An Iterative Extension of BLAST Decoding Algorithm for Layered Space–Time Signals

Ke Liu, *Student Member, IEEE,* and Akbar M. Sayeed, *Senior Member, IEEE*

Abstract—We propose an iterative extension of the Bell Laboratory Layered Space–Time (BLAST) algorithm and its variant, vertical BLAST (VBLAST). A characteristic feature of the BLAST-type algorithm is that symbol decisions with low reliability are fed back to decode other symbols. Both performance analysis based on Gaussian approximation of residual interference, and simulation results demonstrate that error propagation due to unreliable decision feedback can severely limit system performance. The extended algorithm exploits inherent signal diversity in BLAST to mitigate residual interference, thus overcoming the performance bottleneck due to error propagation. It yields an impressive performance gain over BLAST. In particular, the extension of BLAST with zero-forcing interference nulling admits a simple QR implementation and exhibits excellent performance with low complexity.

Index Terms—Bell Laboratory Layered Space–Time (BLAST) algorithm, layered space–time processing, multiantenna systems, vertical BLAST (VBLAST).

I. INTRODUCTION

AUGMENTING temporal and frequency dimensions, the spatial dimension afforded by antenna arrays holds great promise for improving wireless system performance. Information-theoretic studies in [1] and [2] lay out a foundation for deploying multiple antennas at both the transmitter and the receiver. In these works, multiple-antenna systems are shown to increase wireless channel capacity. In particular, channel capacity grows at least linearly with the number of transmit antennas, provided that the number of receive antennas is greater than or equal to the number of transmit antennas. The multiantenna channel can be described as a *multiple-input multiple-output* (MIMO) system, where the number of degrees of freedom (DOFs) is given by the product of the number of transmit and receive antennas. However, the relatively high dimensional nature of MIMO systems poses a nontrivial complexity problem on practical system design.

Code design criteria and constructions for quasi-static slow and fast fading channels have been treated extensively in the seminal paper [3]. The general coding approach is most suitable to realize spatial diversity gain for a small number of transmit antennas. A few handcrafted *space–time* codes for two transmit antennas were demonstrated to operate very close to channel outage capacity. However, the prohibitive decoding complexity limits the general coding approach when there are a large number of transmit antennas in the system. Reduced-complexity approaches that build on [3] have also been proposed (see, e.g., [4]). Signal processing techniques can be used to harness channel capacity with affordable complexity. Foschini *et al.* devised a multilayer structure, called Bell Laboratories Layered Space–Time (BLAST) [5], which decodes signals from a particular transmit antenna by using linear filtering and decision feedback to suppress interference from other transmit antennas. Linear filters in BLAST can be designed using minimum mean-square error (MMSE) or zero-forcing (ZF) criterion. Vertical BLAST (VBLAST) [6] represents an improvement over the original BLAST by dynamically optimizing the order of symbol decoding to reduce the effect of erroneous decision feedback. We will call BLAST, including its variants, the *baseline algorithm*. Different variants of the baseline algorithm generally exhibit different performance. BLAST with ZF filtering (ZF-BLAST) ranks the worst among all of the baseline algorithms in terms of performance. However, an efficient QR implementation makes it an attractive low-complexity solution to space–time decoding. On the contrary, VBLAST with MMSE filtering (MMSE-VBLAST) entails the most computational cost, but is capable of achieving excellent performance. In an important contribution [7], Hassibi has derived an efficient square-root algorithm which reduces the complexity of MMSE-VBLAST to roughly the cubic order of the number of transmit antennas. The algorithm is more involved than the simple ZF-BLAST, and hence, a nonnegligible implementation complexity may be expected.

In this paper, we study the performance of the baseline algorithm based on a Gaussian approximation of residual interference. The analysis and simulation identify the error propagation due to unreliable decision feedback as a bottleneck in limiting system performance. We observe that the quality of symbol decision gradually improves as the baseline algorithm proceeds—the last decoded symbols enjoy the highest reliability. However, potential benefits of having more reliable decisions at the last decoded symbols are largely unexploited in the baseline algorithm. Therefore, we propose an iterative extension, the *extended algorithm*, which subtracts signals due to these symbols from the received signal, and hence, reduces interference toward other symbols. The extended algorithm successively refines decisions for every transmitted symbol using a baseline algorithm. The iterative BLAST extension proposed in

K. Liu is with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: liuk@ece.osu.edu).

A. M. Sayeed is with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53705 USA (e-mail: akbar@engr.wisc.edu).

this paper belongs to a larger family of iterative decision-feedback methods. The exact analysis on such decision-feedback systems seems very difficult. However, improving the quality of feedback decisions is key to performance enhancement. Among all of the variants of the extended algorithm, we focus on the extension of ZF-BLAST (EXT-ZF-BLAST) because of its relatively small complexity and clear interpretation of algorithm behavior. It is shown via simulations that the extended algorithm is able to significantly reduce the effects of error propagation. Moreover, the extended algorithm is also very efficient, in that few iterations are needed to achieve an impressive gain. In our investigation of several extended algorithms, we found that simple EXT-ZF-BLAST can easily outperform much more complicated dynamically ordering algorithms such as ZF-VBLAST. However, the study confirms the superior performance of MMSE-VBLAST; it yields a substantial gain over ZF-BLAST, and its iterative extension only produces a marginal improvement. The contribution of this paper is in studying the effects and performance gains associated with the extended algorithm. From a complexity point of view, simple EXT-ZF-BLAST based on QR implementation performs reasonably well, and holds the merit of low complexity. If the complexity of MMSE-VBLAST can be afforded, the Hassibi algorithm for MMSE-VBLAST should be applied [7].

The outline of the paper is as follows. Section II briefly reviews various baseline algorithms. The extended algorithm is formalized in Section III. The ZF-BLAST algorithm and its extension are presented from a particular implementation, QR decomposition, which concisely represents the layered space–time structure in BLAST and simplifies signal processing and decoding. Performance of the baseline or extended algorithm can be analyzed by approximating residual interference as Gaussian variables. We use ZF-BLAST and its extension to illustrate the proposed analysis technique. Simulation results are provided in Section IV to demonstrate the strength of the extended algorithm and to investigate performance of different algorithm variants. Finally, we make some concluding remarks in Section V.

