Reduced-Dimension MAP Turbo-BLAST Detection

Akrum Elkhazin, Konstantinos (Kostas) N. Plataniotis, and Subbarayan Pasupathy, *Fellow, IEEE*

Abstract—The Bell Labs layered space–time (BLAST) architecture is a simple and efficient multiantenna coding structure that can achieve high spectral efficiency. Many BLAST detectors require more receiver antennas than transmitter antennas. We propose two novel turbo-processing BLAST detectors that can operate in systems with fewer receiver antennas than transmitter antennas. Both detectors are based on the group-detection strategy. The first proposed detector, the reduced-dimension maximum *a posteriori* (RDMAP) detector uses a dynamically formed group for each bit decision, while the second proposed detector, the group maximum *a posteriori* (GMAP) uses a static grouping. For both detectors, a maximum *a posteriori* (MAP) decision is made using a group of transmitted symbols, and the remaining signal contribution is treated as interference. The interference is characterized as nonzero mean colored-noise source that is whitened before a decision is made. Both proposed detectors are generalizations of the MAP detector and the turbo-processing minimum mean-squared error (MMSE) detector in Sellathurai and Haykin, and Abe and Matsumoto. An uncoded bit-error rate analysis for an independent Rayleigh fading environment is also presented. Simulated results are presented which show that both the RDMAP and GMAP detectors have a performance improvement over the MMSE detector, especially in systems having an excess number of transmitter antennas.

Index Terms—Iterative methods, multidimensional signal detection, multiple-input multiple-output (MIMO) systems, nonlinear detection.

I. INTRODUCTION

THE use of multiple transmitter and receiver antennas can provide a tremendous capacity increase in wireless fading environments [1]. The Bell Labs layered space–time (BLAST) architecture is a simple and efficient coding structure that can take advantage of the multiple-input multiple-output (MIMO) channel capacity. A simple interference cancellation and nulling algorithm (ICNA) can be used in detection for both the original diagonal BLAST (D-BLAST) architecture and simplified vertical BLAST (V-BLAST) [4]. In systems with several more receiver antennas than transmitter antennas, even an uncoded V-BLAST system can largely take advantage of the available channel capacity [5]. ICNA, however, gives rise to increasingly poor performance as the number of receiver antennas approaches the number of transmitter antennas, and becomes infeasible for systems with an excess number of transmitter antennas. Systems having an excess number of transmitter antennas thus demand a more complex

detection structure, and exist in different scenarios. For example, in the downlink of a cellular systems, size limitations often make it infeasible to have a mobile station with many receiving antennas. A similar situation can exist when there is more than one transmitter and a single receiver, thus, the total number of transmitting antennas can easily exceed the number of receiving antennas.

There are several detection strategies that can be applied to systems having an excess number of transmitter antennas. An optimal solution is the maximum-likelihood (ML) detector. The ML detector unfortunately has an exponential complexity that grows with R^n, where R is the constellation size and n is the number of transmitter antennas. The ML detector can be implemented using the sphere-decoding algorithm [6] with no performance loss and a complexity reduction. Although the average case complexity of the sphere-decoding algorithm can be polynomial at high signal-to-noise ratio (SNR) [7], sphere decoding still has a exponential worst-case complexity. Suboptimal ML detectors have been applied to BLAST systems using tree-search algorithms [8] and group detection strategies [9], [10].

