixed, $\mu = 5.10^{-8}$, $\mu_{LMS} = 5.10^{-4}$. The constellation sizes are adjusted according to the SNR. Comparison with the theoretical variance predicted by the analysis and with the CRB.

σ_n^2 , that is, an interference-to-noise ratio. It is clear that, for low interference-to-noise ratio, the proposed method provides better performance. On the contrary, when the ratio becomes large (the noise is low or the power of interference too high), the algorithm does not perform well with respect to the LMS algorithm. The reason is that, for lower noise, the sign of the symbol error no longer provides enough information on the amplitude of the interference coefficients. In conclusion, this algorithm is well suited when the noise is approximately of the same amplitude as the remaining interference from crosstalk. So this is perfectly suited to the issue of interest, since, because of the precoding, the interference coefficients that we try to estimate are usually lower than the noise.

For the same set of simulations, Figure 5 shows the performance of the transmission after computing a new precoder with the available estimations. The average SIR (signal-to-interference ratio) and the average SNIR (signal-to-noise-and-interference ratio) before and after the updated precoder are compared. The bottom curve is always the value before the updated precoder and the top curve is the corresponding result after the updated precoder. The results are presented as a function of the SNR that would be available if the interference was totally removed. The figure shows the good performance obtained by the estimation technique. The resulting precoder decreases the interference to at least 10 dB below the noise. Using the SNIR, the corresponding throughput loss for the given tone can be computed for both methods—the proposed one and the LMS method using pilots. As expected, the proposed method brings some gain when the interference-to-noise ratio is low. The bit rate loss can be up to 3-4 times

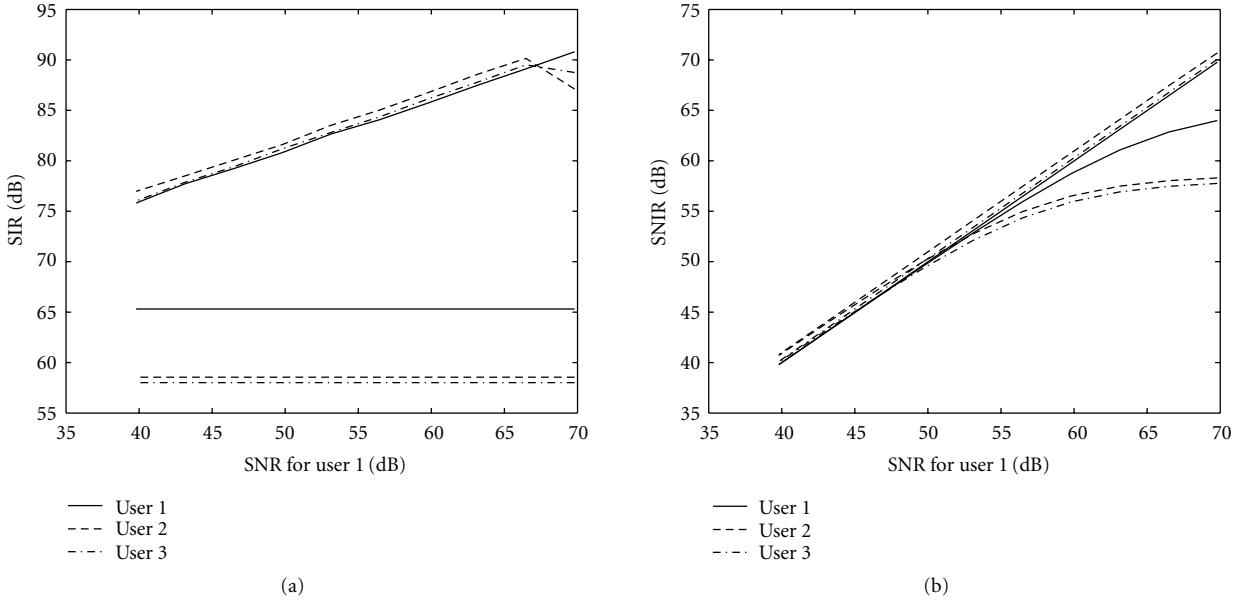


FIGURE 5: (a) Signal-to-interference ratio for different users, averaged over 1000 simulations. (b) Signal-to-noise-and-interference ratio averaged over the same set of simulations. The bottom curve is the result before the updated precoder and the upper curve is the result after the updated precoder.

lower. This bit rate loss is given in Figure 6 using, as an example, a system with 4096 tones and 20 MHz bandwidth. For these values,⁶ the throughput on the studied tone would be on the order of 50–100 kbps depending on the noise variance.

