Reduced-Complexity Receiver Structures for Space–Time Bit-Interleaved Coded Modulation Systems

Inkyu Lee, *Senior Member, IEEE*, and Carl-Erik W. Sundberg, *Fellow, IEEE*

***Abstract*—Transmission efficiency in radio channels can be considerably improved by using multiple transmit and receive antennas and employing a family of schemes called space–time (ST) coding. Both extended range and/or improved bandwidth efficiency can be achieved, compared with a radio link with a single transmit and receive antenna. Bit-interleaved coded modulation schemes give diversity gains on fading channels with higher order modulation constellations combined with conventional binary convolutional codes also for the case of a single transmit and receive antenna radio link. In this paper, we study a family of flexible bandwidth-efficient ST coding schemes which combine these two ideas in a narrowband flat-fading channel and single-carrier modems. We address receiver complexity for the case of a large number of transmit antennas and higher order modulation constellations. Especially, we focus on practical configurations, where the number of transmit antennas is greater than that of receive antennas. Simplified receivers using tentative decisions are proposed and evaluated by means of simulations. Tradeoffs between complexity reduction and performance loss are presented. We emphasize systems that are of particular interest in applications where the number of transmit antennas exceeds the number of receive antennas. A system with four transmit antennas with an eight-fold complexity reduction and a performance loss of about 1 dB is demonstrated.**

***Index Terms*—Coded modulation, iterative decoding, multiple antennas, reduced-complexity receiver, space–time codes (STCs).**

I. Introduction

RADIO channels have a number of severe impairments, such as fading, noise, and interference [1]. In a number of applications, increased demand for extended range and/or improved bandwidth efficiency have been expressed for digital systems. In recent years, a variety of space–time codes (STCs) for schemes with multiple transmit and receive antennas have been presented, see, for example, [2]–[7]. These systems exhibit improved performance in terms of range and/or bandwidth efficiency on noisy fading radio channels, compared with single transmit and receive antenna links.

Paper approved by H. El-Gamal, the Editor for Space–Time Coding and Spread Spectrum of the IEEE Communications Society. Manuscript received March 12, 2003; revised December 21, 2004. This work was supported by the Ministry of Information and Communication (MIC), Korea, under the Information Technology Research Center (ITRC) support program, supervised by the Institute of Information Technology Assessment (IITA). This paper was presented in part at the IEEE International Conference on Vehicular Technology, May 2003.

I. Lee is with the School of Electrical Engineering, Korea University, Seoul, Korea (e-mail: inkyu@korea.ac.kr).

C.-E. W. Sundberg is with SundComm, Chatham, NJ 07928 USA (e-mail: cews@ieee.org).

Digital Object Identifier 10.1109/TCOMM.2006.887491

Coded modulation schemes were originally designed to enhance time diversity on single-antenna radio links [8]–[11]. In order to obtain diversity gains with higher order modulation symbols, bit-interleaved coded modulation (BICM) systems [12]–[14], and multilevel coding methods with multistage decoding [15], [16] were designed using binary convolutional codes. For narrowband flat-fading radio channels, ST-BICM schemes were first proposed by Tonello [17], [18] with multiple-antenna systems. For wideband systems, a scheme combining orthogonal frequency-division multiplexing (OFDM) with ST coding based on BICM methods was proposed to obtain a flexible family of systems for wireless LAN applications [19].

In this paper, we study a family of ST-coded systems based on BICM [17]. These are very flexible systems based on conventional binary convolutional codes, which we earlier introduced and studied for multicarrier systems in [19] and [20]. In contrast to many other ST-coded systems, changes in rates are relatively simple in these bit-interleaved systems. A modification in the binary convolutional code and the corresponding receiver trellis is what is needed. Different number of transmit antennas can also be incorporated without substantial modifications. Unlike other ST-coding schemes such as the ST trellis codes (STTCs) proposed in [2], where each code trellis for a given configuration was handcrafted, the ST-BICM system design is quite flexible. Also, high modulation levels are not suitable for the STTC in [2], whereas the trellis design for the ST-BICM is independent of the constellation size. Furthermore, it has been shown in [17] that the ST-BICM outperforms the STTC scheme for many different system configurations.

The near-optimum receiver consists of a multiple-input multiple-output (MIMO) demapper followed by a maximum *a posteriori* (MAP) decoder. Decoding is performed by iterating between the demapper and the decoder. This is referred to as a decoder using the turbo principle [21], and can also be considered as a serially concatenated coded system configuration [22], [23], where the demapper block and the convolutional decoder are connected by a deinterleaver. The MIMO demapper complexity is closely related to the number of bits/constellation points in the modulation mapper as well as the number of transmit antennas, and the MAP decoder complexity is related to the memory (number of states) in the binary convolutional code [19].

Recently, some architectures have been proposed for reducing the receiver complexity in several MIMO structures. In uncoded systems, a layered ST architecture known as vertical Bell Labs

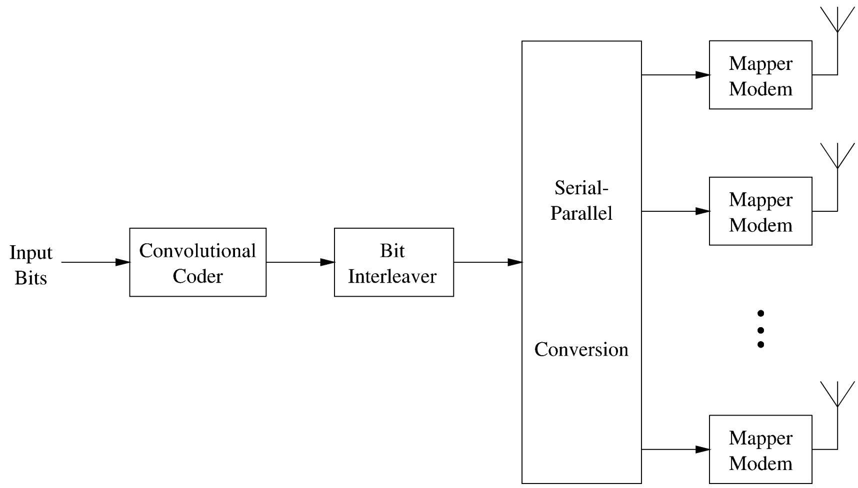


Fig. 1. ST-BICM transmitter structure.

layered space–time (V-BLAST) [24] achieves a high data rate using multiple transmit antennas based on a nulling/cancelling algorithm. To reduce the complexity of the ST-BICM system, the list-sphere decoder [25] or the tree-search algorithm [26] have been applied in the demapper block. Unfortunately, all of these approaches require the MIMO system to have a larger number of receive antennas than transmit antennas. In many wireless applications, it is more desirable for a mobile terminal to have a smaller number of antennas than a base station or an access point. As the downlink throughput is of interest in most cases, this naturally favors the system configuration with the number of receive antennas being smaller than the number of transmit antennas. For the cases where there are more transmit antennas than receive antennas, the sphere decoding algorithm (SDA) [27] and the probabilistic data association (PDA) method [28] have been proposed to achieve near-maximum-likelihood (ML) performance. However, these schemes still cannot be applied to the ST-BICM, since these algorithms generate only hard decisions. Also, the complexity of the SDA grows substantially when there are more transmit antennas than receive antennas [29]. In this paper, we focus on a simplified receiver structure which can be applied to systems equipped with more transmit antennas than receive antennas, where most other reduced-complexity approaches cannot be employed.