Turbo processing is an effective way to improve the performance of a BLAST receiver. Layered turbo processing receivers have been employed in [11]–[13], which consist of an iterative exchange of soft information between the detector and the decoders. The optimal soft-input soft-output (SISO) detector is the maximum *a posteriori* (MAP) detector, which, like the ML detector, has an exponential complexity that grows with R^n. For systems with a large number of transmitter antennas, a computationally more feasible minimum mean-squared error (MMSE) detector has been proposed, first in the multiuser context [14], [15], and later applied to BLAST [2], [3]. The detector discussed in [2] and [3] uses soft interference cancellation with an instantaneous MMSE filter that only suppresses residual interference that has not been cancelled. A similar detector is proposed in [16], which additionally employs a reduced-set MAP detector by forcing tentative decisions. For the detector in [2] and [3], much of the energy from interfering symbols is cancelled after the first iteration, and the instantaneous MMSE filter only needs to suppress energy from uncertain symbols. The performance of the SISO MMSE detector improves dramatically with successive iterations, but falls short of the MAP detector performance, especially in systems having a significant excess of transmitter antennas.

In this paper, we propose two novel SISO detectors, termed the reduced-dimension MAP (RDMAP) detector and group MAP (GMAP) detector, which are both based on a group detection strategy. These detectors bridge the performance gap between MAP and MMSE detectors in systems with an excess number of transmitter antennas, and are developed for quadrature amplitude modulation (QAM). The RDMAP and GMAP detectors operate by dividing the set of transmitted symbols into two groups: a MAP group and an interfering group. In the RDMAP detector, the MAP
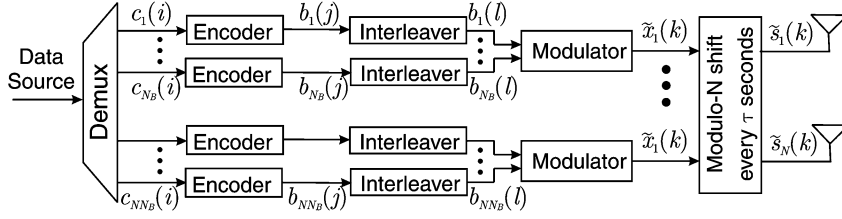
Fig. 1. Layered space–time transmitter.

and interfering groups are formed dynamically for each bit decision. The GMAP detector reduces the complexity of the RDMAP detector by using a static set of disjoint MAP groups and jointly produces a bit decision for each bit in a group. A novel greedy algorithm is proposed to form the MAP groups for the GMAP detector. For both detectors, the symbols in the interfering group are treated as an interfering noise source that is whitened by applying an appropriate filter. The prior probabilities for the interfering symbols are used to determine the mean of the interfering noise source. The size of the MAP group N_G is an adjustable parameter that determines the detector complexity. Through the choice of this parameter, we show that the RDMAP and GMAP detectors are a generalization of both the MAP detector and MMSE detector in [2] and [3].

In this paper, we develop the RDMAP and GMAP detectors for a BLAST transmission across a flat-fading channel. These detectors can also be applied to threaded space–time (TST) codes [17] and universal space–time codes [18] that have recently been developed as generalizations of BLAST. In addition, the RDMAP and GMAP can be easily extended to the frequency-selective channel, following the formulation in [3] by stacking symbols from different time instances together. Finally, we distinguish our group detection strategy from that in [9], as the solution in [9] does not use a noise-whitening filter, and from the solution in [10], as we incorporate prior information into the whitening filter.

An uncoded bit-error rate (BER) analysis of the RDMAP and GMAP detectors is also presented in this paper for an independent Rayleigh fading channel. By approximating the MAP decision rule with the ML decision rule, the union bound on the symbol-error rate (SER) probability can be expressed as the sum of Hermitian quadratic forms in complex Gaussian variables. The BER is obtained from the probability density function (pdf) of the summed Hermitian quadratic forms.

The remainder of this paper is organized as follows. Section II provides a system model that includes the layered transmitter, channel model, and turbo-processing receiver structure. Section III describes the detector design, starting with the MAP detector, followed by the RDMAP and GMAP detectors. An uncoded BER analysis is developed in Section IV, with a complexity analysis and simulated results in Section V. Finally, Section VI contains a summary and concluding remarks.