The following notations are used throughout this paper. Let $(\cdot)^{\dagger}$ and $(\cdot)^{\top}$ denote the Hermitian (complex) transpose and transpose, respectively. A k-dimensional (complex) real Gaussian random distribution with mean $\mathbf{m}$ and correlation matrix $\mathbf{R}$ is denoted by $(\mathrm{C})\mathrm{N}^k(\mathbf{m}, \mathbf{R})$. Let $a_1, \ldots, a_k$ be independent, identically distributed (i.i.d.) $\mathrm{N}^1(0, 0.5)$ random variables. Then, $\sum_{i=1}^{k} a_i^2$ is a chi-squared random variable with k DOFs, denoted by χ_k^2. The expectation is denoted by $\mathrm{E}[\cdot]$, and "orthogonal to" by $\perp$. The $k \times k$ identity matrix is denoted by $\mathbf{I}_k$.

II. BASELINE PROCESSING FOR LAYERED SPACE–TIME SIGNALS

Consider a narrowband multiantenna system with N transmit antennas and M receive antennas, denoted by an (N, M) system. At a given discrete time instant, the M-dimensional received signal $\mathbf{r}$ and the N-dimensional transmitted signal $\mathbf{s}$ are related by

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{w} \tag{1}$$

where the noise vector $\mathbf{w} \sim \mathrm{CN}^M(\mathbf{0}, \mathbf{I}_M)$ is assumed to be uncorrelated in time as well. The channel matrix $\mathbf{H}$ in (1) represents couplings between different pairs of transmit and receive antennas. In rich scattering environments, elements of the channel matrix can be modeled as i.i.d. $\mathrm{CN}^1(0, 1)$ random variables. We adopt a quasi-static approximation of the fading channel, that is, the channel remains unchanged during a *coherence period* (T_{coh} discrete time) and changes independently from one period to another. In this paper, the channel state information $\mathbf{H}$ is available at the receiver, but not known to the transmitter.

The same signal constellation with average power E_s is used at every transmit antenna. Since the noise has unit variance in the channel model, the signal-to-noise ratio (SNR) is then given by

$$\mathrm{SNR} = NE_s. \tag{2}$$

Note that the total transmitted power is fixed independent of the number of transmit antennas.

A layered space–time system modulates incoming bit streams onto symbols and transmits them across all of the transmit antennas. For simplicity, we assume that symbols are uncoded, that is, no error-control code is used in the conversion from information bits into transmitted symbols. Essential ideas in this paper would remain unchanged in the case of coded systems. The uncoded layered space–time signal can be regarded as a special example of *space–time codes* in [3], which gave an upper bound of the pairwise error probability of decoding $\hat{\mathbf{X}}$ while $\mathbf{X}$ is transmitted

$$P(\mathbf{X} - \hat{\mathbf{X}}) \leq \left(\prod_{i=1}^{N} \left(1 + \frac{\lambda_i E_s}{4} \right) \right)^{-M} \tag{3}$$

where λ_i for $1 \leq i \leq N$ are eigenvalues of $\mathbf{A} = (\mathbf{X} - \hat{\mathbf{X}})(\mathbf{X} - \hat{\mathbf{X}})^{\dagger}$. Let $r = \mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{X} - \hat{\mathbf{X}})$. The number of nonzero eigenvalues is r in (3). Therefore, if every codeword difference matrix has rank not less than r, the corresponding space–time code can guarantee an rM order of diversity. Since a codeword in layered space–time systems is simply the uncoded transmitted signal vector, it is easy to see that the codeword difference between two distinct codewords is only of rank 1. Thus, one can conclude from (3) that (uncoded) layered space–time signals attain a diversity of M, which is the number of receive antennas. This implies that uncoded multiantenna systems could deliver good performance while operating at a high rate, if a large number of receive antennas are available. However, such performance would require the use of maximum-likelihood (ML) decoding at the receiver. As N increases, ML decoding quickly becomes impractical. Hence, alternative decoding algorithms of low complexity, although suboptimum, are sought instead. If transmit antennas are treated as "users," the channel (1) resembles that of *multiple-access channels*, which suggests that a vast amount of multiuser detection literature [8], [9] can be carried over to the decoding of layered space–time signals. BLAST and its variant VBLAST are those signal-processing algorithms that draw insights from multiuser detection techniques.

A. BLAST

The BLAST [5] algorithm is centered around the notion of *interference nulling* and *interference cancellation*. Suppose the processing starts from s_N, the symbol transmitted at the Nth transmit antenna. Rewrite (1) as

$$\mathbf{r} = \sum_{n=1}^{N} \mathbf{h}_n s_n + \mathbf{w} \tag{4}$$

where $\mathbf{h}_n$ is the nth column of $\mathbf{H}$ for $1 \leq n \leq N$. Hence, the signal due to the symbol s_N is in the direction of $\mathbf{h}_N$, and the interference toward s_N lies within a linear subspace spanned by $\mathbf{h}_1, \ldots, \mathbf{h}_{N-1}$. Interference nulling tries to reduce the amount of interference toward s_N by applying a linear filter $\mathbf{f}$, the nulling vector, to the received signal. Typically, the linear filter is derived under the ZF or MMSE criterion, which gives rise to ZF-BLAST or MMSE-BLAST, respectively. In ZF-BLAST, interference from other transmit antennas is completely removed by projecting the received signal into a direction orthogonal to the interference subspace. Let $\mathbf{H}_{\sim n}$ and $\mathcal{H}_{\sim n}$ denote the matrix obtained from $\mathbf{H}$ by removing its nth column and the corresponding linear subspace. Therefore, ZF-BLAST requires $\mathbf{f} \perp \mathcal{H}_{\sim N}$. It can be solved by

$$\mathbf{f}_{\mathrm{zf}} = \left(\mathbf{I} - \mathbf{H}_{\sim N}(\mathbf{H}_{\sim N}^{\dagger}\mathbf{H}_{\sim N})^{-1}\mathbf{H}_{\sim N}^{\dagger}\right)\mathbf{h}_N. \tag{5}$$

