

The Extended-Window Channel Estimator for Iterative Channel-and-Symbol Estimation

Renato R. Lopes

DSPCom, DECOM, FEEC, University of Campinas (UNICAMP), 400 Albert Einstein Avenue, 13083-970 Campinas, Sao Paulo, Brazil
Email: rlopes@decom.fee.unicamp.br

John R. Barry

School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0250, USA
Email: barry@ece.gatech.edu

Received 29 April 2004; Revised 21 September 2004

The application of the expectation-maximization (EM) algorithm to channel estimation results in a well-known iterative channel-and-symbol estimator (ICSE). The EM-ICSE iterates between a symbol estimator based on the forward-backward recursion (BCJR equalizer) and a channel estimator, and may provide approximate maximum-likelihood blind or semiblind channel estimates. Nevertheless, the EM-ICSE has high complexity, and it is prone to misconvergence. In this paper, we propose the extended-window (EW) estimator, a novel channel estimator for ICSE that can be used with any soft-output symbol estimator. Therefore, the symbol estimator may be chosen according to performance or complexity specifications. We show that the EW-ICSE, an ICSE that uses the EW estimator and the BCJR equalizer, is less complex and less susceptible to misconvergence than the EM-ICSE. Simulation results reveal that the EW-ICSE may converge faster than the EM-ICSE.

Keywords and phrases: blind channel estimation, EM algorithm, maximum-likelihood estimation, iterative systems.

1. INTRODUCTION

Channel estimation is an important part of communications systems. Channel estimates are required by equalizers that minimize the bit error rate (BER), and can be used to compute the coefficients of suboptimal but lower-complexity equalizers such as the minimum mean-squared error (MMSE) linear equalizer (LE) [1], or the decision-feedback equalizer (DFE) [1]. Traditionally, a sequence of known bits, called a training sequence, is transmitted for the purpose of channel estimation [1]. These known symbols and their corresponding received samples are used to estimate the channel. However, this approach, known as trained estimation, ignores received samples corresponding to the information bits, and thus does not use all the information available at the receiver. Alternatively, semiblind estimators [2] use every available channel output for channel estimation. Thus, they outperform estimators based solely on the channel outputs corresponding to training symbols, and require a shorter training sequence. Channel estimation is still

possible even if no training sequence is available, using a technique known as blind channel estimation.

An important class of algorithms for blind and semiblind channel estimation is based on the iterative strategy depicted in Figure 1 [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14], which we call *iterative channel-and-symbol estimation* (ICSE). In these algorithms, an initial channel estimate is used by a *symbol estimator* to provide initial estimates of the first-order (and possibly also the second-order) statistics of the transmitted symbol sequence. These estimates are used by a *channel estimator* to improve the initial channel estimates. The process is then repeated. The hope is that several iterations between these two low-complexity estimators will lead to estimates that nearly maximize the joint likelihood function.

The application of the expectation-maximization (EM) algorithm, also known as the Baum-Welch algorithm [15, 16], to the blind channel estimation problem results in the canonical ICSE that fits the framework of Figure 1. An EM iterative channel-and-symbol estimator (EM-ICSE) was first reported in [4], and it has some useful properties. First, it generates a sequence of estimates with nondecreasing likelihood, so that the channel estimates are capable of approaching the maximum-likelihood (ML) estimates.

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

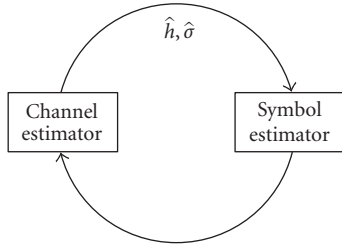


FIGURE 1: Iterative channel-and-symbol estimator.

Second, its symbol estimator is based on the forward-backward recursion of Bahl et al. (BCJR) [17], which minimizes the probability of decision error. Third, the EM-ICSE may be easily modified to exploit, in a natural and nearly optimal way, any a priori information the receiver may have about the transmitted symbols. This a priori information may arise because of pilot symbols (e.g., in semiblind estimation) or error-control coding (e.g., in the context of turbo equalization [6, 7, 8, 9]).

The application of iterative channel estimation to turbo equalization is particularly important, since it leads to channel estimates that benefit from the presence of channel coding, thus performing well at low signal-to-noise ratios [6, 7, 8, 9]. This is particularly important because powerful codes such as turbo codes [18, 19] allow reliable communication at extremely low signal-to-noise ratios, which only exacerbates the estimation problem for traditional channel estimators that ignore the existence of coding, as is the case with most blind channel estimation techniques.

The EM-ICSE has two main drawbacks that we address in this paper: its tendency to converge to inaccurate channel estimates, and its high computational complexity. The problem of convergence to inaccurate estimates arises because the EM-ICSE necessarily generates a sequence of estimates with nondecreasing likelihood. This property makes the EM-ICSE susceptible to getting trapped in a local maximum of the likelihood function. Also, the EM-ICSE has two sources of complexity. First, the EM channel estimator involves the computation and inversion of a square matrix whose order is equal to the channel length. Second, and more important, the complexity of the EM symbol estimator is exponential in the channel length. In [11, 12], ICSEs are proposed that reduce the complexity of the EM-ICSE by introducing a low-complexity symbol estimator. However, these works focus only on the symbol estimator, and use the same channel estimator as the EM-ICSE, resulting in a computational complexity that grows with the square of the channel memory.

In this work, we focus on the channel estimator of Figure 1. We will propose the simplified EM channel estimator (SEM), a channel estimator for ICSE that avoids the matrix inversion of the EM channel estimator. More importantly, an ICSE based on the SEM channel estimator does not require the BCJR equalizer, and thus may be implemented with any number of low-complexity alternatives to the BCJR algorithm, such as those proposed in [20, 21]. Since the complexity of the SEM channel estimator is linear in the channel memory, the overall complexity of an ICSE based on the SEM

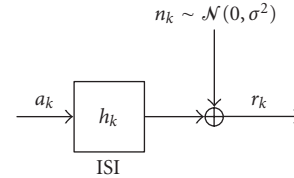


FIGURE 2: Channel model.

channel estimator is also linear if a linear-complexity equalizer is used.