II. SYSTEM MODEL

A. Layered Space–Time Transmitter

Consider the transmitter structure in Fig. 1 for a layered space–time architecture [2] having N transmitter antennas. Binary data is demultiplexed into NN_B bit streams that are independently encoded and interleaved, where N_B is the number of bits per symbol to be modulated on the in-phase and quadrature phase signals. Let $\{c_n(i)\}$, $\{b_n(j)\}$, and $\{b_n(l)\}$ denote the the set of raw binary data, channel coded output, and interleaver output, respectively, for the nth bit stream. The change of indexing variable for j to l indicates the interleaving process. Groups of N_B bit streams are then modulated using an M-ary Gray coded modulation function. The nth modulated output or layer is given by $x_n(k) = M_{\text{ary}}(b_{(n-1)N_B+1}(2k)\ldots b_{nN_B}(2k)) + \sqrt{-1}M_{\text{ary}}(b_{(n-1)N_B+1}(2k+1)\ldots b_{nN_B}(2k+1))$, where M_{ary} is the Gray-coded M-ary modulating function whose output is $\{-2^{N_B-1}-1, -2^{N_B-1}+3, \ldots, 2^{N_B-1}+1\}$. For quaternary phase-shift keying (QPSK), the nth-layer modulated output is given by $x_n(k) = \{2b_n(2k)-1\} + \sqrt{-1}\{2b_n(2k+1)-1\}$. The modulated output from each layer is passed through a modulo-N shifter that changes every τ seconds, and with no loss of generality, we set τ equal to the symbol period. The transmitted signal on antenna n is given by $\tilde{s}_n(k) = \tilde{x}_\alpha(k)$, $\alpha = (n-k) \bmod N$.

B. Channel Model

We consider a flat-fading MIMO channel with N transmitter antennas and M receiver antennas sampled at symbol rate. The channel output at the mth receiver antenna at the kth time index is given by

$$\tilde{r}_m(k) = \sum_{n=1}^{N} \tilde{h}_{mn}\tilde{s}_n(k) + \tilde{v}_m(k) \tag{1}$$

where $\tilde{h}_{mn}$ is the channel gain between the nth transmitter antenna and the mth receiver antenna, and $\tilde{v}_m(k)$ is an additive white Gaussian noise (AWGN) source of variance σ^2. We define the SNR at the channel output as $\text{SNR} = \sum_{n=1}^{N} E|\tilde{h}_{mn}|^2 E_s/\sigma^2 = NE|\tilde{h}_{mn}|^2 E_s/\sigma^2$, where $E_s = E[|\tilde{s}_n(k)|^2] = 2$ for QPSK. Stacking measurements in the spatial domain results in

$$\tilde{\mathbf{r}}(k) \equiv [\tilde{r}_1(k), \ldots, \tilde{r}_M(k)]^T = \hat{\mathbf{H}}\tilde{\mathbf{s}}(k) + \tilde{\mathbf{v}}(k) \tag{2}$$

where $\{\hat{\mathbf{H}}\}_{mn} = \tilde{h}_{mn}$, $\tilde{\mathbf{s}}(k) = [\tilde{s}_1(k), \ldots, \tilde{s}_N(k)]^T$ and $\tilde{\mathbf{v}}(k) = [\tilde{v}_1(k), \ldots, \tilde{v}_M(k)]^T$. It is convenient to transform the complex channel equation in (2) into the real matrix equation

$$\mathbf{r}(k) = \mathbf{H}\mathbf{s}(k) + \mathbf{v}(k) \tag{3}$$

where $\mathbf{r}(k) = [\Re\{\tilde{\mathbf{r}}^T(k)\}\Im\{\tilde{\mathbf{r}}^T(k)\}]^T$, $\mathbf{s}(k) = [\Re\{\tilde{\mathbf{s}}^T(k)\}\Im\{\tilde{\mathbf{s}}^T(k)\}]^T$, $\mathbf{v}(k) = [\Re\{\tilde{\mathbf{v}}^T(k)\}\Im\{\tilde{\mathbf{v}}^T(k)\}]^T$, and