Now, these simulation results can be compared to the analytical developments presented in Section 5. First, the relation (37) between the error variance σ_e^2 (averaged over all interference coefficients) and the corresponding achievable SIR after precoding needs to be verified. For the set of simulations described above, using the error variance from Figure 4 and putting it into (37), the results, in terms of SIR, are provided in Figure 7. It shows a very good correspondence with the averaged SIR from Figure 5 obtained through the simulations by actually computing the precoder for each channel estimation.

Then, we can compare the analytical evaluation (54) of the estimation error variance σ_e^2 to the simulations. Figure 4 shows the theoretical error variance from (54) in dotted lines. It appears that the prediction provides a very good approximation of the achievable estimation error variance. The difference is probably due to non-steady-state behavior. As a matter of fact, the stepsize is usually taken as the smallest value that provides a good convergence on the limited number of samples K available. Hence, even though the convergence is acceptable, the steady-state behavior is usually not reached on that limited number of samples.

It is interesting to note that, for fixed stepsizes, the error variance for the proposed algorithm is proportional to the *square root* of the noise variance (54) while the error variance of the LMS is directly proportional to the noise variance [15]. This is predicted by the analytical results and is very well verified in the simulations: the two curves are almost straight lines with different slopes. This behavior also explains why the proposed method is mainly attractive for a low interference-to-noise ratio.

6. ITERATIVE APPROXIMATE ML ALGORITHM

The biggest drawback of the adaptive algorithm presented above is the necessity to adequately choose the stepsize. As in any LMS-like algorithm, the performance is sensitive to the choice of the stepsize μ , see (54), and the error variance can be reduced by decreasing μ . However the convergence speed is also dependent on the stepsize, and decreasing it tends to slow down the convergence of the algorithm. Hence, the achievable performance is usually still several dB from the CRB due to the fact that the convergence has to be assured on a limited number of samples.

In order to achieve better performance, a block algorithm is presented in this section. The principle is to try to approach the true ML estimate of the interference coefficients given the entire blocks of K symbols. So, instead of simplifying the steepest descent algorithm, as it is done in Section 3, the full gradient, involving the summation over all K observations is used at each step in the iteration. The iterative algorithm can

⁶ Another assumption on these would only change the results by a fixed factor.

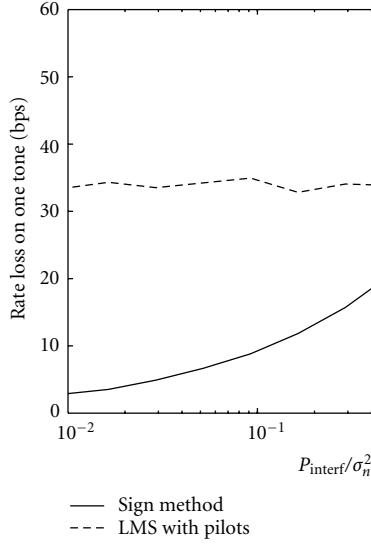


FIGURE 6: Throughput loss for one tone corresponding to the obtained signal-to-noise-and-interference ratio (using standard formulas), for a system with 4096 tones and 20 MHz bandwidth.

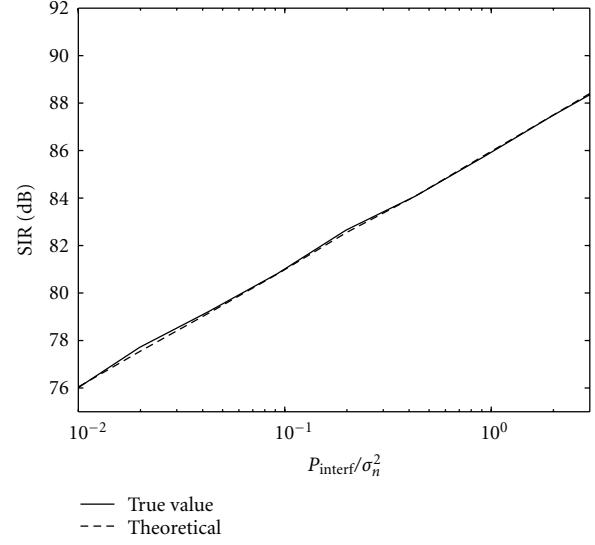


FIGURE 7: Comparison between the true averaged SIR (after updating the precoder) and the theoretical SIR given the estimation error variance, for user 1.

simply be written as

$$\hat{\mathbf{h}}_i^{n+1} = \hat{\mathbf{h}}_i^n + \sum_{k=0}^{K-1} \mu z_i^k D \left(-\frac{z_i^k \hat{\mathbf{h}}_i^n \mathbf{u}^k}{\sqrt{\sigma_{n,i}^2}} \right) (\mathbf{u}^k)^T. \quad (55)$$

Note that, now, for each iteration n , the summation is done over all k . The stepsize can be kept the same as in the LMS-like adaptive algorithm from Section 3. The LMS-like algorithm can be used once on the whole block to provide the initial estimate of the iterative ML algorithm.