We will also investigate the convergence of an ICSE based on the SEM estimator. We will see that, after misconvergence, the SEM channel estimates may have a structure that can be exploited to escape the local maximum of the likelihood function. We then propose the extended-window (EW) channel estimator, a simple modification to the SEM channel estimator that exploits this structure and greatly decreases the probability of misconvergence, without significantly affecting the computational complexity.

This paper is organized as follows. In Section 2 we present the channel model and describe the problem we will investigate. In Section 3, we briefly review the EM-ICSE. In Section 4, we propose the SEM estimator, a linear-complexity channel estimator for ICSE that is not intrinsically linked to a symbol estimator. In Section 5, we propose the EW estimator, an extension to the SEM estimator of Section 4 that is less likely than EM to get trapped in a local maximum of the joint likelihood function. In Section 6, we present some simulation results, and we draw some conclusions in Section 7.

2. CHANNEL MODEL AND PROBLEM STATEMENT

Consider the transmission of K zero-mean, uncorrelated symbols a_k belonging to some alphabet \mathcal{A} , with unit energy $E[|a_k|^2] = 1$, across a dispersive channel with memory μ and additive-white Gaussian noise. The received signal at time k can be written as

$$r_k = \mathbf{h}^T \mathbf{a}_k + n_k, \quad (1)$$

where $\mathbf{h} = (h_0, h_1, \dots, h_\mu)^T$ represents the channel impulse response, $\mathbf{a}_k = (a_k, a_{k-1}, \dots, a_{k-\mu})^T$, and n_k represents white Gaussian noise with variance σ^2 . Let $\mathbf{a} = (a_0, a_1, \dots, a_{K-1})$ and $\mathbf{r} = (r_0, r_1, \dots, r_{N-1})$ denote the input and output sequences, respectively, where $N = K + \mu$. The resulting channel model is depicted in Figure 2.

Notice that, as far as channel estimation is concerned, the assumption that the transmitted symbols are uncorrelated is not too restrictive. Indeed, most training sequences are chosen so as to satisfy this assumption (thus minimizing the Cramér-Rao bound [22]) and the presence of an interleaver in most coded systems also ensures that the transmitted sequence is approximately uncorrelated. In other words, for channel estimation purposes, assuming that the transmitted

symbols are uncorrelated does not exclude the presence of a training sequence or of a channel code. As we will see, it is the symbol estimator in Figure 1 that exploits the presence of a training sequence or of a channel code.

This paper concerns the joint estimation of \mathbf{a} , \mathbf{h} , and σ relying solely on the received signal \mathbf{r} . Ideally, we would like to solve the joint-ML channel estimation and symbol detection problem, that is, find

$$(\hat{\mathbf{a}}_{\text{ML}}, \hat{\mathbf{h}}_{\text{ML}}, \hat{\sigma}_{\text{ML}}) = \arg \max \log p_{\mathbf{h}, \sigma}(\mathbf{r}|\mathbf{a}), \quad (2)$$

where $\log p_{\mathbf{h}, \sigma}(\mathbf{r}|\mathbf{a})$ is the log-likelihood function, defined as the logarithm of the pdf of the received signal \mathbf{r} conditioned on the channel input \mathbf{r} and parameterized by \mathbf{r} and σ . Intuitively, the ML estimates are those that best explain the received sequence, in the sense that we are less likely to observe the channel output if we assume any other set of parameters to be correct, that is, $p_{\mathbf{h}, \sigma}(\mathbf{r}|\mathbf{a}) \geq p_{\mathbf{h}_{\text{ML}}, \sigma_{\text{ML}}}(\mathbf{r}|\mathbf{a}_{\text{ML}})$ for all \mathbf{h} , σ , \mathbf{a} . Besides this intuitive interpretation, ML estimates have many interesting theoretical properties [22].

It is noteworthy that the maximization in (2) should be performed over the set of valid transmitted sequences. Thus, the joint-ML channel-and-symbol estimation problem in (2) incorporates all possible scenarios: fully trained estimation (all of \mathbf{a} is known); semiblind estimation without coding (parts of \mathbf{a} are known, unknown parts of \mathbf{a} can be any sequence of symbols); semiblind estimation with coding (parts of \mathbf{a} are known, \mathbf{a} must be a valid codeword); blind estimation without coding (none of \mathbf{a} is known, \mathbf{a} can be any sequence of symbols); and blind estimation with coding (none of \mathbf{a} is known, \mathbf{a} must be a valid codeword).

Unfortunately, a direct solution to the problem in (2) is too complex. Therefore, this paper focuses on iterative techniques that provide an approximate solution to (2) with reasonable computational complexity. In the sequel, we review the EM-ICSE, an ICSE that computes a sequence of estimates with nondecreasing likelihood and that, with proper initialization or if the likelihood function is well-behaved, will converge to the ML estimates.

3. THE EM-ICSE

The EM algorithm [15, 16] provides an iterative solution to the blind identification problem in (2) that fits the paradigm of Figure 1, as first reported in [4]. The EM channel estimator (see Figure 1) for the $(i+1)$ th iteration of the EM-ICSE is defined by

$$\hat{\mathbf{h}}_{(i+1)} = \mathbf{R}_i^{-1} \mathbf{p}_i, \quad (3)$$

$$\begin{aligned} \hat{\sigma}_{(i+1)}^2 &= \frac{1}{N} \sum_{k=0}^{N-1} \mathbb{E} \left[|r_k - \hat{\mathbf{h}}_{(i+1)}^T \mathbf{a}_k|^2 | \mathbf{r}; \hat{\mathbf{h}}_{(i)}, \hat{\sigma}_{(i)}^2 \right] \\ &= \frac{1}{N} \sum_{k=0}^{N-1} |r_k|^2 - 2\hat{\mathbf{h}}_{(i+1)}^T \mathbf{p}_i + \hat{\mathbf{h}}_{(i+1)}^T \mathbf{R}_i \hat{\mathbf{h}}_{(i+1)}, \end{aligned} \quad (4)$$

where

$$\mathbf{R}_i = \frac{1}{N} \sum_{k=0}^{N-1} \mathbb{E} [\mathbf{a}_k \mathbf{a}_k^T | \mathbf{r}; \hat{\mathbf{h}}_{(i)}, \hat{\sigma}_{(i)}^2], \quad (5)$$

$$\mathbf{p}_i = \frac{1}{N} \sum_{k=0}^{N-1} r_k \mathbb{E} [\mathbf{a}_k | \mathbf{r}; \hat{\mathbf{h}}_{(i)}, \hat{\sigma}_{(i)}^2]. \quad (6)$$