$$\mathbf{H} = \begin{bmatrix} \Re\{\tilde{\mathbf{H}}\} & -\Im\{\tilde{\mathbf{H}}\} \\ \Im\{\tilde{\mathbf{H}}\} & \Re\{\tilde{\mathbf{H}}\} \end{bmatrix} \tag{4}$$
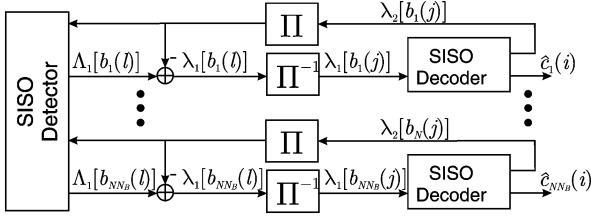
Fig. 2.   Turbo-processing layered receiver.

is the $M' \times N'$ real channel matrix with $M' = 2M$ and $N' = 2N$.

C. Turbo BLAST Receiver

The block diagram for the turbo-processing BLAST receiver is shown in Fig. 2. The receiver consists of a SISO symbol detector, a set of NN_B SISO channel decoders, and an interleaver and deinterleaver between each decoder and the detector. There are modulo-N shifters at the input and output of the detector that have been omitted from Fig. 2 for clarity. In each iteration, the detector produces an *a posteriori* probability (APP) for each coded bit in the form of a log-likelihood ratio (LLR) that is given by

$$\Lambda_1[b_n(l)] = \log \frac{P(b_n(l) = 1|\mathbf{r}(k))}{P(b_n(l) = 0|\mathbf{r}(k))} \equiv \lambda_1[b_n(l)] + \lambda_2^p[b_n(l)] \tag{5}$$

where $\lambda_1[b_n(l)]$ is the *extrinsic* information that is fed to the channel decoder for the nth bit stream, and $\lambda_2^p[b_n(l)]$ is the *a priori* information provided by the nth channel decoder. The channel decoders produce an extrinsic LLR for each coded bit as

$$\lambda_2[b_n(j)] = \log \frac{P(b_n(j) = 1|\lambda_1[b_n(i)], i \neq j)}{P(b_n(j) = 0|\lambda_1[b_n(i)], i \neq j)}. \tag{6}$$

After a sufficient number of iterations, estimates for the uncoded bits can be obtained as

$$\hat{c}_n(i) = \text{sgn}\left(\log \frac{P(c_n(i) = 1|\lambda_1[b_n(j)])}{P(c_n(i) = 0|\lambda_1[b_n(j)])} \right). \tag{7}$$

The channel decoders can be efficiently implemented using a SISO APP module [19].

III. DETECTOR DESIGN

The optimal APP detector can be derived by evaluating (5) over all possible symbol vectors. From (3), the signal component of the real-value channel observation $\mathbf{r}(k)$ is determined by a symbol vector $\mathbf{s}(k)$, and is formed from a bit vector $\mathbf{b} = [b_1, \ldots, b_{NN_B}]^T$. Let S be the set of possible $\mathbf{s}(k)$, and let S_i^n be a subset of S having $b_n = i, i \in \{0, 1\}$. The APP for a coded bit b_n can be evaluated in the log domain as

$$\Lambda_1[b_n] = \log \frac{P(b_n = 1|\mathbf{r}(k))}{P(b_n = 0|\mathbf{r}(k))}$$

$$= \log \frac{\sum_{\mathbf{s} \in S_1^n} P(\mathbf{s}) \exp\left(\frac{-\|\mathbf{r}(k) - \mathbf{H}\mathbf{s}\|^2}{2\sigma^2} \right)}{\sum_{\mathbf{s} \in S_0^n} P(\mathbf{s}) \exp\left(\frac{-\|\mathbf{r}(k) - \mathbf{H}\mathbf{s}\|^2}{2\sigma^2} \right)}. \tag{8}$$

The evaluation of (8) can be computationally prohibitive for systems with a large number of transmitter antennas. There are a

total $|S| = 2^{2NN_B}$ terms to be summed in the numerator and denominator of (8). In Sections III-A and B, we derive two reduced-complexity detectors that are based on reducing the size of the symbol vector $\mathbf{s}$.