Although the ZF interference nulling completely removes interference, it weakens signal as well especially when $\mathbf{f}$ and $\mathbf{h}_N$ form a large angle so that the projection of $\mathbf{h}_N$ onto $\mathbf{f}$ has small power. As an improvement, MMSE-BLAST chooses the linear filter to minimize $\mathrm{E}|\mathbf{f}^{\dagger}\mathbf{r} - s_N|^2$. Note that $\mathrm{E}\mathbf{s}\mathbf{s}^{\dagger} = E_s\mathbf{I}$. The solution is given by

$$\mathbf{f}_{\mathrm{mmse}} = E_s\left(\mathbf{I} + E_s\mathbf{H}\mathbf{H}^{\dagger}\right)^{-1}\mathbf{h}_N. \tag{6}$$

After interference nulling, the decision of s_N, denoted by $\hat{s}_N$, is generated from $\tilde{s}_N = \mathbf{f}^{\dagger}\mathbf{r}$. Then, the algorithm reconstructs the signal due to s_N and subtracts it from the received signal $\mathbf{r}$. In other words, the interference from s_N toward the rest of the undecoded symbols is canceled. If the above interference cancellation is perfect, the system has one less transmit antenna

$$\mathbf{r}^{(N-1)} = \mathbf{r} - \mathbf{h}_N \hat{s}_N = \sum_{n=1}^{N-1} \mathbf{h}_n s_n + \mathbf{w} = \mathbf{H}^{(N-1)}\mathbf{s}^{(N-1)} + \mathbf{w} \tag{7}$$

where the superscript denotes the reduced order of the problem. Applying the same interference nulling and cancellation procedures to the order-reduced problem, that is, an $(N-1, M)$ system, BLAST successively decodes symbols transmitted at every transmit antenna.

B. VBLAST

Close examination of BLAST reveals that decision feedback comes from "weaker" symbols that experience more interference from other symbols. For example, symbol s_N is interfered by all of the rest of $N - 1$ symbols. However, it is the first one to be decoded in BLAST, and its decision is used in the subsequent interference cancellation. The unreliable decision results in nonnegligible error propagation, which can cost a severe performance penalty. Thus, an improvement can be that

the decoding algorithm chooses the "best" symbol at each decoding step. VBLAST [6] ordered symbols according to their mean square error (MSE) after interference nulling. It chooses the one with minimal MSE to perform actual decoding and interference cancellation.

In order to determine the order among symbols, one needs to determine the nulling vector for every symbol. Fortunately, all of the nulling vectors can be computed at once. Let $\mathbf{f}_n$ denote the nulling vector for the nth symbol. The interference nulling for each symbol is performed as

$$\tilde{\mathbf{s}} = \mathbf{F}^{\dagger}\mathbf{r} \tag{8}$$

where the nth column of $\mathbf{F}$ is $\mathbf{f}_n$. The filter matrix $\mathbf{F}$ for ZF or MMSE nulling is given by [6], [7]

$$\mathbf{F}_{\mathrm{zf}} = \mathbf{H}\left(\mathbf{H}^{\dagger}\mathbf{H}\right)^{-1}$$
$$\mathbf{F}_{\mathrm{mmse}} = E_s\mathbf{H}\left(\mathbf{I} + E_s\mathbf{H}^{\dagger}\mathbf{H}\right)^{-1}. \tag{9}$$

The MSE of interference nulling for every symbol is given by

$$\mathrm{E}\left[(\mathbf{s}-\tilde{\mathbf{s}})(\mathbf{s}-\tilde{\mathbf{s}})^{\dagger}\right] = \begin{cases} \left(\mathbf{H}^{\dagger}\mathbf{H}\right)^{-1}, & \text{ZF Nulling} \\ E_s\left(\mathbf{I}+E_s\mathbf{H}^{\dagger}\mathbf{H}\right)^{-1}, & \text{MMSE Nulling.} \end{cases} \tag{10}$$

VBLAST decodes the "strongest" symbol whose MSE is smallest among all of the symbols, thus improving the quality of decision feedback. After the symbol has been decoded, its interference toward other symbols is subtracted. Ordering and interference nulling/cancellation operations repeat until all of the symbols have been decoded. As we will demonstrate via simulations, dynamic ordering in VBLAST can effectively mitigate error propagation due to unreliable decision feedback.

In this paper, BLAST-type algorithms will be termed baseline algorithms, as they are not iterative. We also use the names such as ZF-BLAST and MMSE-VBLAST to categorize different baseline variants.

III. EXTENDED BLAST DECODING

As discussed above, error propagation due to imperfect decision feedback in the baseline algorithm represents a bottleneck in system performance. Although dynamic ordering, as in the VBLAST, is an effective method to overcome this bottleneck, there exist other remedies that yield low-complexity solutions. Roughly speaking, the last symbol to be decoded in the baseline algorithm would experience the least amount of interference from other symbols, because interference from the other symbols has been canceled. Thus, its decisions are relatively more reliable. However, this observation is not used in the baseline algorithm. So, a simple way to improve system performance is to subtract from the received signal the signal of the last decoded symbol after an initial application of the baseline algorithm. This forms a feedback flow in the reverse order, which we term *loopback*. Assuming good loopback cancellation, loopback operation can effectively remove the contribution of one transmit antenna, which helps to improve the next iteration of the baseline algorithm, since the number of interfering sources is reduced by one. Loopback operation defines the distinct characteristics of the extension of the BLAST-type algorithm, which we call the extended algorithm. Moreover, the algorithm can be configured to loopback more symbols, and thus operates in an

iterative fashion. We formalize the proposed extension in the following algorithm.

Algorithm 1 (Extended Algorithm): Let I_{loopback} denote the number of loopback iterations. Fix a baseline algorithm to be used in space–time decoding. Maintain a set $\mathcal{L}$ whose elements are the symbols used in loopback operations.

Step 1) Decode the original (N, M) system using the baseline algorithm. Add the last decoded symbol into $\mathcal{L}$.

Step 2) Subtract the signal due to the symbols in $\mathcal{L}$ from previous iterations of the baseline algorithm.

Step 3) Apply the baseline algorithm to generate decisions for symbols that are not in $\mathcal{L}$.

Step 4) Update symbol decision for every element in $\mathcal{L}$ by subtracting interference from all the other symbols. Add the last decoded symbol in Step 3) into $\mathcal{L}$.

Step 5) Repeat Steps 2)–4) by I_{loopback} times.

Note that $I_{\text{loopback}} = 0$ corresponds to the baseline algorithm, and different variants of the baseline algorithm give rise to different variants of the extended algorithm. For example, EXT-ZF-BLAST stands for the extended algorithm with ZF-BLAST as its baseline algorithm.