The EM symbol estimator (see Figure 1) provides the values of $\hat{\mathbf{a}}_k^{(i)} = \mathbb{E}[\mathbf{a}_k | \mathbf{r}; \hat{\mathbf{h}}_{(i)}, \hat{\sigma}_{(i)}^2]$ and $\mathbb{E}[\mathbf{a}_k \mathbf{a}_k^T | \mathbf{r}; \hat{\mathbf{h}}_{(i)}, \hat{\sigma}_{(i)}^2]$ that are required by (5) and (6). The a posteriori expected values in (5) and (6) are computed assuming that $\hat{\mathbf{h}}_{(i)}$ and $\hat{\sigma}_{(i)}^2$ are the actual channel parameters. Notice that $\hat{a}_k = \mathbb{E}[a_k | \mathbf{r}; \hat{\mathbf{h}}_{(i)}, \hat{\sigma}_{(i)}^2]$ is the a posteriori MMSE estimate of a_k , and we refer to \hat{a}_k as a *soft symbol estimate*.

Also, note that \mathbf{R}_i and \mathbf{p}_i of (5) and (6) can be viewed as estimates of the a posteriori autocorrelation matrix of the transmitted sequence and the cross-correlation vector between the transmitted and received sequences, respectively. Thus, (3) and (4) are similar to the MMSE-trained channel estimator [22], in which \mathbf{R}_i and \mathbf{p}_i are computed with the actual transmitted sequence.

The computation of the expected values in (5) and (6) require the knowledge of the a posteriori probabilities $\mathbb{E}[\mathbf{a}_k | \mathbf{r}; \hat{\mathbf{h}}_{(i)}, \hat{\sigma}_{(i)}^2]$ and $\mathbb{E}[\mathbf{a}_k \mathbf{a}_k^T | \mathbf{r}; \hat{\mathbf{h}}_{(i)}, \hat{\sigma}_{(i)}^2]$. For an uncoded system, these can be exactly computed with the forward-backward recursion or BCJR algorithm [17]. Because the computational complexity of this algorithm grows exponentially with the channel length, some authors [11, 12] have proposed lower-complexity alternatives that compute approximations to these a posteriori probabilities. In other words, the algorithms of [11, 12] are approximations to the EM-ICSE that also fit the framework of Figure 1, and that are also based on the channel estimator of (3), (4), (5), and (6).

Unfortunately, in the presence of a channel code, an exact computation of \mathbf{R}_i and \mathbf{p}_i is prohibitively complex. The most common solution in this case is to modify the EM-ICSE, using a turbo equalizer as the symbol estimator [6]. In other words, for coded systems, $\mathbb{E}[\mathbf{a}_k | \mathbf{r}; \hat{\mathbf{h}}_{(i)}, \hat{\sigma}_{(i)}^2]$ and $\mathbb{E}[\mathbf{a}_k \mathbf{a}_k^T | \mathbf{r}; \hat{\mathbf{h}}_{(i)}, \hat{\sigma}_{(i)}^2]$ are based on the decoder output. Similarly, the presence of training symbols is easily handled by the symbol estimator, which only has to set the training symbols as deterministic constants when computing \mathbf{R}_i and \mathbf{p}_i . Based on these two observations, we see that the channel estimator of the EM-ICSE always ignores the presence of a training sequence or of a channel code. It is the symbol estimator that exploits the structure of the transmitted symbols to improve their estimates.

4. A SIMPLIFIED EM CHANNEL ESTIMATOR

In this section, we propose the simplified EM estimator (SEM), an alternative to the EM channel estimator in (3), (4), (5), and (6) that avoids the computation of \mathbf{R}_i and the matrix inversion of (3). To derive the SEM estimator, we note that, from channel model (1) and the uncorrelatedness assumption, we get $h_n = \mathbb{E}[r_k a_{k-n}]$. This expected value may

be computed by conditioning on \mathbf{r} , yielding

$$E[r_k a_{k-n}] = E[E[r_k a_{k-n} | \mathbf{r}]] = E[r_k E[a_{k-n} | \mathbf{r}]], \quad (7)$$

where the last equality follows from the fact that r_k is a constant given \mathbf{r} . Note that the channel estimator has no access to $E[a_k | \mathbf{r}]$, which requires exact channel knowledge. However, based on the iterative paradigm of Figure 1, at the i th iteration the channel estimator does have access to $\tilde{a}_k^{(i)} = E[a_k | \mathbf{r}; \hat{\mathbf{h}}_{(i)}, \hat{\sigma}_{(i)}^2]$. Replacing this value in (7), and also replacing a time average for the ensemble average, leads to the following channel estimator:

$$\hat{h}_n^{(i+1)} = \frac{1}{N} r_k \tilde{a}_{k-n}^{(i)} \quad \text{for } n = 0, 1, \dots, \mu. \quad (8)$$

Notice that in (8) the channel is estimated by correlating the received signal with the soft symbol estimates \tilde{a}_k . This is similar to the fully trained channel estimator of [23, 24], known as channel probing, except that the training symbols have been replaced by their soft estimates.

As for estimating the noise variance, let $\hat{a}_k^{(i)}$ be a hard decision of the k th transmitted symbol, chosen as the element of \mathcal{A} closest to $\tilde{a}_k^{(i)}$. Also, define the vector $\hat{\mathbf{a}}_k^{(i)} = (\hat{a}_k^{(i)}, \hat{a}_{k-1}^{(i)}, \dots, \hat{a}_{k-\mu}^{(i)})^T$. We propose to compute $\hat{\sigma}_{(i+1)}^2$ using

$$\hat{\sigma}_{(i+1)}^2 = \frac{1}{N} \sum_{k=0}^{N-1} |r_k - \hat{\mathbf{h}}_{(i+1)}^T \hat{\mathbf{a}}_k^{(i)}|^2. \quad (9)$$

Notice that in (9) we use hard instead of soft symbol estimates. In our simulations, we found that doing so improved convergence speed.