It has been observed by simulations [19] that an iterative decoding scheme is crucial where the number of transmit antennas exceeds that of receive antennas. When the number of receive antennas is at least the same as that of transmit antennas, reasonable performance can be achieved without the iterative decoding scheme. In that case, the overall receiver complexity could be lower, as there is no need to process soft information from the previous iteration. However, in the system configuration we are interested in, the receiver complexity issue becomes more important, as an iterative decoding structure is now required.

In this paper, we present a low-complexity receiver structure which can reduce the complexity significantly in the demapper block with an aid of simple minimum mean-square error (MMSE) estimations and tentative decisions. We demonstrate that the associated loss in performance is small. Furthermore, the proposed reduced-complexity receiver design does not put any restrictions on the number of transmit/receive antennas, unlike other reduced-complexity approaches. Therefore, the proposed scheme can be applied to practical systems where the number of receive antennas is smaller than the number of transmit antennas. An MMSE-based approach was presented in [30] in the context of iterative decoding of coded code-division multiple-access (CDMA) systems. Our approach differs from this in that we are using tentative decisions to improve the demapper accuracy.

A large number of different system configurations are evaluated for reduced-complexity systems and compared with their full-complexity counterparts using computer simulations. We demonstrate a system with four transmit antennas which have an eight-fold complexity reduction in the demapper block at a loss of about 1 dB in performance. All results in this paper are given for the narrowband flat-fading case. Such issues as optimization of the binary convolutional code in these systems, as well as mapper choices, are not dealt with in this paper. The results can be generalized to the wideband case with frequency-selective fading using OFDM [19]. The systems and the reduced-complexity receivers can also, in principle, be used for other channels, such as fast fading.

The outline of the paper is as follows. After the introduction, the system model is presented in Section II. In Section III, we discuss the diversity order for the ST-BICM scheme. The simplified receiver structures are presented in Section IV. Performance evaluation by system simulations are given in Section V. Finally, the paper is terminated by a discussion and conclusions section.

II. System Model

Let N_t and N_r be the number of transmit and receive antennas, respectively, assuming $N_t \geq 2$ and $N_r \geq 1$. Fig. 1 shows a typical transmitter structure of the ST-BICM system considered in this paper. The modem constellations used are M-ary phase-shift keying (M-PSK) or M-ary quadrature amplitude modulation (M-QAM), with M being the number of

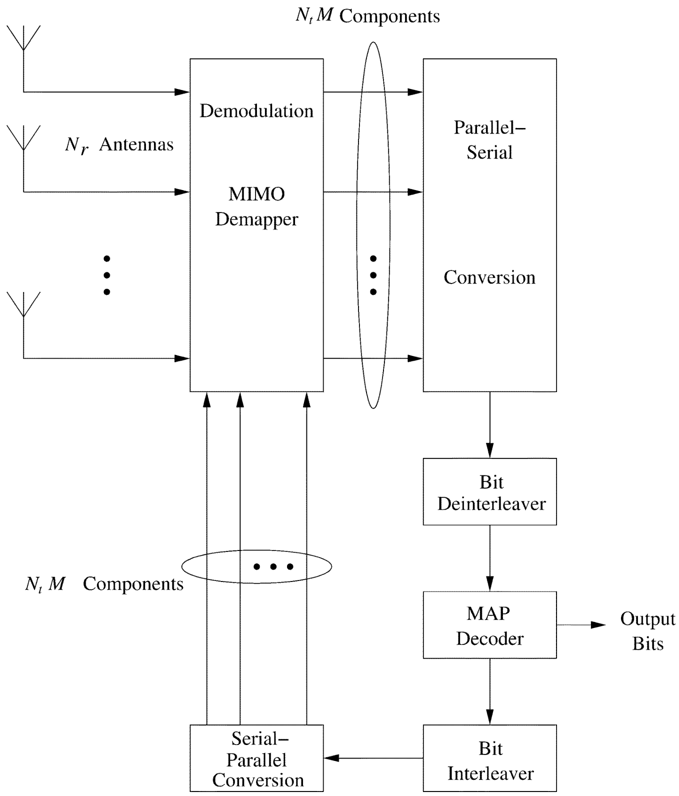


Fig. 2. ST-BICM receiver structure.

symbols in the constellation. Here we primarily consider narrowband systems. Wideband channels could be considered by using OFDM rather than single-carrier modems. The spectral efficiency of the system is [17] $R_T = R_C \cdot N_t \cdot M_b$ b/s/Hz, where R_C is the rate of the convolutional code used, and M_b is defined as $\log_2 M$.

Fig. 2 shows a typical receiver structure. The MIMO demapper and the MAP decoder for the convolutional code are the main building blocks. The near-optimum decoder performs iterative decoding. For detailed illustration of the structures, we refer to [17], [19], and [25].

The received signal at the kth time slot from the jth receive antenna is represented by

$$y_k^j = \sum_{i=1}^{N_t} h_k^{i,j} x_k^i + n_k^j, \quad \text{for } j = 1, 2, \ldots, N_r \tag{1}$$

where x_k^i is the transmitted symbol at the ith transmit antenna at the kth time slot. The signal has the symbol energy E_s. The channel coefficient $h_k^{i,j}$ is the equivalent channel response of the link between the ith transmit antenna and jth receive antenna at the kth time slot, and is assumed to be complex Gaussian with unit variance and zero mean (Rayleigh fading). Finally, in (1), n_k^j is a sequence of independent, identically distributed (i.i.d.) complex zero-mean Gaussian variables with variance $N_0/2$ per dimension.

In this paper, we will focus on the block-fading channel model [31], where fading coefficients are static over a block of transmitted symbols, and independent over blocks.

Denoting N_s as the size of the packet, (1) can be written as

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{n}_k, \quad \text{for } k = 1, 2, \ldots, N_s$$

where we define $\mathbf{y}_k = [y_k^1 \ldots y_k^{N_r}]^T, \mathbf{x}_k = [x_k^1 \ldots x_k^{N_t}]^T, \mathbf{n}_k = [n_k^1 \ldots n_k^{N_r}]^T$ and

$$\mathbf{H}_k = \begin{bmatrix} h_k^{1,1} & \cdots & h_k^{N_t,1} \\ \vdots & \ddots & \vdots \\ h_k^{1,N_r} & \cdots & h_k^{N_t,N_r} \end{bmatrix}.$$

III. Diversity Order for ST-BICM

For ST-BICM systems under the exact feedback assumption in the block-fading narrowband flat-fading channel, it has been derived that the pairwise error probability (PEP) at high signal-to-noise ratio (SNR) is upper bounded by [17], [32]

$$P(\mathbf{x} \to \hat{\mathbf{x}}) \le \left(\prod_{i=1}^{N_t} \frac{E_s}{4N_0} d_E^i(\mathbf{x}, \hat{\mathbf{x}}) \right)^{-N_r}$$

where we define $d_E^i(\mathbf{x}, \hat{\mathbf{x}})$ as the sum of the squared Euclidean distances computed on the subsequences transmitted over the ith antenna.