A. EXT-ZF-BLAST

Among many variants of the extended algorithm, we primarily focus on EXT-ZF-BLAST because of its relatively low implementation complexity through QR decomposition [7], [10], [11].

We may assume that the number of transmit antennas is not larger than the number of receive antennas, that is, $N \leq M$. Essentially the same idea applies for the case of $N > M$, with a small modification given in [4]. The QR decomposition [12] of the channel matrix $\mathbf{H}$ is

$$\mathbf{H} = \mathbf{Q}\mathbf{G} \qquad (11)$$

where $\mathbf{Q}$ is an $M \times M$ unitary matrix and the $M \times N$ upper triangular matrix $\mathbf{G}$ has the following form:

$$\mathbf{G} = \begin{pmatrix} g_{1,1} & \cdots & \cdots \\ 0 & \ddots & \vdots \\ \vdots & 0 & g_{N,N} \\ 0 & \cdots & 0 \end{pmatrix}. \qquad (12)$$

It can be shown [1], [4]

$$\begin{aligned} |g_{j,j}|^2 &\sim \chi^2_{2(M-j+1)}, & 1 \leq j \leq N \\ |g_{j,k}|^2 &\sim \chi^2_2, & k > j, 1 \leq j \leq N. \end{aligned} \qquad (13)$$

The upper triangle structure of $\mathbf{G}$ simplifies the ZF-BLAST algorithm considerably. All of the interference-nulling operations in BLAST can be done in one step prior to decoding by multiplying $\mathbf{r}$ by $\mathbf{Q}^\dagger$. More specifically

$$\mathbf{y} = \mathbf{Q}^\dagger\mathbf{r} = \mathbf{G}\mathbf{s} + \mathbf{z} \qquad (14)$$

where $\mathbf{z} = \mathbf{Q}^\dagger\mathbf{w} \sim \text{CN}^M(\mathbf{0}, \mathbf{I})$. Since $\mathbf{G}$ is an upper triangular matrix, symbols with larger indexes avoid interference from symbols with small indexes, which exactly reflects the pattern of interference nulling in BLAST. We summarize the implementation in the following algorithm.

Algorithm 2 (QR Implementation of ZF-BLAST): The processing proceeds as follows.

Step 1) Perform the unitary transformation in (14)

Step 2) Decode symbol of the Nth transmit antenna according to $y_N = g_{N,N}s_N + w_N$. Denote the corresponding decision by $\hat{s}_N$.

Step 3) The algorithm successively decodes symbols from $N-1$ to the first transmit antenna. Given n ($N-1 \geq n \geq 1$), subtract interference generated by $s_{n+1}, \ldots, s_N$ as

$$\tilde{y}_n = y_n - \sum_{k=n+1}^{N} g_{n,k}\hat{s}_k$$

$$= g_{n,n}s_n + \sum_{k=n+1}^{N} g_{n,k}(s_k - \hat{s}_k) + z_n. \qquad (15)$$

Decision $\hat{s}_n$ is then generated from $\tilde{y}_n$.

Algorithm 2 concisely represents the key BLAST-type signal-processing techniques. Assuming perfect interference cancellation, the algorithm creates N noninterfering one-dimensional (1-D) subchannels, corresponding to every transmit antenna, as described in [5]. The nth subchannel can be expressed as

$$\tilde{y}_n = g_{n,n}s_n + z_n, \quad 1 \leq n \leq N \qquad (16)$$

where $|g_{n,n}|^2 \sim \chi^2_{2(M-n+1)}$, that is, the nth subchannel has diversity order of $M - n + 1$. Besides the layered space–time processing, codes can be designed to improve system performance (see, e.g., [13]). However, as we elaborated before, the decision from the "weakest" subchannel, that is, the one with lowest diversity, is used to create a subchannel with higher diversity. Therefore, the performance of those subchannels with higher diversity may be compromised, due to error propagation. The extended algorithm EXT-ZF-BLAST can be used to overcome the performance bottleneck while maintaining low complexity of QR implementation.

We use a (6, 6) system as an example to explain the operation of EXT-ZF-BLAST. *Algorithm 2* is applied to generate decisions of the first subchannel, which is the strongest one. The signal emitted by the first transmit antenna can be estimated from the decisions. We subtract it from the received signal, which now can be treated as being generated from a (5, 6) system. If we apply *Algorithm 2* for this system, the diversity order of all subchannels, ranging from 2 to 6, is increased by one level, compared with the original system. Improved subchannels produce better decisions, which, in turn, help the decoding of the first subchannel. This completes the first loopback operation, after which, the diversity order of all subchannels (starting from 1 to 6) is 6, 6, 5, 4, 3, and 2, respectively. We see that the second subchannel has full diversity order besides the first subchannel. We can start the next loopback operation by canceling signals from both the first and the second subchannels, thus effectively forming a (4, 6) system. Define the depth of loopback operation to be I_{loopback}, that is, subchannel 1 to I_{loopback} are used successively in the loopback cancellation. Since the processing structure evolves after each loopback operation, the final structure has the property that the first $I_{\text{loopback}} + 1$ subchannels have full diversity order, while the rest have five down to the $1 + I_{\text{loopback}}$ level of

diversity. The formal EXT-ZF-BLAST is given in the following algorithm for an (N, M) system.

Algorithm 3 (EXT-ZF-BLAST): For $1 \leq n \leq N$, we define

$$\mathbf{H}_n = (\mathbf{h}_{n+1}, \ldots, \mathbf{h}_N) \tag{17}$$

where $\mathbf{h}_n$ is the nth column of $\mathbf{H}$. Also, let $\mathbf{s}_n = (s_{n+1}, \ldots, s_N)^\top$. Denote the loopback depth by $I_{\text{loopback}} \leq n_T - 1$. The algorithm proceeds as follows.

Step 1) Decode each symbol using *Algorithm 2* for the original (N, M) system. Denote the initial decision of the symbol s_n by $\hat{s}_n$.