Remark 1. Combining the estimates (8) into a single vector, we find that $\hat{\mathbf{h}}_{(i+1)} = (\hat{h}_0^{(i+1)}, \dots, \hat{h}_\mu^{(i+1)})^T = \mathbf{p}_i$. Thus, we may view (8) as a simplification of the EM estimate $\mathbf{R}_i^{-1} \mathbf{p}_i$ that avoids matrix inversion by approximating \mathbf{R}_i by \mathbf{I} . This approximation is reasonable, since \mathbf{R}_i is a posteriori estimate of the autocorrelation matrix of the transmitted vector, which, due to the uncorrelatedness assumption, is close to the identity for large N . Furthermore, since this approximation results in a channel estimator that is less complex than the EM channel estimator defined in (3) and (4), we refer to the channel estimator defined by (8) and (9) as the simplified EM estimator (SEM).

Remark 2. The SEM channel estimator requires only the soft symbol estimates $\tilde{a}_k^{(i)}$, so that an ICSE based on the SEM estimator may be represented as in Figure 3. Note that any equalizer that produces soft symbol estimates can be used, which allows for an even lower-complexity implementation of an SEM-based ICSE, using equalizers such as those proposed in [20, 21].

Remark 3. It is interesting to notice that, while substituting the actual values of \mathbf{h} or \mathbf{a} for their estimates will always improve the performance of the iterative algorithm, the same is not true for σ . Indeed, substituting σ for $\hat{\sigma}$ will often result

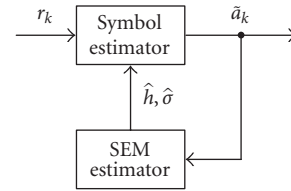


FIGURE 3: Iterative channel-and-symbol estimation with the SEM channel estimator.

in performance degradation. Intuitively, one can think of $\hat{\sigma}$ as playing two roles: in addition to measuring σ , it also acts as a measure of reliability of the channel estimate $\hat{\mathbf{h}}$. In fact, consider a decomposition of the channel output:

$$r_k = \hat{\mathbf{h}}^T \mathbf{a}_k + (\mathbf{h} - \hat{\mathbf{h}})^T \mathbf{a}_k + n_k. \quad (10)$$

The term $(\mathbf{h} - \hat{\mathbf{h}})^T \mathbf{a}_k$ represents the contribution to r_k from the estimation error. By using $\hat{\mathbf{h}}$ to model the channel in the BCJR algorithm, we are in effect lumping the estimation error with the noise, resulting in an effective noise sequence with variance larger than σ^2 . It is thus appropriate that $\hat{\sigma}$ should exceed σ whenever $\hat{\mathbf{h}}$ differs from \mathbf{h} . Alternatively, it stands to reason that an unreliable channel estimate should translate to an unreliable symbol estimate, regardless of how well $\hat{\mathbf{h}}^T \mathbf{a}_k$ matches r_k . Using a large value of $\hat{\sigma}$ in the BCJR equalizer ensures that its output will have a small reliability. Fortunately, the noise variance estimate produced by (9) measures the energy of both the second and the third term in (10). If $\hat{\mathbf{h}}$ is a poor channel estimate, $\tilde{\mathbf{a}}$ will also be a poor estimate for \mathbf{a} , and convolving $\tilde{\mathbf{a}}$ and $\hat{\mathbf{h}}$ will produce a poor match for \mathbf{r} , so that (9) will produce a large estimated noise variance.

5. THE EXTENDED-WINDOW CHANNEL ESTIMATOR

Misconvergence is a common characteristic of ICSEs, especially in blind systems. To illustrate this problem, consider estimating the channel $\mathbf{h} = (1, 2, 3, 4, 5)^T$ with a BPSK constellation and $\text{SNR} = \|\mathbf{h}\|^2 / \sigma^2 = 20$ dB. An ICSE based on the BCJR symbol estimator and the SEM channel estimator converges to $\hat{\mathbf{h}}^{(20)} = (2.1785, 3.0727, 4.1076, 5.0919, 0.1197)^T$ after 20 iterations, with $K = 1000$ bits, with initialization $\hat{\mathbf{h}}^{(0)} = (1, 0, 0, 0, 0)^T$ and $\hat{\sigma}_{(0)}^2 = 1$. Although the algorithm fails, $\hat{\mathbf{h}}^{(20)}$ is seen to roughly approximate a shifted (or delayed) and truncated version of the actual channel. A possible explanation for this behavior is that the channel is maximum phase, while we used a minimum phase initialization. This phase mismatch between $\hat{\mathbf{h}}$ and the initialization $\hat{\mathbf{h}}^{(0)}$ introduces a delay that cannot be compensated for by the iterative scheme. In fact, after convergence, a_k is approximately $\text{sign}(\tilde{a}_{k+1})$, and h_0 can be accurately estimated by correlating r_k with \tilde{a}_{k+1} . However, because the delay n in (8) is limited to the narrow window $0, \dots, \mu$, this correlation is never computed. This observation leads us to propose

the *extended-window* (EW) channel estimator, in which (8) is computed for a broader range of n .

To determine how much the correlation window must be extended, consider two extreme cases. First, suppose $\mathbf{h} \approx (0, \dots, 0, 0, 1)^T$, so that $r_k \approx a_{k-\mu} + n_k$. Also, assume that $\hat{\mathbf{h}} \approx (1, 0, 0, \dots, 0)^T$. In this case, assuming a BPSK constellation, the symbol estimator output is $\hat{a}_k = \tanh(r_k/\sigma^2)$. Hence, assuming a large SNR, $\hat{a}_k \approx a_{k-\mu}$, so to estimate h_0 and h_μ we must compute (8) for $n = -\mu$ and $n = 0$, respectively. Likewise, if $\mathbf{h} \approx (1, 0, 0, \dots, 0)^T$ and $\hat{\mathbf{h}} \approx (0, \dots, 0, 0, 1)^T$, the symbol estimator output \hat{a}_k is such that $\hat{a}_k \approx a_{k+\mu}$, so to estimate h_0 and h_μ we must compute (8) for $n = \mu$ and $n = 2\mu$, respectively. These observations, based on two extreme cases, suggest the *extended-window* (EW) channel estimator, which computes

$$g_n = \frac{1}{N} \sum_{k=0}^{N-1} r_k \hat{a}_{k-n}^{(i)} \quad \text{for } n = -\mu, \dots, 2\mu. \quad (11)$$