It is clear from the above bound that the maximum diversity order for the ST-BICM systems in block-fading channels is equal to $N_t N_r$. Note that this will only be achieved if the Hamming distances associated with all transmit antennas are nonzero [17]. It is worth noting that improved diversity can be achieved by exploiting the frequency domain (OFDM) and/or time domain (longer interleavers) [1], [19].

Let us denote L_d as the number of instances $1 \le i \le N_t$ such that $d_E^i(\mathbf{x}, \hat{\mathbf{x}})$ is nonzero. Then, the achievable diversity order in the ST-BICM is equal to $L_d N_r$. For example, if any of the transmit-antenna subsequences have an all-zero Hamming distance, the achieved degree of diversity for the system is $L_d N_r$, which is less than the full diversity $N_t N_r$. Therefore, the condition to achieve full diversity for the ST-BICM is as follows [32]:

$$d_E^i(\mathbf{x}, \hat{\mathbf{x}}) \neq 0 \quad \text{for } i = 1, 2, \ldots, N_t. \tag{2}$$

For a given spectral efficiency, one may use smaller constellations in conjunction with high-rate codes. By using smaller signal constellations at a given N_t, lower complexity receivers are achieved, as the demapper complexity depends on the constellation size. Normally, good high-rate codes are obtained by applying a puncturing mechanism to low-rate codes [33]. However, it has been shown in [32] that the punctured code under certain conditions fails to achieve the full diversity for the ST-BICM systems. In other words, the diversity order L_d achieved by the ST-BICM scheme with punctured codes is less than the maximum achievable N_t.

This is somewhat unfortunate for the design of the ST-BICM system, as the choice of higher modulation is the only option left for achieving a higher spectral efficiency. The higher modulation level, in turn, increases the demapper complexity, since the number of candidates to search grows exponentially with the modulation level. Thus, this contributes to motivating us to look for a simplified receiver structure, especially for the demapper block, as the MAP decoder block complexity is independent of the modulation level.

IV. Reduced-Complexity Receiver Structure

First, we illustrate the optimum demapper operation which computes the log-likelihood ratio (LLR) for the individual bit. Let b_k^n be the bit that is mapped into the nth bit position ($n = 1, 2, \ldots, N_t M_b$) in the input symbol vector $\mathbf{x}_k$. For example, the ith input symbol x_k^i is mapped by the input bits $b_k^{(i-1)M_b+1}, b_k^{(i-1)M_b+2}, \ldots, b_k^{iM_b}$. We denote $L(b_k^n)$ as the LLR value for the bit b_k^n. To distinguish the LLR values from the demapper and the decoder, $L_m(b_k^n)$ and $L_c(b_k^n)$ are used, respectively. Also, $L^e(b_k^n)$ is defined as extrinsic information LLR values obtained by subtracting the input LLR values from the output $L(b_k^n)$.

Let the set $\mathcal{S}_d^n$, $d = \pm 1$, be a set of all symbol vectors with a $+1$ or -1 value of bit b_k^n, respectively. Denoting $\mathbf{b}_k$ and $\mathbf{L}_k^e$ as column vectors comprised of b_k^n and the extrinsic LLR values from the MAP decoder $L_c^e(b_k^n)$, the LLR values for the demapper are given by [17], [19], [25]

$$\begin{aligned} L_m(b_k^n) &= \log \frac{P\left(b_k^n = +1|\mathbf{y}_k, \mathbf{H}_k\right)}{P\left(b_k^n = -1|\mathbf{y}_k, \mathbf{H}_k\right)} \\ &= \log \frac{\sum_{\mathbf{x}_k \in \mathcal{S}_{+1}^n} p(\mathbf{x}_k, \mathbf{y}_k, \mathbf{H}_k)}{\sum_{\mathbf{x}_k \in \mathcal{S}_{-1}^n} p(\mathbf{x}_k, \mathbf{y}_k, \mathbf{H}_k)} \end{aligned} \tag{3}$$

where, after normalization, we use the joint probability density

$$p(\mathbf{x}_k, \mathbf{y}_k, \mathbf{H}_k) = \exp\left(-\frac{1}{N_0}\|\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k\|^2 + \frac{1}{2}\mathbf{b}_k^T\mathbf{L}_k^e\right).$$

Note that the total number of candidates to search in (3) is $2^{N_t \cdot M_b}$. Obviously, the complexity of the optimum demapper increases exponentially with the number of transmit antennas and the modulation level. In this section, we propose a new scheme to simplify the computation in the demapper block which dominates the complexity of the ST-BICM receiver.

Basically, in order to reduce the size of the search set, we use tentative decisions. Suppose that reliable tentative decisions are available for P bit positions $n_1, n_2, \ldots, n_p$, and let $\hat{b}_k^{n_1}, \hat{b}_k^{n_2}, \ldots, \hat{b}_k^{n_p}$ be the hard decisions for the corresponding bits. Then, define $\mathcal{L}^P$ as a set of all symbol vectors with bits in those positions equal to the tentative decisions

$$\mathcal{L}^P = \left\{\mathbf{x}_k \,|b_k^{n_1} = \hat{b}_k^{n_1}, b_k^{n_2} = \hat{b}_k^{n_2}, \ldots, b_k^{n_p} = \hat{b}_k^{n_p}\right\}.$$

Now we carry out the candidate search in (3) over a reduced set $\mathcal{L}^P \cap \mathcal{S}_d^n$ instead of $\mathcal{S}_d^n$. As a result, the size of the set which the demapper needs to search over decreases to $2^{N_t M_b - P}$. The tentative decisions can be obtained from the LLR values computed from the previous iteration. However, at the first iteration, *a priori* information for the received signal is not yet available at the demapper. Thus, we rely on the MMSE estimation. It turns out that the choice of reliable tentative decisions in the first iteration is crucial for the performance of the iterative decoder. Thus, we start illustrating the MMSE-estimation scheme for the first iteration.

A. MMSE Estimation at the First Iteration

The goal here is to generate the tentative decisions and to provide a figure of merit to identify which decisions are most reliable. Then, we compute the LLR values over a set where those bit positions are fixed. Since the demapper at the first iteration does not have any *a priori* information, we apply an MMSE-estimation filter $\mathbf{w}_{k,i}$ to the received signal $\mathbf{y}_k$ to get an estimate of the transmitted symbol x_k^i. Thus, we need to minimize the estimation error, defined as

$$e_k^i = x_k^i - \mathbf{w}_{k,i}^* \mathbf{y}_k$$

where $\mathbf{w}_{k,i}$ is a column vector of length N_r which estimates x_k^i.