Step 2) For $i = 1$ to I_{loopback}, do the following:

a) Subtract signals due to symbols from s_1 to s_i as

$$\tilde{\mathbf{r}} = \mathbf{r} - \sum_{n=1}^{i} \mathbf{h}_n \hat{s}_n = \mathbf{H}_i \mathbf{s}_i + \sum_{n=1}^{i} \mathbf{h}_n (s_n - \hat{s}_n) + \mathbf{w} = \mathbf{H}_i \mathbf{s}_i + \tilde{\mathbf{w}}. \tag{18}$$

b) Apply *Algorithm 2* to the $(N - i, M)$ system expressed in (18) to generate decisions $\hat{\mathbf{s}}_i = (\hat{s}_{i+1}, \ldots, \hat{s}_N)^\top$.

c) From $k = i$ to 1, cancel interference from other symbols as

$$\tilde{\mathbf{r}}_k = \mathbf{r} - \sum_{n \neq k} \mathbf{h}_n \hat{s}_n. \tag{19}$$

The updated decision $\hat{s}_k$ is generated from $\tilde{\mathbf{r}}_k$.

B. Approximate Performance Analysis

The subchannels created by the layered space–time processing from *Algorithm 2* are coupled through decision feedback. Since subchannels with higher indexes have less diversity, information bits transmitted through them need more protection. If the corresponding decisions suffer considerable degradation in quality, the higher diversity gain afforded by subchannels with lower indexes cannot be realized, due to nonnegligible interference. We could analytically study the performance of the baseline algorithm and its extension by modeling residual interference in every step of the algorithm as Gaussian variables, and evaluating the performance based on this approximation. In this section, we use ZF-BLAST ant its extension to illustrate the method. For simplicity, we assume that binary phase-shift keying (BPSK) is used.

A Dth-order diversity channel can be described as

$$y = hx + w \tag{20}$$

where $|h|^2 \sim \chi_{2D}^2$ and $w \sim \text{CN}^1(0, \sigma^2)$. Let E_s denote signal power, that is, $E_s = \text{E}|x|^2$. The $\rho = E_s/\sigma^2$ is the average SNR per diversity branch. The probability of error P_e for BPSK modulation given channel realization h is

$$P(\rho|h) = \text{Q}\left(\sqrt{2|h|^2\rho}\right) \tag{21}$$

where $\text{Q}(x) = \int_x^\infty (1/2\pi) e^{-(t^2/2)} dt$ is the tail integration of Gaussian function. The average P_e, $P(\rho)$, is the expectation of $P(\rho|h)$ in (21) with respect to random variable $|h|^2$, and is given in [14] as

$$P(\rho) = \left[\frac{1}{2}(1-\mu)\right]^D \sum_{d=0}^{D-1} \binom{D-1-d}{d} \left[\frac{1}{2}(1+\mu)\right]^d \tag{22}$$

where $\mu = \sqrt{\rho/(1+\rho)}$.

Recall that the second term in (15) is the interference term due to imperfect decision feedback. In order to quantify the effect of interference, we approximate it as a Gaussian random variable, denoted by $\tilde{w}_n$. Suppose P_e for s_k is p_k, that is, $\Pr(s_k \neq \hat{s}_k) = p_k$. Given equal probability of $\sqrt{E_s}$ and $-\sqrt{E_s}$, it is easy to verify that

$$\text{E}[s_k - \hat{s}_k] = 0 \tag{23}$$

$$\text{E}\left[|s_k - \hat{s}_k|^2\right] = 4 p_k E_s. \tag{24}$$

Therefore, the mean and variance of $\tilde{w}_n$ are given by

$$\text{E}[\tilde{w}_n] = 0 \tag{25}$$

$$\text{E}\left[|\tilde{w}_n|^2\right] = 4 E_s \sum_{k=n}^{N} p_k \tag{26}$$

where we have used the fact that $g_{j,k} \sim \chi_2^2$, and it is independent of $s_k - \hat{s}_k$.

Now we are ready to give an algorithm that calculates P_e of all subchannels under Gaussian approximation for interference.

Algorithm 4 (Performance Approximation for ZF-BLAST): Given SNR, $E_s = \text{SNR}/N$ in (2).

Step 1) Calculate p_N of Nth subchannel by (22) with diversity order $D = M - N + 1$ and $\rho = E_s$.

Step 2) Calculate p_k successively where $N - 1 \geq k \geq 1$. For given k, interference power $\tilde{\sigma}^2$ is calculated by (26). Then p_k is obtained using (22) with $D = N - k + 1$ and $\tilde{\rho} = E_s/(1 + \tilde{\sigma}^2)$.

Step 3) The overall P_e is obtained by averaging across all the subchannels, that is, $P_b(\text{SNR}) = \sum_{k=1}^{N} p_k/N$.

Similarly, we can analytically study the performance of EXT-ZF-BLAST by approximating the residual interference after decision feedback as Gaussian random variables. Hence, we have the following algorithm parallel to *Algorithm 4*.

Algorithm 5 (EXT-ZF-BLAST): Let $\mathbf{p} = (p_1, \ldots, p_{n_T})^\top$ be a vector of P_e's of all subchannels at each step in *Algorithm 3* with loopback parameter I_{loopback}. Denote $\mathbf{p}_n = (p_{n+1}, \ldots, p_N)^\top$. Given SNR, $E_s = \text{SNR}/N$ by (2).

Step 1) Calculate $\mathbf{p}$ using *Algorithm 4* for the original (N, M) system using ZF-BLAST.

Step 2) For $i = 1$ to I_{loopback}, do the following.

By Gaussian approximation of residual interference, $\tilde{\mathbf{w}}$ in (18) is distributed as $\text{CN}^M(0, \tilde{\sigma}^2)$ with

$$\tilde{\sigma}^2 = 1 + 4 E_s \sum_{n=1}^{i} p_n. \tag{27}$$

Apply *Algorithm 4* for the $(N - i, M)$ system in (18) to generate $\mathbf{p}_i$.

From $k = i$ to 1, update p_k using (22) with diversity order $D = M$ and $\rho = E_s/(1 + 4 E_s \sum_{j \neq k} p_j)$ in view of (19).