By doing this, we ensure that $\mathbf{g} = (g_{-\mu}, \dots, g_{2\mu})^T$ has $\mu + 1$ entries that estimate the desired correlations $E[r_k a_{k-n}]$, for $n \in \{0, \dots, \mu\}$. Its remaining terms are an estimate of $E[r_k a_{k-n}]$ for $n \notin \{0, \dots, \mu\}$, and hence should be close to zero. Therefore, we define the EW channel estimates by

$$\hat{\mathbf{h}}_{(i+1)} = (g_\delta, \dots, g_{\delta+\mu})^T, \quad (12)$$

where the delay parameter $\delta \in \{-\mu, \dots, \mu\}$ is chosen so that $\hat{\mathbf{h}}_{(i+1)}$ represents the $\mu + 1$ consecutive coefficients of \mathbf{g} with highest energy. In other words, δ is chosen to maximize $\|\hat{\mathbf{h}}_{(i+1)}\|^2$.

Notice that after convergence we expect that $g_\delta \approx h_0$. Comparing (7) and (11), we note that this is equivalent to saying that $a_k \approx \hat{a}_{k-\delta}^{(i)}$. This delay must be taken into account in the estimation of the noise variance. With that in mind, we propose to estimate σ^2 using a modified version of (9), namely

$$\hat{\sigma}_{(i+1)}^2 = \frac{1}{N} \sum_{k=0}^{N-1} |r_k - \hat{\mathbf{h}}_{(i+1)}^T \hat{\mathbf{a}}_{k-\delta}^{(i)}|^2. \quad (13)$$

5.1. Computational complexity

We now compare the computational complexity of the EW channel estimator of (11), (12), and (13) to that of the EM channel estimator of (3) and (4). We ignore the cost of computing \hat{a}_k , and we consider the complexity in terms of sums and multiplications per received symbol.

For each received symbol, the EW algorithm performs $3\mu + 1$ multiplications and $3\mu + 1$ additions to compute the vector \mathbf{g} in (11). The division by N , as well as the computation of δ , is done only once per block of N received symbols, and thus can be ignored. The computation of each term in the summation in (13) involves $\mu + 2$ multiplications and the same number of sums. Hence, the total computational cost of the EW channel estimator is $4\mu + 4$ multiplications and $4\mu + 4$ sums.

For the EM channel estimator, we consider that $E[\mathbf{a}_k \mathbf{a}_k^T | \mathbf{r}; \hat{\mathbf{h}}_{(i)}, \hat{\sigma}_{(i)}^2] \approx E[\mathbf{a}_k | \mathbf{r}; \hat{\mathbf{h}}_{(i)}, \hat{\sigma}_{(i)}^2] E[\mathbf{a}_k | \mathbf{r}; \hat{\mathbf{h}}_{(i)}, \hat{\sigma}_{(i)}^2]^T$. This approximation is used in [11, 12], and allows for a simpler complexity comparison. With this simplification, and noting that $E[\mathbf{a}_k \mathbf{a}_k^T | \mathbf{r}; \hat{\mathbf{h}}_{(i)}, \hat{\sigma}_{(i)}^2]$ is a symmetric matrix, we see that the computation of \mathbf{R}_i in (5) requires $(\mu + 1)\mu/2$ multiplications and an equal number of sums per received symbol. On the other hand, the computation of \mathbf{p}_i in (6) requires $\mu + 1$ multiplications and sums per received symbol. The linear system in (3) is solved only once, so that its cost can be ignored. The same can be said about most of the operations in (4), except for its first term, which requires 1 multiplication and sum per received symbol. Thus, the total cost of this approximate EM channel estimator is $\mu^2/2 + 3\mu/2 + 2$ multiplications and sums per received symbol.

6. SIMULATION RESULTS

In this section, we use simulations to compare the performance of the fully blind EM-ICSE and the fully blind EW-ICSE, assuming both ICSEs use the BCJR symbol estimator. The results presented in this section all correct for the aforementioned shifts in the estimates. In other words, when computing channel estimation error or BER, the channel and symbol estimates were shifted to best match the actual channel or the transmitted sequence. Note that this shift was done only for the purpose of computing the errors, and hence did not affect the estimates in the iterative procedure.

For comparison purposes, we also consider fully trained channel estimators, in which all the transmitted bits are assumed known by the channel estimator. We consider the fully trained MMSE estimator which, as discussed in Section 3, can be seen as a trained version of the EM channel estimator. We also consider channel probing which, as discussed in Section 4, can be seen as the trained counterpart of the EW channel estimator. In the simulations of the trained estimators, we use the same block of received samples to estimate the channel (assuming that all transmitted symbols are known) and to estimate the transmitted symbols (with the BCJR equalizer, using the trained channel estimates).

As a first test of the EW-ICSE, we simulated the transmission of $K = 600$ BPSK symbols over the channel $\mathbf{h} = (-0.2287, 0.3964, 0.7623, 0.3964, -0.2287)^T$ from [12]. To stress the fact that the EW-ICSE is not sensitive to initial conditions, we initialized $\hat{\mathbf{h}}$ randomly using $\hat{\mathbf{h}}_{(0)} = \mathbf{u}\hat{\sigma}_{(0)}/\|\mathbf{u}\|$, where $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\hat{\sigma}_{(0)}^2 = \sum_{k=0}^{N-1} |r_k|^2/2N$. By assigning half of the received energy to the signal and half to the noise, we are essentially initializing the SNR estimate to 0 dB. In Figure 4, we show the convergence behavior of the EW-ICSE estimates, averaged over 100 independent runs of this experiment using $\text{SNR} = \|\mathbf{h}\|^2/\sigma^2 = 9$ dB. Only the convergence of \hat{h}_0 , \hat{h}_1 , and \hat{h}_2 is shown; the behavior of \hat{h}_3 and \hat{h}_4 is similar to that of \hat{h}_2 and \hat{h}_0 , respectively, but we show only the coefficients with worse convergence. The shaded regions around the channel estimates correspond to plus and minus one standard deviation. For comparison, we show the average behavior of the EM channel estimates in Figure 5.