Let $\mathbf{e}_k$ be a column vector of length N_t, comprised of the estimation error e_k^i. By combining the column vectors $\mathbf{w}_{k,i}$ into an N_r by N_t matrix, it is straightforward to find the optimum estimation matrix $\mathbf{W}_k$ which minimizes the mean-square error (MSE) $E(\|\mathbf{e}_k\|^2)$ as

$$\mathbf{W}_k = (\mathbf{H}_k\mathbf{H}_k^* + \alpha\mathbf{I})^{-1}\mathbf{H}_k \tag{4}$$

where α is defined as $\alpha = 1/\text{SNR} = N_0/E_S$.

Note that in case of $N_r = 1$, the optimum estimate filter matrix becomes equivalent to the scaled channel matrix as

$$\mathbf{W}_k = \frac{1}{\|\mathbf{H}_k\|^2 + \alpha}\mathbf{H}_k.$$

Also, using the matrix inversion lemma[1] [34], (4) can be written as

$$\mathbf{W}_k = \mathbf{H}_k(\mathbf{H}_k^*\mathbf{H}_k + \alpha\mathbf{I})^{-1}.$$

Then, it follows that the MMSE-estimate vector of $\mathbf{x}_k$ is obtained by

$$\hat{\mathbf{x}}_k = \mathbf{W}_k^*\mathbf{y}_k = (\mathbf{H}_k^*\mathbf{H}_k + \alpha\mathbf{I})^{-1}\mathbf{H}_k^*\mathbf{y}_k. \tag{5}$$

After some matrix manipulations, the covariance matrix of the estimation error $\mathbf{e}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$ is expressed by

$$\mathbf{R}_e = E(\mathbf{e}_k\mathbf{e}_k^*) = N_0(\mathbf{H}_k^*\mathbf{H}_k + \alpha\mathbf{I})^{-1}. \tag{6}$$

Denoting σ_i^2 as the MSE for $\hat{x}_k^i$, σ_i^2 is the ith diagonal element of the covariance matrix $\mathbf{R}_e$. Note that only one matrix inversion for $(\mathbf{H}_k^*\mathbf{H}_k + \alpha\mathbf{I})^{-1}$ is required to compute both the MMSE-estimation filter $\mathbf{W}_k$ and the covariance matrix $\mathbf{R}_e$.

Now we can generate the tentative decisions for the transmitted bits based on the MMSE estimate. Instead of simply slicing the MMSE estimate $\hat{\mathbf{x}}_k$, we proceed further to get a better decision. Denoting $\hat{L}(b_k^n)$ as the soft-output value based on these MMSE estimates $\hat{\mathbf{x}}_k$, the soft output for each bit can be computed as

$$\hat{L}(b_k^n) = \log \frac{P\left(b_k^n = +1|\hat{\mathbf{x}}_k\right)}{P\left(b_k^n = -1|\hat{\mathbf{x}}_k\right)} = \log \frac{\sum_{\mathbf{x}_k \in \mathcal{S}_{+1}^n} p(\hat{\mathbf{x}}_k|\mathbf{x}_k)}{\sum_{\mathbf{x}_k \in \mathcal{S}_{-1}^n} p(\hat{\mathbf{x}}_k|\mathbf{x}_k)}.$$

[1] $(\mathbf{A} + \mathbf{BC})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{CA}^{-1}\mathbf{B} + \mathbf{I})^{-1}\mathbf{CA}^{-1}$

Here the joint probability $p(\hat{\mathbf{x}}_k|\mathbf{x}_k)$ can be approximated by a complex Gaussian distribution. Thus, by applying the Gaussian probability density function, and after normalization, the joint probability is obtained by

$$p(\hat{\mathbf{x}}_k|\mathbf{x}_k) = \exp(-(\hat{\mathbf{x}}_k - \mathbf{x}_k)^* \mathbf{R}_e^{-1}(\hat{\mathbf{x}}_k - \mathbf{x}_k)).$$

To simplify the computation, we further neglect off-diagonal terms in the correlation matrix $\mathbf{R}_e$. Then, $\mathbf{R}_e$ is approximated to a diagonal matrix, with σ_i^2 being the ith diagonal element of the matrix. Thus, the computation for the above expression reduces to

$$p(\hat{\mathbf{x}}_k|\mathbf{x}_k) = \prod_{i=1}^{N_t} \exp\left(-\frac{1}{\sigma_i^2}\left|\hat{x}_k^i - x_k^i\right|^2\right).$$

Now the soft-output computation can be carried out in the form of a single scalar symbol instead of a vector. For $n = (i-1)M_b + 1, (i-1)M_b + 2, \ldots, iM_b$, we use a notation $b_k^{n|i}$ to indicate that the nth bit position falls into the ith symbol. After cancelling out the symbols which are not related to $b_k^{n|i}$, the soft-output computation now becomes

$$\hat{L}\left(b_k^{n|i}\right) = \log \frac{\sum_{x_k^i \in \mathcal{M}_{+1}^{n|i}} \exp\left(-\frac{1}{\sigma_i^2}\left|\hat{x}_k^i - x_k^i\right|^2\right)}{\sum_{x_k^i \in \mathcal{M}_{-1}^{n|i}} \exp\left(-\frac{1}{\sigma_i^2}\left|\hat{x}_k^i - x_k^i\right|^2\right)} \quad (7)$$

where the set $\mathcal{M}_d^{n|i}$ is defined as $\mathcal{M}_d^{n|i} = \{x_k^i | b_k^{n|i} = d\}$.

Note that the soft-output computation in (7) can be carried out simply over $M = 2^{M_b}$ candidates. By applying the max-log approximation [35], [36], we can further simplify the computation of the soft-output values. Given the MMSE estimate $\hat{\mathbf{x}}_k$, the simplified scheme transforms the $M = 2^{M_b}$ search in (7) into a simpler M_b boundary problem with very little performance loss.

The soft-output values for all the transmitted bits obtained by the method illustrated above can be directly passed onto the MAP decoder when the number of transmit antennas is at least the same as that of receive antennas. However, when the condition on the antenna numbers is not met, the overall performance of such a system using the MMSE estimates becomes unacceptable, as the accuracy of the estimates is degraded. Note that even when more complicated estimation methods without approximation are applied, they do not work well when there are more transmit than receive antennas, which is the case this paper is focused on. Thus, instead we use the MMSE-estimation-based soft-output computation described above only for determining the most reliable tentative decisions.

Now all the soft outputs for $N_t M_b$ bits are available using (7). Instead of passing them all onto the decoder block, we only choose the P most reliable tentative decisions. Here we will explain how to identify those P bit decisions. Obviously, if incorrect hard decisions are used in determining a candidate set $\mathcal{L}^P$, the overall performance degrades due to error propagation in the demapper. Thus, the accuracy of the tentative decisions is essential for the performance. To this end, we now address the criteria for choosing the most reliable decisions, and derive the probability of bit-decision error.