A. Reduced-Dimension MAP Detector

In order to reduce the number of terms to be summed in the MAP decision, we perform a MAP decision using a subset of the elements of the symbol vector $\mathbf{s}(k)$ and treat the contribution of the remaining signal contribution as interference. For a signal vector $\mathbf{s} = \mathbf{s}(k)$, define the MAP group as the set $G = \{\alpha_1, \ldots, \alpha_{N_G}\}$ of N_G integers corresponding to indexes of elements in $\mathbf{s}$ used in the MAP decision. Further define an interfering group as a complementary set $\bar{G} = \{\beta_1, \ldots, \beta_{N_{\bar{G}}}\}$ of $N_{\bar{G}}$ integers corresponding to indexes of elements in $\mathbf{s}$ such that $G \bigcap \bar{G} = \emptyset$ and $G \bigcup \bar{G} = \{1, \ldots, N'\}$. For a particular choice of G, the channel output can be expressed as

$$\mathbf{r} = \mathbf{H}_G \mathbf{s}_G + \mathbf{H}_{\bar{G}} \mathbf{s}_{\bar{G}} + \mathbf{v} \tag{9}$$

where $\mathbf{s}_G = [s_{\alpha_1}, \ldots, s_{\alpha_{N_G}}]^T$ is the reduced-dimension signal vector, $\mathbf{s}_{\bar{G}} = [s_{\beta_1}, \ldots, s_{\beta_{N_{\bar{G}}}}]^T$ is the interference vector, $\mathbf{H}_G = [\mathbf{h}_{\alpha_1}, \ldots, \mathbf{h}_{\alpha_{N_G}}]$, $\mathbf{H}_{\bar{G}} = [\mathbf{h}_{\beta_1}, \ldots, \mathbf{h}_{\beta_{N_{\bar{G}}}}]$, and $\mathbf{h}_i$ is the ith column of $\mathbf{H}$. The time index k has been omitted for clarity. The contribution of the interference and Gaussian noise can be treated as a colored noise source. Let $\mathbf{w} = \mathbf{H}_{\bar{G}} \mathbf{s}_{\bar{G}} + \mathbf{v}$ be the colored noise source whose mean is $\bar{\mathbf{w}} = \text{E}[\mathbf{w}] = \mathbf{H}_{\bar{G}} \hat{\mathbf{s}}_{\bar{G}}$, where $\hat{\mathbf{s}}_{\bar{G}} = [\hat{s}_{\bar{G}1}, \ldots, \hat{s}_{\bar{G}N_{\bar{G}}}]$ and $\hat{s}_{\bar{G}i} = \text{E}[s_{\bar{G}i}]$ is evaluated using the prior probabilities from the channel decoders as

$$\text{E}[s_{\bar{G}i}] = \sum s_{\bar{G}i} P(s_{\bar{G}i}). \tag{10}$$

For a symbol mapping $s = M_{\text{ary}}(c_1 \ldots c_{N_B})$, the symbol probability is $P(s) = \prod_{n=1}^{N_B} P(b_n = c_n)$, with $P(b_n = 0) = 1/(1 + \exp(\lambda_2^p[b_n]))$. For QPSK modulation, (10) reduces to $\text{E}[s] = \tanh(\lambda_2^p[b]/2)$. The covariance of $\mathbf{w}$ is given by