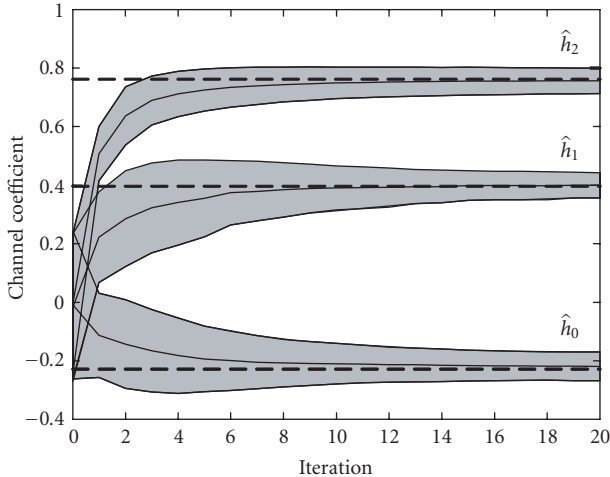


FIGURE 4: Estimates of $\mathbf{h} = (-0.2287, 0.3964, 0.7623, 0.3964, -0.2287)^T$, produced by the EW-ICSE. Dashed lines correspond to the actual channel coefficients.

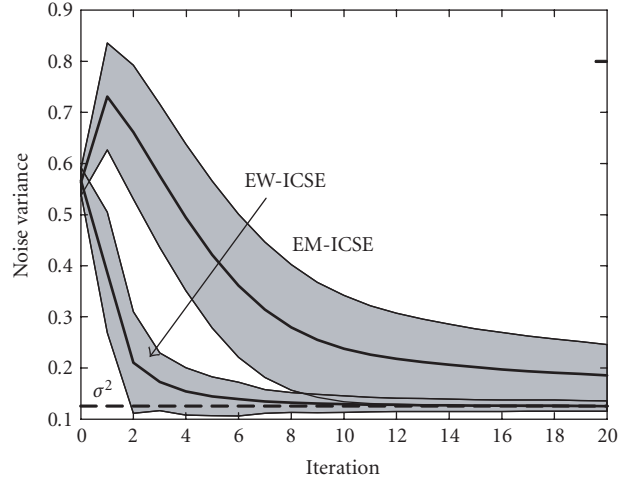


FIGURE 6: Estimates of σ^2 , produced by the EW-ICSE and the EM-ICSE.

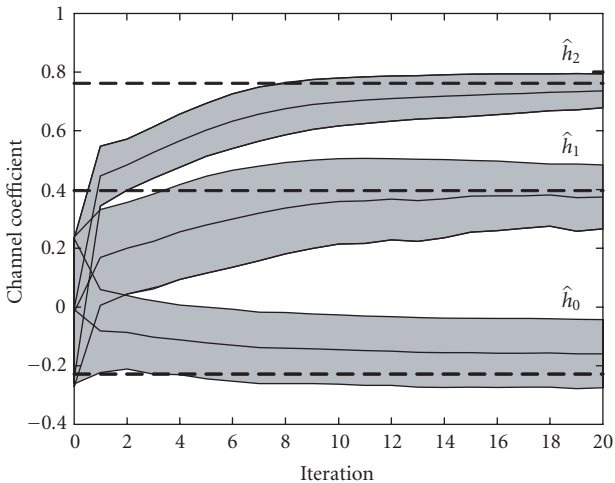


FIGURE 5: EM estimates of $\mathbf{h} = (-0.2287, 0.3964, 0.7623, 0.3964, -0.2287)^T$. Dashed lines correspond to the actual channel coefficients.

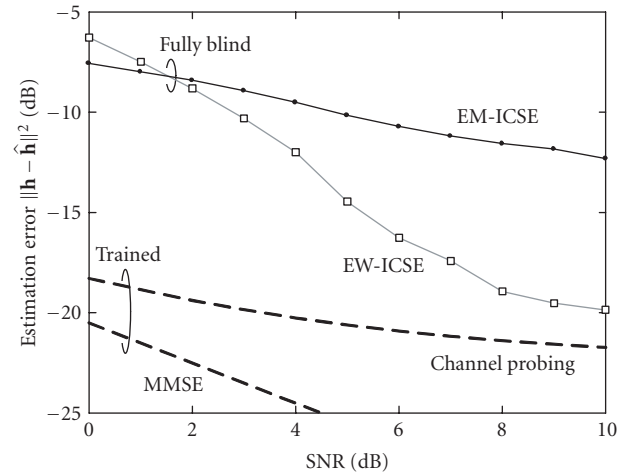


FIGURE 7: Estimation error for the EM-ICSE and EW-ICSE, after 20 iterations. Also shown are the performances of the trained channel probing and trained MMSE estimates.

Unlike the good performance of the EW-ICSE, the EM estimates even fail to converge in the mean to the correct estimates, especially \hat{h}_0 . This happens because the EM-ICSE often gets trapped in local maxima of the likelihood function [16], while the EW-ICSE avoids many of these local maxima. The better convergence behavior of the EW-ICSE is even more clear in Figure 6, where we show the noise variance estimates. Also, Figures 4, 5, and 6 suggest that the EW-ICSE converges faster than the EM-ICSE.

In Figure 7 we show the channel estimation error for the EW-ICSE and the EM-ICSE estimates as a function of SNR, after 20 iterations. The number of iterations is enough for both the EM-ICSE and the EW-ICSE to converge in this case. We also show the estimation errors of the trained MMSE estimates and the trained channel probing estimates. The results are averaged over 100 independent runs of this experiment.

In Figure 8, we show the average BER. Again, as we can see in Figures 7 and 8, the EW-ICSE performs better than the EM-ICSE.

It is interesting to notice in Figures 7 and 8 that for high enough SNR the performance of the EW-ICSE approaches that of its trained counterpart, the channel probing estimator. One might also expect the performance of the EM-ICSE to approach that of its trained counterpart, the MMSE algorithm. However, as we can see from Figures 7 and 8, the EM-ICSE performs worse than channel probing, which is in turn worse than the MMSE estimator. The difference between the EM and MMSE estimates may be explained by the misconvergence of the EM-ICSE.