IV. NUMERICAL RESULTS

Fig. 1 shows simulation performance of the baseline space–time processing (ZF-BLAST) for a (6, 6) uncoded system. T_{coh} is chosen to be 10 000 time instants. For each static channel coherence period, $\mathbf{H}$ is generated randomly by sampling independent $\text{CN}^1(0, 1)$ variables for all its entries. Information bits are modulated using BPSK onto six transmit
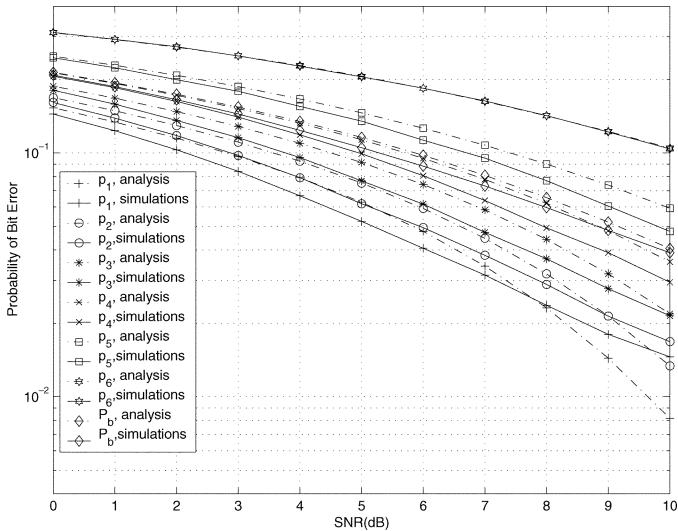
Fig. 1. Baseline space–time processing performance for a (6, 6) BPSK uncoded system. p_k where $1 \leq k \leq 6$ denotes probability of bit error for kth subchannel. Simulated performance, plotted as the solid line, is obtained by averaging over 10 000 channel static periods with $T_{\text{coh}} = 10\,000$. Performance from approximate analysis is also plotted in dashed lines.
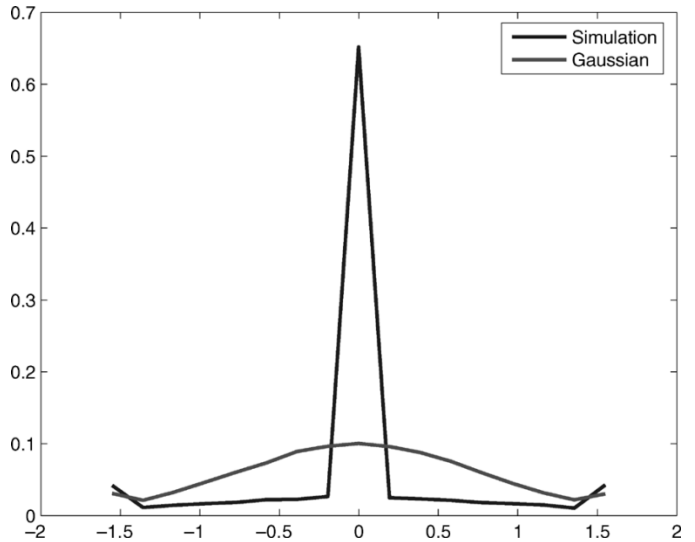

Fig. 2. Histogram of the interference signal corresponding to the first subchannel of a (6, 6) BLAST simulation. The horizontal scale is with respect to the data variance.
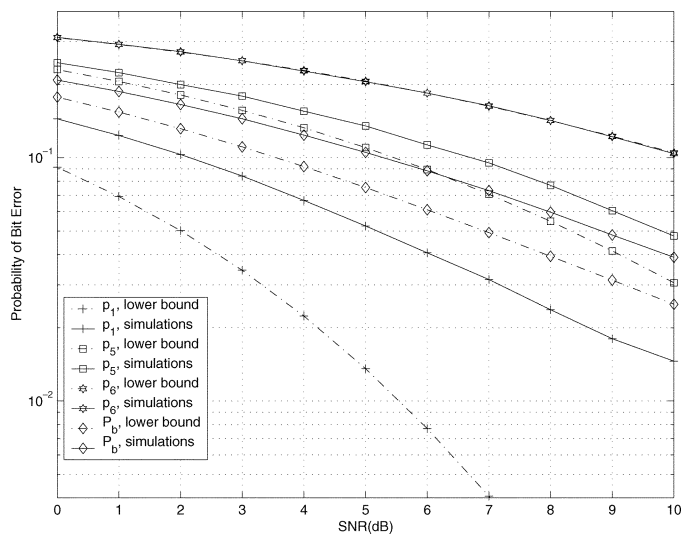

Fig. 3. Ideal subchannel performance assuming perfect decision feedback, the lower bound for P_e is compared with actual simulated performance for the simulation in Fig. 1.

antennas at every time instant. *Algorithm 2* is applied at the receiver. P_e is averaged over 10 000 channel coherence periods, and thus system performance is estimated over 10^8 time instants. We plot overall P_e and P_e for each subchannel from simulations in Fig. 1. We also plot the corresponding analytical performance curves from *Algorithm 4*. It is seen that our approximate analysis deviates from simulation data at higher SNR. However, they are very close for overall P_e. Furthermore, we observe from the figure that subchannels with higher indexes tend to have worse performance, since those subchannels have less diversity.

The Gaussian approximation is motivated by the argument of the central limit theorem, as the interference consists of a sum of individual error signals. However, the approximation is not accurate for small values of antenna dimension. Fig. 2 plots the actual histogram of the interference, collected from BLAST simulations. It is seen that the interference has much higher concentration around the center than a corresponding Gaussian distribution with the same mean and variance. This may explain the discrepancy between the approximate analysis and simulation data. However, as clearly demonstrated in Fig. 1, the Gaussian approximation is able to capture the essential behavior of decoding algorithm, thus serving as a useful tool to study the qualitative algorithm characteristics.

Imperfect decision feedback in the baseline algorithm compromises the performance of the subchannels with large diversity order. The ideal P_e performance of each subchannel assuming perfect interference cancellation is computed using (22), and serves as a lower bound on the actual P_e performance. In Fig. 3, we plot both the ideal and the actual P_e for several subchannels in the (6, 6) system used in the above simulation. Comparing ideal performance with actual simulation data, we can see that imperfect decision feedback severely compromises the performance of subchannels with large diversity order. Moreover, the overall system performance is limited by the worst

subchannel, as evident from the figure. Therefore, the worst subchannel in layered space–time structures is the bottleneck that limits system performance. Also, the significant gap between ideal and actual performance of subchannels with high diversity indicates significant room for improvement.