Considering that the hard decisions $\hat{b}_k^n$ are obtained by slicing the estimate $\hat{x}_k^i$, the probability of decision error in $\hat{b}_k^n$ is dependent on the probability of estimation error in $\hat{x}_k^i$. Therefore, the probability that the hard decision $\hat{b}_k^n$ is correct is equal to

$$P_c = (1 - P_e) \cdot P\left\{b_k^n = \hat{b}_k^n\right\}$$

where P_e is the probability that the MMSE estimate $\hat{x}_k^i$ results in an incorrect decision. Using the $Q(\cdot)$ function [1], [37], P_e is computed by

$$P_e \approx N_e Q\left(\frac{d_{\min}}{2\sigma_i}\right)$$

where N_e represents the average number of nearest neighbors, and $d_{\min}$ stands for the minimum distance for a constellation. Both N_e and $d_{\min}$ can be determined from the given constellation. For example, N_e and $d_{\min}$ for 16-QAM with $E_s = 10$ are 3 and 2, respectively.

Finally, using the relationship between $\hat{L}(b_k^n)$ and $P\{b_k^n = \hat{b}_k^n\}$, the figure of merit which we will use for identifying the most reliable decisions is given by

$$\left(1 - N_e Q\left(\frac{d_{\min}}{2\sigma_i}\right)\right) \cdot \frac{1}{1 + e^{-|\hat{L}(b_k^n)|}} \quad (8)$$

where the computation of $Q(\cdot)$ can be carried out using an accurate approximation in [37].

In actual implementations, the figure of merit (8) can be obtained by a simple lookup table. Note that the MSE for the MMSE estimate σ_i^2 is only related to P_e. If the P_e term is neglected, then the criteria of the bit-decision reliability is simplified to the magnitude of the soft output $|\hat{L}(b_k^n)|$ at the expense of performance degradation. Also note that the soft-output computation in (7) is used only for computing the figure of merit in (8), and not used directly at all.

To determine the P most reliable bit positions, we compute the probability of correct decision for each bit position from (8), and choose the P positions with the largest P_c. After identifying the P tentative decisions, we finally return to (3) to compute the LLR values over the set $\mathcal{L}^P \cap \mathcal{S}_d^n$. Several approximations have been made to arrive at the probability of correct decision P_c in (8). However, since only a small number of hard decisions from the MMSE estimation are involved in the final LLR computation, the overall performance is relatively immune to the accuracy of the approximations.

Fig. 3 summarizes the LLR-value-generation scheme described above for the first iteration. For each ith transmitted symbol, we first obtain the MMSE estimate $\hat{x}_k^i$ from (5), and the corresponding MSE σ_i^2 from (6). Based on these values, we then compute the soft-output values $\hat{L}(b_k^{n|i})$ for the transmitted bits from (7). Now we can compute the figure of merit for the decision reliability in (8). After picking the P positions with the largest P_c, those bit positions are fixed with the tentative decisions. Finally, the LLR values for every bit position except those fixed positions are computed in (3) over a set with a reduced size.

With the practical number of transmit antennas $(N_t \leq 4)$, the MMSE computation in (5) is simple. Therefore, compared with the overall demapper complexity, the computation complexity related to the MMSE estimation in the proposed scheme is quite small.

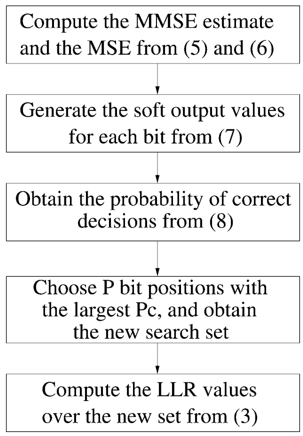


Fig. 3. LLR value generation for the first iteration.

It should be emphasized again that the main focus of this paper is practical system configurations where there are more transmit than receive antennas. Therefore, other schemes which were proposed for systems where the number of receive antennas is at least equal to that of transmit antennas (for example, the single-input single-output (SISO) system approach in [38]) cannot be applied.

B. Complexity Reduction in Subsequent Iterations

Once the demapper generates the LLR values using the technique described in the previous section at the first iteration, the MAP decoder can produce the LLR values for the transmitted bits for the subsequent iterations. Therefore, it is straightforward to apply the same principle in the demapper block in the subsequent iterations to reduce the size of a set of candidates to search.

After the demapper computes the LLR values, only the extrinsic part of the LLR values are transferred to the MAP decoder through a deinterleaver, following the turbo principle. Then, the MAP decoder updates the LLR values $L_c(b_k^n)$ for the transmitted bit b_k^n. Also, after subtracting the input LLR values, the extrinsic information LLR values $L_c^e(b_k^n)$ are sent back to the demapper through an interleaver. Even though the demapper block updates the LLR computation using the extrinsic information $L_c^e(b_k^n)$, the overall LLR value $L_c(b_k^n)$ should be used to determine the P most reliable bit positions, since the magnitude of the LLR value $|L_c(b_k^n)|$ indicates the decision reliability. Therefore, in the second and following iterations, the demapper chooses those P bit positions with the largest LLR values $|L_c(b_k^n)|$, and then at each iteration, a set $\mathcal{L}^P$ is determined based on those fixed P bit positions. Finally, the LLR values for bit positions except for those P fixed positions are computed from (3) over the set $\mathcal{L}^p \cap \mathcal{S}_d^n$.

It has been observed [39] that the logarithm of a summation of exponential terms is well-approximated by taking only the maximum value of the exponent. Therefore, the LLR values for the bits except those P fixed positions can be determined by

$$L_m\left(b_k^n\right) = \max_{\mathbf{x}_k \in \mathcal{L}^p \cap \mathcal{S}_{+1}^n} \left(-\frac{1}{N_0} \|\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k\|^2 + \frac{1}{2} \mathbf{b}_k^T \mathbf{L}_k^e \right) - \max_{\mathbf{x}_k \in \mathcal{L}^p \cap \mathcal{S}_{-1}^n} \left(-\frac{1}{N_0} \|\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k\|^2 + \frac{1}{2} \mathbf{b}_k^T \mathbf{L}_k^e \right).$$

Note that for those P fixed positions, the LLR values computed at the decoder $L_c(b_k^n)$ are used again as the LLR values for the demapper. If any of the hard decisions $\hat{b}_k^n$ are incorrect, the LLR computation becomes inaccurate, and eventually, the magnitude of the LLR values gets smaller. It is readily checked by noting that the first term $-(1/N_0)\|\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k\|^2$ inside the maximization represents the Euclidean distance, and that the second term $(1/2)\mathbf{b}_k^T \mathbf{L}_k^e$ stands for the level of the reliability. In the presence of incorrect decisions, both terms decrease.

Thus, in general, any incorrect decisions result in the decrease in the magnitude of the LLR value. Simulation results suggest that a better performance is obtained by scaling up the LLR values, except for the P fixed bit positions. Therefore, we now maximize over $-(\gamma/N_0)\|\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k\|^2 + (\gamma/2)\mathbf{b}_k^T \mathbf{L}_k^e$, where γ is a scaling factor which can be optimized through simulations. Typical values of γ range from 1.2 to 2. It is found that, in general, the higher the modulation level is, the greater γ becomes.

The maximum possible value for P is $N_t M_b$, which corresponds to the hard-decoding method presented in [18]. However, in that work, the hard-decoding processing was carried out in the decoder part only for the second and following iterations. The complexity associated with the demapper part, which is much greater than the decoder block complexity, was not reduced for the first iteration.