It should be pointed out that even though the channel estimates provided by the MMSE algorithm are better than the channel probing estimates, the BER of both estimates is

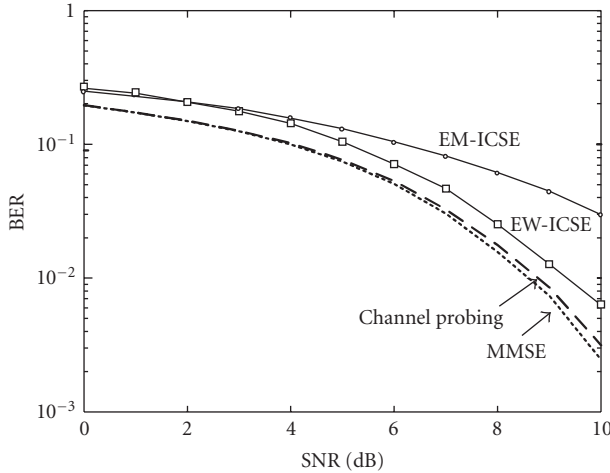


FIGURE 8: Bit error rate versus SNR using EM and EW estimates after 20 iterations. Also shown is the performance resulting from the use of the trained channel probing and trained MMSE estimates.

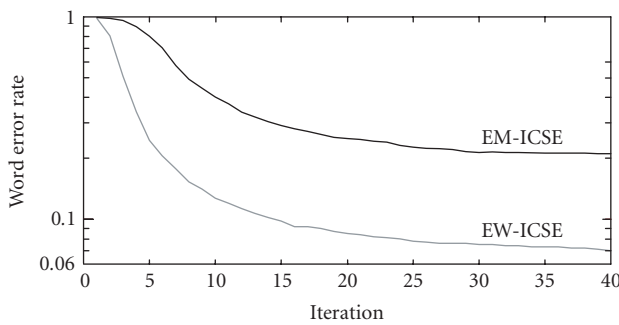


FIGURE 9: WER for the EW-ICSE and the EM-ICSE, averaged over 1000 random channels.

similar. In other words, the channel probing estimates are “good enough,” and the added complexity of the MMSE estimator does not have much impact on the BER performance in the SNR range considered here. Finally, we observed that the BER performance of a BCJR equalizer with channel knowledge cannot be distinguished from that of a BCJR equalizer using the MMSE estimates.

To further support the claim that the EW-ICSE avoids most of the local maxima of the likelihood function that trap the EM-ICSE, we ran both the EM-ICSE and the EW-ICSE on 1000 random channels of memory $\mu = 4$, generated as $\mathbf{h} = \mathbf{u}/\|\mathbf{u}\|$, where $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The estimates were initialized to $\hat{\sigma}_{(0)}^2 = \sum_{k=0}^{N-1} |r_k|^2 / 2N$ and $\hat{\mathbf{h}}^{(0)} = (0, \dots, 0, \hat{\sigma}_{(0)}, 0, \dots, 0)^T$, that is, the center tap of $\hat{\mathbf{h}}^{(0)}$ is initialized to $\hat{\sigma}_{(0)}$. We used SNR = 18 dB, and blocks of $K = 1000$ BPSK symbols. In Figure 9 we show the word error rate (WER) (fraction of blocks detected with errors) of the EW-ICSE and the EM-ICSE versus iteration. It is again clear that the EW-ICSE outperforms the EM-ICSE. It should be noted that in this example the equalizer based on the channel probing estimates was able to detect all transmitted sequences correctly.

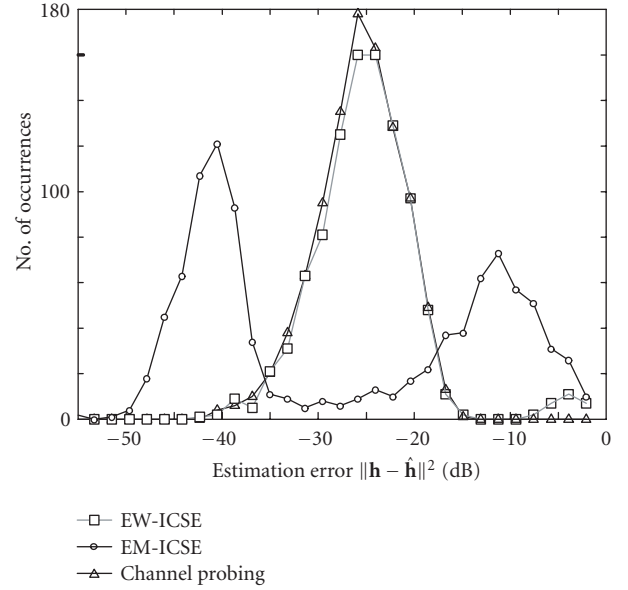


FIGURE 10: Histograms of estimation errors for the EW-ICSE and the EM-ICSE over an ensemble of 1000 random channels.

The better performance of the EW estimates can also be seen in Figure 10, where we show histograms of the estimation errors (in dB) for the channel probing, the EW, and the EM estimates, computed after 50 iterations. We see that while only 3% of the EW estimates have an error larger than -16 dB, 35% of the EM estimates have an error larger than -16 dB. In fact, the histogram for the EW estimates is very similar to that of the channel probing estimates, which again shows the good convergence properties of the EW-ICSE. It is also interesting to note in Figure 10 that the EM estimates have a bimodal behavior: the estimation errors produced by the EM-ICSE are grouped around -11 dB and -43 dB. These groups are respectively better than and worse than the channel probing estimates. This bimodal behavior can be explained by the fact that the EM algorithm often converges to inaccurate estimates, leading to large estimation errors. On the other hand, when the EM algorithm does work, it works very well.

7. CONCLUSIONS

We presented the EW channel estimator, a linear-complexity channel estimator for ICSE. We have shown that this technique can be seen as a modification of the EM channel estimator. A key feature of the EW estimator is its extended window, which greatly improves the convergence behavior of ICSEs based on the EW estimator, avoiding most of the local maxima of the likelihood function that trap the EM-ICSE. Furthermore, the computational complexity of the EW estimator grows linearly with the channel memory, as opposed to the quadratic complexity of the EM channel estimator. Additionally, the EW estimator may be used with any soft-output equalizer. This allows for even further complexity reduction when compared to the EM-ICSE, which requires

a BCJR equalizer. However, simulations show that, despite its good convergence properties, the EW-ICSE is not globally convergent. The problem of devising an iterative strategy that is guaranteed to always avoid misconvergence, regardless of initialization, remains open.