Fig. 4 demonstrates performance improvement with loopback cancellation in the extended space–time algorithm (EXT-ZF-BLAST) for a (6, 6) system. As before, bits are sent using BPSK at each transmit antenna. We vary the loopback depth I_{loopback} from 1 to 5. The performance of the baseline algorithm (ZF-BLAST) is also included as a comparison. As is evident from the figure, a large performance gain is achieved by using loopback cancellation. For example, the system with full loopback at 4 dB already achieves P_e of the BLAST system at 9 dB, resulting in a 5-dB saving. We note that performance gain increases as SNR increases, which projects more power savings at higher SNR.
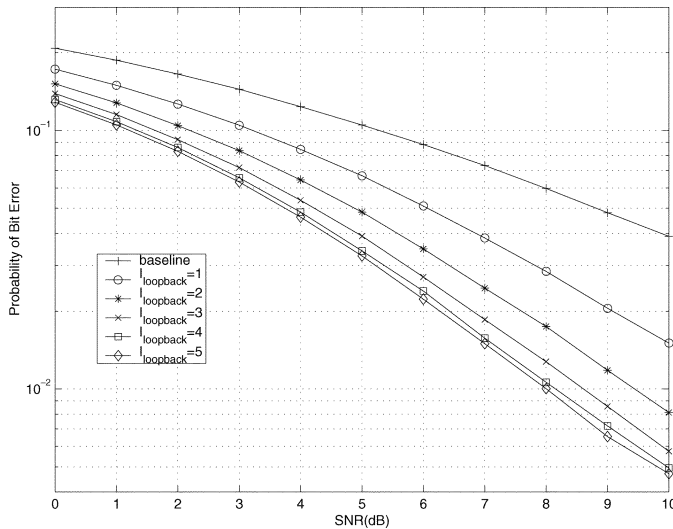
Fig. 4. Simulated (6, 6) BPSK uncoded system performance using EXT-ZF-BLAST with different I_{loopback}. Performance of the baseline algorithm is also included for comparison. Dramatic improvement in performance compared with the baseline algorithm is evident.

Moreover, a few loopback cancellations are sufficient. Fig. 4 shows that the algorithm is able to achieve large performance improvement with only one or two levels of loopback.

Different layered space–time decoding algorithms entail different complexity. MMSE filtering is more expensive than the ZF filtering, because of the QR implementation used by the latter. Dynamic order in the algorithm adds additional complexity. We are interested in comparing the performance of different variants of the baseline (extended) algorithms. As demonstrated by the simulations and analysis above, ZF-BLAST has the minimal complexity, but its performance is limited. Since EXT-ZF-BLAST maintains the low-complexity advantage as its baseline algorithm, we single out EXT-ZF-BLAST in our simulation study.

Fig. 5 plots the performance of ZF-BLAST, ZF-VBLAST, EXT-ZF-BLAST, MMSE-VBLAST, and EXT-VBLAST on a simulated (4, 4) system with 4-ary quadrature amplitude modulation (4-QAM). It is seen that one loopback operation in EXT-ZF-BLAST is sufficient to bring significant gain over its baseline algorithm ZF-BLAST. EXT-ZF-BLAST outperforming ZF-VBLAST suggests that the cost of dynamic ordering with ZF filtering is unnecessary. The performance enhancement due to loopback is critically linked to the relative improvement in terms of the quality of feedback decisions. In an MMSE-VBLAST system, each subchannel decoding is dynamically optimized to best exploit channel diversity. Thus, the loopback can only bring in a marginal gain on decision quality, which is demonstrated in the figure, showing that MMSE-VBLAST exhibits exceptional performance and its iterative extension seems not worthy of the extra cost. However, the implementation of MMSE-VBLAST is much more complicated, compared with that of ZF-BLAST and EXT-ZF-BLAST. Therefore, EXT-ZF-BLAST could well serve as a low-complexity option for layered space–time processing. As we increase the number of antennas in the system, the performance gain of EXT-ZF-BLAST gradually increases.
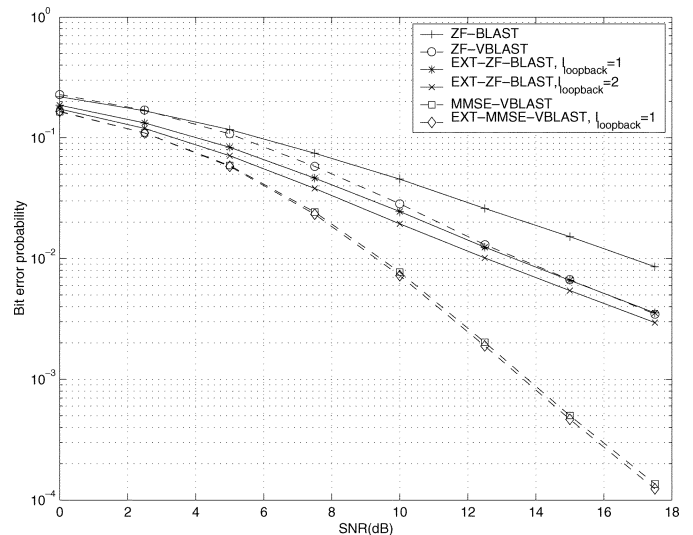

Fig. 5. Performance comparison between variants of baseline (extended) algorithm on a simulated (4, 4) 4-QAM system. Extended algorithm is seen to improve upon its corresponding baseline algorithm. EXT-ZF-BLAST outperforms ZF-VBLAST, which suggests that the ordering overhead could be avoided in this case. It is also evident from the figure that MMSE-BLAST ranks the best in various algorithms, and its extension only brings marginal performance gain.
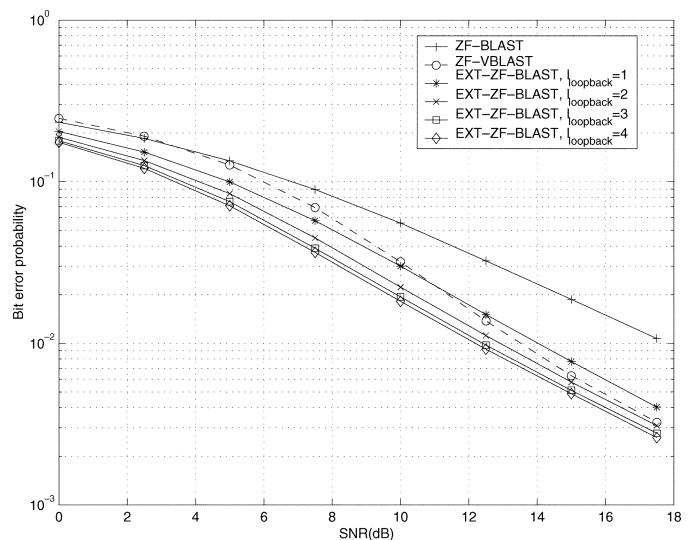

Fig. 6. Performance comparison between variants of baseline (extended) algorithm on a simulated (6, 6) 4-QAM system. Extended algorithm outperforms ZF-VBLAST.