The simulations show that this algorithm usually converges in approximately 10 iterations. The convergence can easily be observed through the amplitude of the corrections. Figure 8 shows the results (error variance) of the iterative ML procedure, with 10 iterations following the initial estimation based on the adaptive LMS-like algorithm. It appears that the results are very close to the CRB, as could be expected from an ML algorithm using a large number of observations K . This confirms that the iterative procedure converges sufficiently in around 10 iterations.

7. CONCLUSIONS

We have proposed a new scheme for the tracking of FEXT channel coefficients in downstream VDSL. This scheme is intended for systems with uncoordinated receivers and coordination at the transmitter, using some kind of precoding scheme to remove the influence of FEXT. The principle is to feed back some limited amount of information about the received signals from the receivers to the transmitter in order to allow the estimation of the crosstalk channels at the transmitter (where the precoder needs to be computed). We have proposed a tracking algorithm, based on the maximum

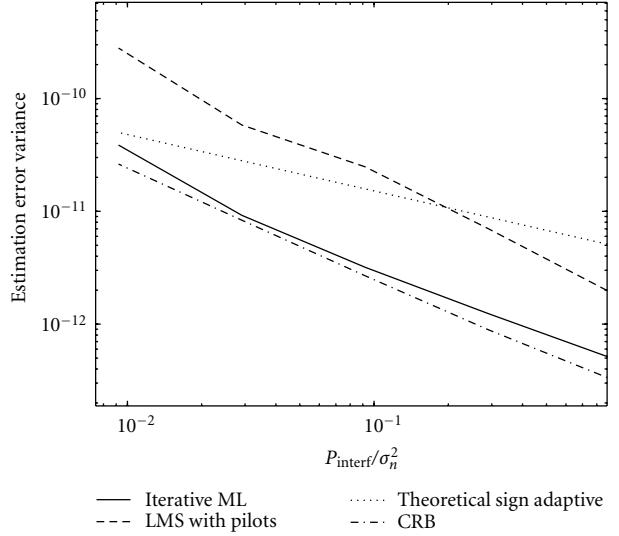


FIGURE 8: Results of the iterative ML algorithm (error variance) for 10 iterations, compared to the CRB and other methods.

likelihood principle, using the sign of the “symbol error” as feedback. We have computed the Cramer-Rao bound and shown that, for a given number of bits to be used as feedback or pilots, the proposed structure exhibits a better potential when the ratio between the interference power and the noise power is low. The simulation results and the analysis have confirmed that the method performs better than a classical scheme using pilot symbols when this ratio is low, which is the case in the problem of interest, due to the presence of the precoder. Analytical results that have been provided

closely approximate the performance of the scheme. Finally, an improved version that has been presented exhibits a performance close to the Cramer-Rao bound. It must be noted that all the results presented here have been provided for real symbols, but the algorithm can easily be extended to complex symbols.

ACKNOWLEDGMENTS

We would like to thank the U-BROAD project partners and in particular the people from France Telecom R&D Labs for providing the crosstalk measurements used in this paper. This research was supported in part by the Commission of the EC under Contract FP6 IST1-506790 (U-BROAD). Parts of this paper were presented at the International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, Pa, USA, March 2005.