REFERENCES

- [1] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital Communications*, Kluwer Academic Publishers, Norwell, Mass, USA, 3rd edition, 2003.
- [2] J. Ayadi, E. de Carvalho, and D. T. M. Slock, "Blind and semi-blind maximum likelihood methods for FIR multichannel identification," in *Proc. IEEE International Conference on Acoustics, Speech, Signal Processing (ICASSP '98)*, vol. 6, pp. 3185–3188, Seattle, Wash, USA, May 1998.
- [3] M. Feder and J. A. Catipovic, "Algorithms for joint channel estimation and data recovery-application to equalization in underwater communications," *IEEE J. Oceanic Eng.*, vol. 16, no. 1, pp. 42–55, 1991.
- [4] G. K. Kaleh and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Trans. Commun.*, vol. 42, no. 7, pp. 2406–2413, 1994.
- [5] C. Anton-Haro, J. A. R. Fonollosa, and J. R. Fonollosa, "Blind channel estimation and data detection using hidden Markov models," *IEEE Trans. Signal Processing*, vol. 45, no. 1, pp. 241–247, 1997.
- [6] J. Garcia-Frias and J. D. Villasenor, "Combined turbo detection and decoding for unknown ISI channels," *IEEE Trans. Commun.*, vol. 51, no. 1, pp. 79–85, 2003.
- [7] K.-D. Kammeyer, V. Kühn, and T. Petermann, "Blind and nonblind turbo estimation for fast fading GSM channels," *IEEE J. Select. Areas Commun.*, vol. 19, no. 9, pp. 1718–1728, 2001.
- [8] A. O. Berthet, B. S. Ünäl, and R. Visoz, "Iterative decoding of convolutionally encoded signals over multipath Rayleigh fading channels," *IEEE J. Select. Areas Commun.*, vol. 19, no. 9, pp. 1729–1743, 2001.
- [9] R. R. Lopes and J. R. Barry, "Exploiting error-control coding in blind channel estimation," in *IEEE Global Communications Conference (GLOBECOM '01)*, vol. 2, pp. 1317–1321, San Antonio, Tex, USA, November 2001.
- [10] V. Krishnamurthy and J. B. Moore, "On-line estimation of hidden Markov model parameters based on the Kullback-Leibler information measure," *IEEE Trans. Signal Processing*, vol. 41, no. 8, pp. 2557–2573, 1993.
- [11] L. B. White, S. Perreau, and P. Duhamel, "Reduced computation blind equalization for FIR channel input Markov models," in *IEEE International Conference on Communications (ICC '95)*, vol. 2, pp. 993–997, Seattle, Wash, USA, June 1995.
- [12] M. Shao and C. L. Nikias, "An ML/MMSE estimation approach to blind equalization," in *Proc. IEEE International Conference on Acoustics, Speech, Signal Processing (ICASSP '94)*, vol. 4, pp. 569–572, Adelaide, SA, Australia, April 1994.
- [13] H. A. Cirpan and M. K. Tsatsanis, "Stochastic maximum likelihood methods for semi-blind channel estimation," *IEEE Signal Processing Lett.*, vol. 5, no. 1, pp. 21–24, 1998.
- [14] B.-P. Paris, "Self-adaptive maximum-likelihood sequence estimation," in *IEEE Global Communications Conference (GLOBECOM '93)*, vol. 4, pp. 92–96, Houston, Tex, USA, November–December 1993.
- [15] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Annals of Mathematics Statistics*, vol. 41, no. 1, pp. 164–171, 1970.
- [16] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistics Society*, vol. 39, no. 1, pp. 1–38, 1977.
- [17] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. 20, no. 2, pp. 284–287, 1974.
- [18] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: turbo-codes," in *IEEE International Conference on Communications (ICC '93)*, vol. 2, pp. 1064–1070, Geneva, Switzerland, May 1993.
- [19] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenation of interleaved codes: performance analysis, design, and iterative decoding," *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 909–926, 1998.
- [20] M. Tüchler, R. Koetter, and A. C. Singer, "Turbo equalization: principles and new results," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 754–767, 2002.
- [21] R. R. Lopes and J. R. Barry, "Soft-output decision-feedback equalization with a priori information," in *IEEE Global Communications Conference (GLOBECOM '03)*, vol. 3, pp. 1705–1709, San Francisco, Calif, USA, December 2003.
- [22] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, New York, NY, USA, 2nd edition, 1994.
- [23] C. A. Montemayor and P. G. Flikkema, "Near-optimum iterative estimation of dispersive multipath channels," in *IEEE 48th Vehicular Technology Conference (VTC '98)*, vol. 3, pp. 2246–2250, Ottawa, ON, Canada, May 1998.
- [24] M. Sandell, C. Luschi, P. Strauch, and R. Yan, "Iterative channel estimation using soft decision feedback," in *IEEE Global Communications Conference (GLOBECOM '98)*, vol. 6, pp. 3728–3733, Sydney, NSW, Australia, November 1998.

Renato R. Lopes received the B.S. and M.S. degrees from the University of Campinas (UNICAMP), Brazil, in 1995 and 1997, and the Ph.D. degree from the Georgia Institute of Technology, USA, in 2003, all in electrical engineering. He also received an M.A. degree in applied mathematics from the Georgia Institute of Technology, USA, in 2001. During his studies, he was supported by the Brazilian agencies CNPq and CAPES, and held teaching and research assistant positions from 1999 to 2003. He is currently a postdoctoral researcher at UNICAMP, under a grant from FAPESP. His research interests are in the general area of communications theory, including equalization, identification, iterative receivers, and coding theory.



John R. Barry received the B.S. degree in electrical engineering from the State University of New York, Buffalo, in 1986, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1987 and 1992, respectively. Since 1992, he has been with the Georgia Institute of Technology, Atlanta, where he is an Associate Professor in the School of Electrical and Computer Engineering. Currently he is visiting Georgia Tech Lorraine, Metz, France. His research interests include wireless communications, equalization, and multiuser communications. He is a coauthor with E. A. Lee and D. G. Messerschmitt of *Digital Communications*, third edition, Kluwer, Norwell, Mass, 2004, and the author of *Wireless Infrared Communications*, Kluwer, Norwell, Mass, 1994.