We plot the performance of EXT-ZF-BLAST on a simulated (6, 6) system using 4-QAM modulation in Fig. 6. Comparing Fig. 5 with Fig. 6, one can see at $P_e = 10^{-2}$ the extended algorithm achieves a 6-dB gain in a (6, 6) system, while only a 4-dB gain in a (4, 4) system. For the same (6, 6) system as in Fig. 6, we show the performance of MMSE-VBLAST in Fig. 7, which again demonstrates that an extended algorithm for MMSE-VBLAST could not offer much benefit.

V. CONCLUSION

We have presented an iterative extension of the baseline BLAST-type algorithm. The extended algorithm is aimed at mitigating error propagation due to imperfect decision feedback
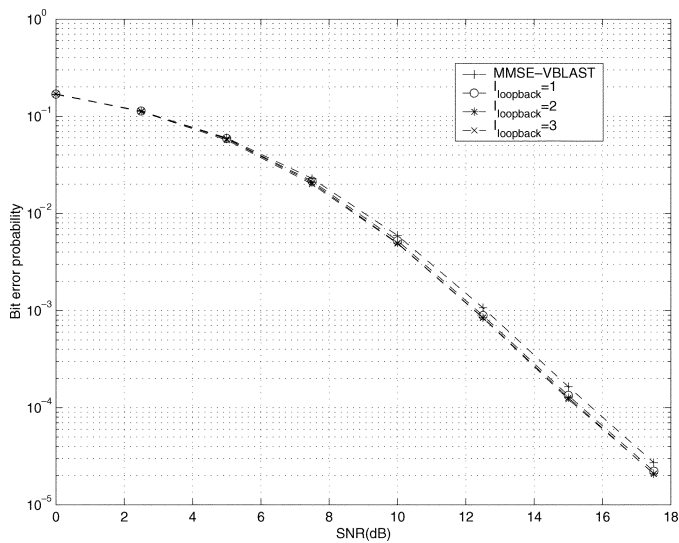
Fig. 7. Performance of EXT-MMSE-VBLAST on a simulated (6, 6) 4-QAM system. The extended algorithm does not seem to significantly improve over the baseline algorithm.

in BLAST. Performance and cost tradeoffs play an important role in choosing the appropriate algorithm for a particular application. Among all of the variants of the baseline (extended) algorithm, EXT-ZF-BLAST seems a good candidate for low-complexity applications, yet is capable of achieving good performance. However, if the complexity is not a primary concern, MMSE-VBLAST gives the best performance.

REFERENCES

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[2] I. E. Telatar, "Capacity of Multi-Antenna Gaussian Channels," AT&T Bell Labs., Tech. Rep., 1995.

[3] 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.

[4] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Combined array processing and space-time coding," *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 1121–1128, May 1999.

[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, no. 2, pp. 41–59, 1996.

[6] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. ISSSE*, Pisa, Italy, Sep. 1998, pp. 295–300.

[7] B. Hassibi, "An efficient square-root algorithm for BLAST," in *Proc. ICASSP*, Istanbul, Turkey, Jun. 2000, pp. 11737–11740.

[8] S. Verdú, *Multiuser Detection*. New York: Cambridge Univ. Press, 1998.

[9] M. K. Varanasi, "Decision feedback multiuser detection: A systematic approach," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 219–240, Jan. 1999.

[10] T. Marzetta, "BLAST training: Estimating channel characteristics for high-capacity space-time wireless," in *Proc. 37th Annu. Allerton Conf.*, Monticello, IL, Sep. 1999, pp. 958–966.

[11] K. Liu and A. M. Sayeed, "Improved layered space-time processing in multiple antenna systems," in *Proc. 39th Annu. Allerton Conf.*, Monticello, IL, Oct. 2001.

[12] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge Univ. Press, 1988.

[13] D. S. Shiu and J. M. Kahn, "Scalable layered space–time codes for wireless communications performance and analyis and design criteria," in *Proc. WCNC*, New Orleans, LA, Sep. 1999, pp. 159–163.

[14] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

Ke Liu (S'99) received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 1999, and the M.S. degree in electrical engineering, the M.A. degree in mathematics, and the Ph.D. degree in electrical engineering from the University of Wisconsin, Madison, in 2001, 2002, and 2004, respectively.

From 2000 to 2004, he was a Research Assistant with the Wireless Communications Laboratory, University of Wisconsin, Madison. Since October 2004, he has been with The Ohio State University, Columbus, where he currently serves as a Postdoctoral Researcher with the Information Processing Systems Laboratory. His research interests include information theory, wireless communications, wireless ad hoc networks, and statistical signal processing.

Dr. Liu was the recipient of the Graduate School Fellowship from 1999 to 2000 and the Harold A. Peterson Graduate Student Paper Award for his work on space–time coding for multiantenna systems. He has served as a reviewer for the IEEE Communications, Signal Processing, and Information Theory Societies.

Akbar M. Sayeed (S'89–M'97–SM'02) received the B.S. degree from the University of Wisconsin, Madison in 1991, and the M.S. and Ph.D. degrees from the University of Illinois at Urbana-Champaign, Urbana, in 1993 and 1996, respectively, all in electrical and computer engineering.

While with the University of Illinois, he was a Research Assistant with the Coordinated Science Laboratory and was also the Schlumberger Fellow in signal processing from 1992 to 1995. During 1996–1997, he was a Postdoctoral Fellow with Rice University, Houston, TX. Since August 1997, he has been with the University of Wisconsin, Madison, where he is currently an Assistant Professor of Electrical and Computer Engineering. His research interests are in wireless communications, sensor networks, statistical signal processing, wavelets, and time-frequency analysis.

Dr. Sayeed was the recipient of the National Science Foundation CAREER Award in 1999 and the Office of Naval Research Young Investigator Award in 2001. He served as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS from 1999 to 2002.