Even though the low-complexity receiver architecture is illustrated in this section assuming a quasi-static channel, obviously the proposed technique is not limited to that case. However, it should be noted that in other channel conditions, such as fast-fading channels or frequency-selective channels, a system based on a puncturing scheme and smaller constellations would be more favorable in terms of the receiver complexity.

V. Simulations

In this section, we present the simulation results for the reduced-complexity receiver scheme illustrated in the previous section. Several different system configurations, including the number of antennas and the spectral efficiency, are considered for the simulations. Since only nonpuncturing systems are considered, the coding rate is chosen as $R_C = 1/N_t$. As a result, the modulation level becomes 2^L, where the spectral efficiency is given as L. Also, the scaling factor γ is fixed to 1.5 throughout the simulations.

First, Fig. 4 presents the simulation result for the two-by-one $(N_t = 2, N_r = 1)$ system with 192 information bits. This short length was chosen because of latency requirements. The y axis exhibits the frame-error rate (FER). If not specified otherwise, optimum free-distance 16-state convolutional codes are employed throughout the simulation section. Also, a randomly generated interleaver of length $N_t N_s \log_2 M$ is used. Longer interleavers are expected to yield somewhat improved performance. A Gray mapping is employed in all simulation cases. In Fig. 4, the curves with $P = 0$ indicate the performance of the optimum iterative decoder. As shown in this plot, a suboptimum scheme with one tentative decision $(P = 1)$ loses about 2 dB, compared with the optimum system at 1% FER.

The performance degradation of the suboptimum system is getting smaller when the number of receive antennas becomes comparable to the transmit antenna number, as we expected.

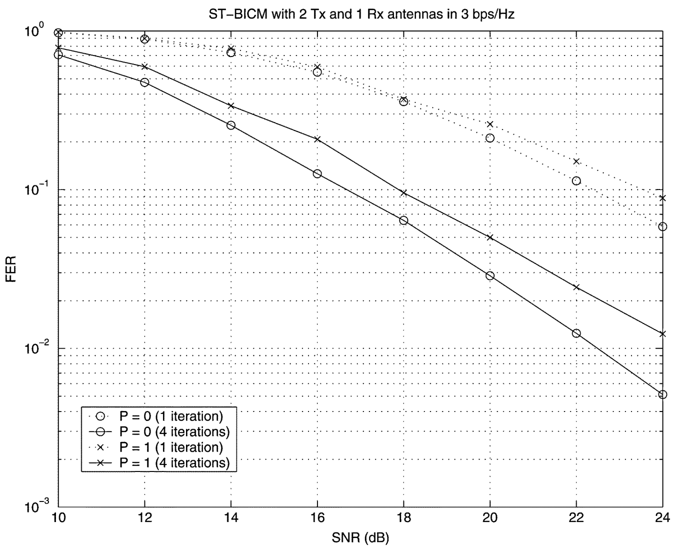


Fig. 4. Comparison of reduced-complexity system.

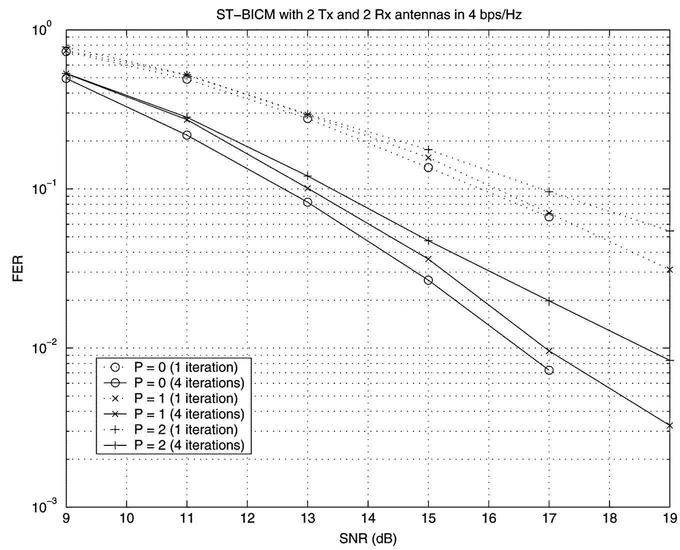


Fig. 6. 2×2 ST-BICM system with 4 bps/Hz.

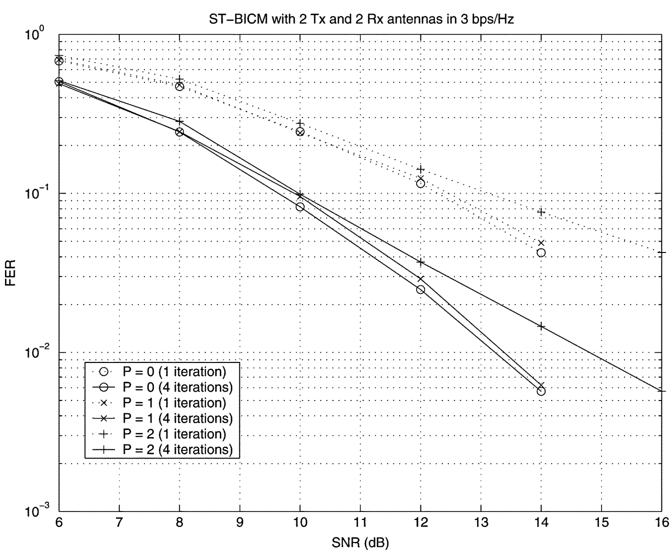


Fig. 5. 2×2 ST-BICM system with 3 b/s/Hz.

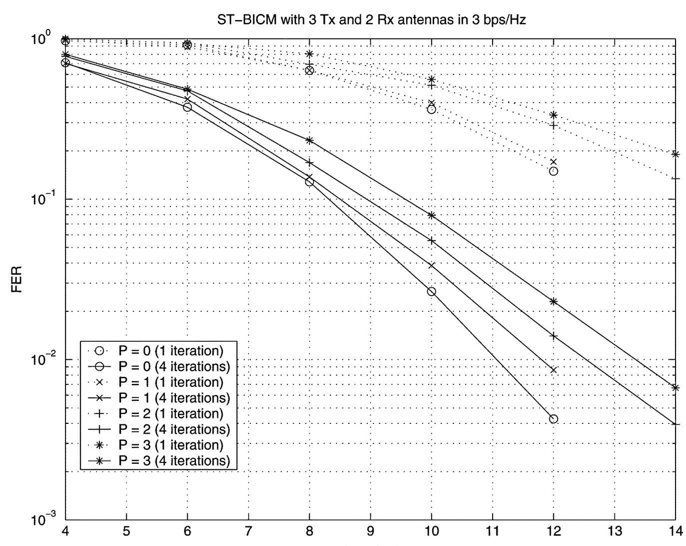


Fig. 7. 3×2 ST-BICM system with 3 b/s/Hz.

Figs. 5 and 6 show the simulation results for the 2-by-2 system with 3 and 4 b/s/Hz, respectively. For both cases, the suboptimum schemes with $P = 1$ are just a few tenths of a decibel away from the optimum case, and with $P = 2$, the performance loss is less than 2 dB. It is also noted that as the modulation level increases, the performance of the suboptimum scheme degrades. For higher constellation sizes, the Euclidean distance in the constellation points becomes smaller, and this makes the demapper performance less effective.