REFERENCES

- [1] G. Latsoudas and N. D. Sidiropoulos, "On the performance of certain fixed-complexity multiuser detectors in FEXT-limited vectored DSL systems," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, vol. 3, pp. 889–892, Philadelphia, Pa, USA, March 2005.
- [2] R. Cendrillon, M. Moonen, J. Verlinden, T. Bostoen, and G. Ginis, "Improved linear crosstalk precompensation for DSL," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04)*, vol. 4, pp. 1053–1056, Montreal, Quebec, Canada, May 2004.
- [3] G. Ginis and J. M. Cioffi, "Vectored transmission for digital subscriber line systems," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 5, pp. 1085–1104, 2002.
- [4] A. Leshem and L. Youming, "A low complexity coordinated FEXT cancellation for VDSL," in *Proceedings of 11th IEEE International Conference on Electronics, Circuits and Systems (ICECS '04)*, pp. 338–341, Tel-Aviv, Israel, December 2004.
- [5] M. L. Honig, P. Crespo, and K. Steiglitz, "Suppression of near- and far-end crosstalk by linear pre- and post-filtering," *IEEE Journal on Selected Areas in Communications*, vol. 10, no. 3, pp. 614–629, 1992.
- [6] G. Ginis and J. M. Cioffi, "A multi-user precoding scheme achieving crosstalk cancellation with application to DSL systems," in *Proceedings of the 34th Asilomar Conference on Signals, Systems and Computers (ACSSC '00)*, vol. 2, pp. 1627–1631, Pacific Grove, Calif, USA, October 2000.
- [7] R. Cendrillon, M. Moonen, E. Van den Bogaert, and G. Ginis, "The linear zero-forcing crosstalk canceller is near-optimal in DSL channels," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '04)*, vol. 4, pp. 2334–2338, Dallas, Tex, USA, November–December 2004.
- [8] L. Tong and S. Perreau, "Multichannel blind identification: from subspace to maximum likelihood methods," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1951–1968, 1998.
- [9] Y. Zeng and T.-S. Ng, "A semi-blind channel estimation method for multiuser multiantenna OFDM systems," *IEEE Transactions on Signal Processing*, vol. 52, no. 5, pp. 1419–1429, 2004.
- [10] M. Morelli and U. Mengali, "A comparison of pilot-aided channel estimation methods for OFDM systems," *IEEE Transactions on Signal Processing*, vol. 49, no. 12, pp. 3065–3073, 2001.
- [11] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Transactions on Broadcasting*, vol. 48, no. 3, pp. 223–229, 2002.
- [12] D. J. Love and R. W. Heath Jr., "Limited feedback precoding for spatial multiplexing systems," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '03)*, vol. 4, pp. 1857–1861, San Francisco, Calif, USA, December 2003.
- [13] R. W. Heath Jr. and G. B. Giannakis, "Exploiting input cyclostationarity for blind channel identification in OFDM systems," *IEEE Transactions on Signal Processing*, vol. 47, no. 3, pp. 848–856, 1999.
- [14] J. Louveaux and A.-J. van der Veen, "Downstream VDSL channel tracking using limited feedback for crosstalk precompensated schemes," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, vol. 3, pp. 337–340, Philadelphia, Pa, USA, March 2005.
- [15] A. H. Sayed, *Fundamentals of Adaptive Filtering*, John Wiley & Sons, Hoboken, NJ, USA, 2003.
- [16] M. Moeneclaey, "On the true and the modified Cramer-Rao bounds for the estimation of a scalar parameter in the presence of nuisance parameters," *IEEE Transactions on Communications*, vol. 46, no. 11, pp. 1536–1544, 1998.
- [17] N. D. Sidiropoulos, E. Karipidis, A. Lessem, and L. Youming, "Statistical characterization and modelling of the copper physical channel," Tech. Rep., Deliverable D2.1. EU-FP6 STREP project U-BROAD No. 506790, 2004, http://www.metalinkbb.com/site/app/UBoard_Publications.asp?year=2004
- [18] ETSI, "Transmission and Multiplexing (TM); Access transmission systems on metallic access cables; Very high speed Digital Subscriber Line (VDSL); Part 2: Transceiver specification," ETSI TS 101 270-2, 2000.
- [19] J. A. C. Bingham, *ADSL, VDSL, and Multicarrier Modulation*, John Wiley & Sons, New York, NY, USA, 2000.

J. Louveaux received the Electrical Engineering degree and the Ph.D. degree from the Université Catholique de Louvain (UCL), Louvain-la-Neuve, Belgium, in 1996 and 2000, respectively. From 2000 to 2001, he was a Visiting Scholar in the Electrical Engineering Department at Stanford University, Calif. He is currently a Postdoctoral Researcher at the Delft University of Technology, The Netherlands. His research interests are in signal processing for digital communications, mainly synchronization issues and high-bit-rate transmission over wired channels. His current specific interests are in crosstalk cancellation techniques in DSL systems. He serves as an Associate Editor for the IEEE Communications Letters since 2003. He is corecipient of the "Prix Biennal Siemens 2000" and the "Prix Scientifique Alcatel 2005."



A.-J. van der Veen was born in The Netherlands in 1966. He graduated (cum laude) from the Department of Electrical Engineering, Delft University of Technology, in 1988, and received the Ph.D. degree (cum laude) from the same institute in 1993. Throughout 1994, he was a postdoctoral scholar at Stanford University, in the Scientific Computing/Computational Mathematics group and in the Information Systems Lab. At present, he is a Full Professor in the Signal Processing group of DIMEs, Delft University of Technology. He is the recipient of a 1994 and a 1997 IEEE SPS Young Author Paper Award, and was an Associate Editor for the IEEE Transactions on Signal Processing (1998–2001), Chairman of the IEEE SPS SPCOM Technical Committee (2002–2004), and Editor-in-Chief of the IEEE Signal Processing Letters (2002–2005). He is currently the Editor-in-Chief of the IEEE Transactions on Signal Processing. His research interests are in the general area of system theory applied to signal processing, and in particular algebraic methods for array signal processing and signal processing for communications.