Fig. 7 shows the 3-by-2 ST-BICM performance in 3 b/s/Hz for the $P = 0, 1, 2, 3$ cases. It exhibits a smooth tradeoff between the receiver complexity and the performance. For $P = 1, 2, 3$, the size of the candidate set is reduced by a factor of $1/2, 1/4, 1/8$, respectively. With $P = 3$, the total number of candidates to search in the demapper block is reduced from 512 to 64. Even with this significant reduction in the candidate number, the performance is just about 2 dB away from the optimum case.

Finally, the 4-by-2 ST-BICM with 2 b/s/Hz case is presented in Fig. 8. In this case, we use a four-state convolutional code with code polynomials of $[5, 5, 7, 7]$. In this simulation case, the performance gap between the optimum scheme and the proposed scheme is even smaller. Again, with the reduction of the candidate number by $1/8$ in the $P = 3$ case, the proposed suboptimum case is only 1 dB away from the optimum performance.

From the simulation results presented in this section, it is shown that the proposed reduced-complexity scheme works better with a small constellation size. It should be emphasized again that the previous approaches for reducing the receiver complexity [24]–[26] require the number of transmit antennas to be less than or equal to the number of receive antennas. Thus, these schemes do not work for the case where $N_t > N_r$. Further improvements are expected by optimizing γ.

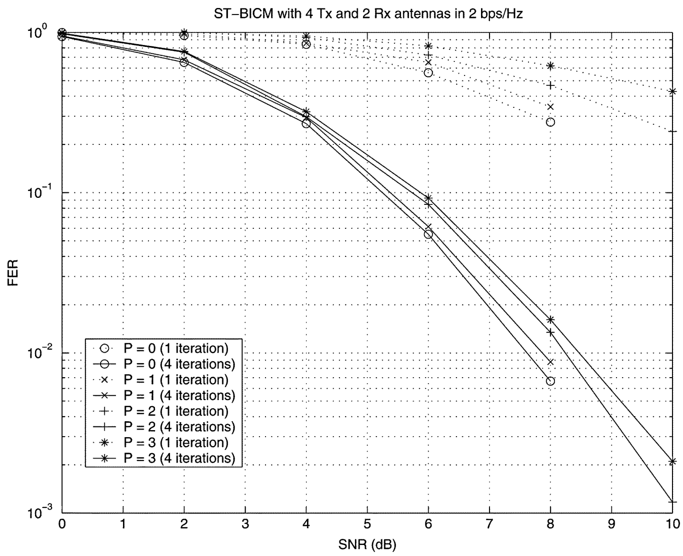


Fig. 8. 4×2 ST-BICM system with 2 b/s/Hz.

VI. Discussion and Conclusions

This paper studies a flexible class of STCs based on BICM systems. We address the receiver-complexity issue, which is a key problem with STCs based on BICM systems with a near-optimum receiver. The MIMO demapper complexity grows exponentially with the number of transmit antennas and with the number of bits per constellation point. Thus, for attractive systems with a large number of transmit antennas and high-bandwidth-efficiency constellation sizes, complexity could be a severe problem. In this paper, we have introduced and studied approaches to simplified suboptimum receivers. We emphasize that unlike prior work, the strategy given in this paper does not require more receive than transmit antennas. We have demonstrated reduced-complexity systems with an eight-fold reduction in complexity and (depending on modulation constellation size) a performance loss of 1–2 dB.

In this paper, perfect knowledge of the channel state information at the receiver side is assumed. The impact of the imperfect channel estimation for the STCs was addressed in [40].

For simplicity, we only considered and simulated the suboptimum reduced-complexity receivers for narrowband flat-fading cases. Generalizations to wideband systems with frequency-selective fading should be straightforward using combinations with OFDM.

Acknowledgment

The authors would like to thank A. M. Chan for the programming work.

References

[1] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1983.
[2] V. Tarokh, N. Seshadri, and A. R. Calderbank, “Space-time codes for high data rate wireless communication: Performance criterion and code construction,” *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.
[3] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, “Space-time block codes from orthogonal designs,” *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.
[4] S. L. Ariyavisitakul, “Turbo space-time processing to improve wireless channel capacity,” *IEEE Trans. Commun.*, vol. 48, no. 8, pp. 1347–1359, Aug. 2000.
[5] G. J. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas,” *Bell Labs. Tech. J.*, vol. 1, pp. 41–59, 1996.
[6] I. E. Telatar, “Capacity of multi-antenna Gaussian channels,” *Eur. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.
[7] S. M. Alamouti, “A simple transmitter diversity scheme for wireless communications,” *IEEE J. Sel. Areas Commun.*, no. 10, pp. 1451–1458, Oct. 1998.
[8] C.-E. W. Sundberg and N. Seshadri, “Coded modulation for fading channels: An overview,” *Eur. Trans. Telecommun.*, vol. 4, pp. 325–334, May–Jun. 1993.
[9] C. Schlegel and D. J. Costello, “Bandwidth efficient coding for fading channels: Code construction and performance analysis,” *IEEE J. Sel. Areas Commun.*, vol. 7, no. 12, pp. 1356–1368, Dec. 1989.
[10] D. Divsalar and M. K. Simon, “The design of trellis coded MPSK for fading channels: Performance criteria,” *IEEE Trans. Commun.*, vol. 36, no. 9, pp. 1004–1012, Sep. 1988.
[11] G. Ungerboeck, “Channel coding with multilevel/phase signals,” *IEEE Trans. Inf. Theory*, vol. IT-28, no. 1, pp. 56–67, Jan. 1982.
[12] E. Zehavi, “8-PSK trellis codes for a Rayleigh channel,” *IEEE Trans. Commun.*, vol. 40, no. 5, pp. 873–884, May 1992.
[13] X. Li and J. A. Ritcey, “Bit-interleaved coded modulation with iterative decoding,” *IEEE Commun. Lett.*, vol. 1, no. 11, pp. 169–171, Nov. 1997.
[14] ——, “Trellis-coded modulation with bit interleaving and iterative decoding,” *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 715–724, Apr. 1999.
[15] H. Imai and S. Hirakawa, “A new multilevel coding method using error correcting codes,” *IEEE Trans. Inf. Theory*, vol. IT-23, no. 5, pp. 371–377, May 1977.
[16] N. Seshadri and C.-E. W. Sundberg, “Multilevel trellis coded modulations for the Rayleigh fading channel,” *IEEE Trans. Commun.*, vol. 41, no. 9, pp. 1300–1310, Sep. 1993.
[17] A. M. Tonello, “Space-time bit-interleaved coded modulation with an iterative decoding strategy,” in *Proc. IEEE Veh. Technol. Conf.*, Sep. 2000, pp. 473–478.
[18] ——, “Performance of space-time bit-interleaved codes in fading channels with simplified iterative decoding,” in *Proc. IEEE Veh. Technol. Conf.*, May 2001, pp. 1357–1361.
[19] I. Lee, A. Chan, and C.-E. W. Sundberg, “Space-time bit-interleaved coded modulation for OFDM systems,” *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 820–825, Mar. 2004.
[20] I. Lee and C.-E. W. Sundberg, “Wireless OFDM systems with multiple transmit and receive antennas with bit interleaved coded modulation,” *IEEE Wireless Commun.*, pp. 80–87, Jun. 2006.
[21] C. Berrou, A. Glavieux, and P. Thitimajshima, “Near Shannon limit error-correcting coding and decoding: Turbo-codes,” in *Proc. IEEE Int. Conf. Commun.*, May 1993, pp. 1064–1070.
[22] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, “Serial concatenation of interleaved codes: Performance analysis, design, and iterative decoding,” *IEEE Trans. Inf. Theory*, vol. 44, no. 5, pp. 909–926, May 1998.
[23] I. Lee, “The effect of a precoder on serially concatenated coding systems with isi channel,” *IEEE Trans. Commun.*, vol. 49, no. 7, pp. 1168–1175, Jul. 2001.
[24] P. W. Wolniansky, “V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel,” in *Proc. URSI Int. Symp. Signals, Syst., Electron.*, Sep. 1998, pp. 295–300.
[25] B. Hochwald and S. T. Brink, “Achieving near-capacity on a multiple-antenna channel,” *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.
[26] Y. L. C. de Jong and T. J. Willink, “Iterative tree search detection for MIMO wireless systems,” in *Proc. IEEE Veh. Technol. Conf.*, Jul. 2002, pp. 1041–1045.
[27] M. Damen, K. Abed-Meraim, and J.-C. Belfiore, “Generalised sphere decoder for asymmetrical space-time communication architecture,” *Electron. Lett.*, pp. 166–167, Jan. 2000.
[28] J. Luo, K. R. Pattipati, P. K. Willett, and F. Hasegawa, “Near-optimal multiuser detection in synchronous CDMA using probabilistic data association,” *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 361–363, Sep. 2001.
[29] A. M. Chan and I. Lee, “A new reduced-complexity sphere decoder for multiple antenna systems,” in *Proc. Int. Conf. Commun.*, Apr. 2002, pp. 460–464.

[30] X. Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1046–1061, Jul. 1999.

[31] R. Knopp and P. A. Humblet, "On coding for block fading channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 189–205, Jan. 2000.

[32] I. Lee and C.-E. W. Sundberg, "Code construction for space time bit interleaved coded modulation systems," in *Proc. Int. Conf. Commun.*, Jun. 2004, pp. 722–726.

[33] D. Haccoun and G. Begin, "High-rate punctured convolutional codes for Viterbi and sequential decoding," *IEEE Trans. Commun.*, vol. 37, no. 11, pp. 1113–1125, Nov. 1989.

[34] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1980.

[35] R. van Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*. London, U.K.: Artech House, 2000.

[36] F. Tosato and P. Bisaglia, "Simplified soft-output demapper for binary interleaved COFDM with application to HIPERLAN/2," in *Proc. Int. Conf. Commun.*, Sep. 2002, pp. 664–668.

[37] P. O. Borjesson and C.-E. Sundberg, "Simple approximation of the error function Q(x)," *IEEE Trans. Commun.*, vol. COM-27, no. 3, pp. 639–643, Mar. 1979.

[38] M. Tuchler, A. Singer, and R. Koetter, "Minimum mean squared error equalization using a priori information," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 673–683, Mar. 2002.

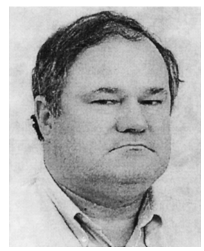
[39] P. Robertson, E. Villebrun, and P. Hoeher, "A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log domain," in *Proc. Int. Conf. Commun.*, 1995, pp. 1009–1013.

[40] V. Tarokh, A. F. Naguib, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Mismatch analysis," in *Proc. Int. Conf. Commun.*, Jun. 1997, pp. 309–313.

Inkyu Lee (S'92–M'95–SM'01) was born in Seoul, Korea, in 1967. He received the B.S. degree (Hon.) in control and instrumentation engineering from Seoul National University, Seoul, Korea, in 1990, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1992 and 1995, respectively.

From 1991 to 1995, he was a Research Assistant at the Information Systems Laboratory, Stanford University. From 1995 to 2001, he was a Member of Technical Staff at Bell Laboratories, Lucent Technologies, where he studied the high-speed wireless system design. He later worked for Agere Systems (formerly Microelectronics Group of Lucent Technologies), Murray Hill, NJ, as a Distinguished Member of Technical Staff from 2001 to 2002. In September 2002, he joined the faculty of Korea University, Seoul, Korea, where he is currently an Associate Professor in the School of Electrical Engineering. He has published over 30 journal papers in IEEE, and has 20 U.S. patents granted or pending. His research interests include digital communications, signal processing, and coding techniques applied to wireless systems.

Dr. Lee currently serves as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS in the area of Wireless Communication Theory, and has also been a Guest Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (Special Issue on 4G Wireless Systems). He received the IT Young Engineer Award as the IEEE/IEEK joint award in 2006. Also in 2006, he received the APCC Best Paper Award.



Carl-Erik W. Sundberg (S'69–M'75–SM'81–F'90) was born in Karlskrona, Sweden, in 1943, and received the M.S.E.E. and the Dr.Techn. degrees from the Lund Institute of Technology, University of Lund, Lund, Sweden, in 1966 and 1975, respectively.

During 1976, he was an ESA Research Fellow at the European Space Research and Technology Centre (ESTEC), Noordwijk, The Netherlands. From 1977 to 1984, he was a Research Professor (Docent) in the Department of Telecommunication Theory, University of Lund. From 1984 to 2000, he was a Distinguished Member of Technical Staff (DMTS) at Bell Laboratories, Murray Hill, NJ, and during 2001, he was a DMTS at Agere Systems, Murray Hill, NJ. He retired from Bell Labs/Agere in December 2001. Since 2002, he has been President and Chief Scientist with SundComm, Chatham, NJ. During part of 2003, 2004, and 2005, he was a Visiting Professor at Korea University, Seoul, Korea. During 2006, he was a Visiting Scientist at DoCoMo Communications Labs, Palo Alto, CA. His research interests include digital communications. He has published over 100 journal papers and contributed over 150 conference papers. He has over 100 US and international patents. He is a coauthor of *Digital Phase Modulation* (New York: Plenum, 1986), *Topics in Coding Theory* (New York: Springer-Verlag, 1989) and *Source-Matched Digital Communications* (New York: IEEE Press, 1996).

In 1986, Dr. Sundberg and his coauthor received the IEEE Vehicular Technology Society's Paper of the Year Award, and in 1989, he and his coauthors were awarded the Marconi Premium Proc. IEE Best Paper Award. Two of his papers were selected for inclusion in the IEEE Communications Society 50th Anniversary Journal Collection Volume 2002. He has been Guest Editor for the IEEE JOURNAL ON SPECIAL AREAS IN COMMUNICATIONS, 1988–1989 and 2004–2005. He is listed in *Marquis Who's Who in America* (New Providence, NJ: Marquis Who's Who).