$$\mathbf{R}_w = \text{E}[(\mathbf{w} - \bar{\mathbf{w}})(\mathbf{w} - \bar{\mathbf{w}})^T] = \mathbf{H}_{\bar{G}} \mathbf{\Omega} \mathbf{H}_{\bar{G}}^T + \frac{\mathbf{I}\sigma^2}{2} \tag{11}$$

where $\Omega = \text{diag}(\omega_1^2, \ldots, \omega_{N_{\bar{G}}}^2)$ and $\omega_i^2 = \text{E}[|s_{\bar{G}i} - \hat{s}_{\bar{G}i}|^2] = \text{E}[s_{\bar{G}i}^2] - \text{E}[s_{\bar{G}i}]^2$. For QPSK, $\omega_i^2 = 1 - \hat{s}_{\bar{G}i}^2$. We have assumed perfect interleaving in (11) such that the elements of $\hat{\mathbf{s}}_{\bar{G}}$ are independent. The noise $\mathbf{w}$ can be whitened by first removing the mean $\bar{\mathbf{w}}$ and then applying an appropriate noise-whitening filter $\mathbf{F} = \mathbf{\Sigma}^{-1/2} \mathbf{Q}^T$, where $\mathbf{\Sigma}$ is a diagonal matrix and $\mathbf{Q}$ is an orthogonal matrix, both obtained from the eigenvalue decomposition of $\mathbf{R}_w = \mathbf{Q}\mathbf{\Sigma}\mathbf{Q}^T$, $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$. The whitened channel observation is given by

$$\mathbf{y} = \mathbf{F}(\mathbf{r} - \bar{\mathbf{w}}). \tag{12}$$

Under the assumption the noise at the filter output is Gaussian, the APP for the bit b_n corresponding to the nth bit in the mapping $\mathbf{s}_G = [M_{\text{ary}}(b_1, \ldots, b_{N_G}), \ldots, M_{\text{ary}}(b_{1+(N_G-1)N_B}, \ldots, b_{N_G N_B})]^T$ can be evaluated as

$$\Lambda_1[b_n] = \log \frac{P(b_n = 1|\mathbf{y})}{P(b_n = 0|\mathbf{y})}$$

$$= \log \frac{\sum_{\mathbf{s}_G \in S_{G1}^n} P(\mathbf{s}_G) \exp\left(\frac{-\|\mathbf{y} - \mathbf{F}\mathbf{H}_G \mathbf{s}_G\|^2}{2} \right)}{\sum_{\mathbf{s}_G \in S_{G0}^n} P(\mathbf{s}_G) \exp\left(\frac{-\|\mathbf{y} - \mathbf{F}\mathbf{H}_G \mathbf{s}_G\|^2}{2} \right)} \tag{13}$$

where S_{Gi}^n is the set of possible $\mathbf{s}_G$ having $b_n = i$. The noise at the filter output is uncorrelated, but it is not Gaussian, since the interference has a non-Gaussian pdf. This is especially true at high SNR, when the noise contribution is almost entirely from inference. The APP in (13) tends to be more confident than the true APP, because of the non-Gaussian noise at the filter output. An overconfident APP can cause the turbo receiver to diverge. In order to prevent divergence, we prescale the prior LLRs according to $\lambda_2^p[b] \leftarrow \text{sgn}(\lambda_2^p[b])|\lambda_2^p[b]|^\kappa$, where $0 < \kappa \leq 1$. Prescaling the prior LLRs leads to an overestimation of the noise power from interfering symbols and compensates for the non-Gaussian noise pdf. The prescaling function tends to scale down large LLR values more than small LLR values, as incorrect large LLR values are more likely to cause divergence.

The choice of the groups G and $\bar{G}$ is critical to the performance of the RDMAP detector. Consider forming G and $\bar{G}$ for bit decisions corresponding to a symbol s_i. The biggest factor affecting the bit decision in (13) is the noise enhancement or signal attenuation associated with the noise-whitening filter $\mathbf{F}$. The noise-whitening filter scales down the channel output in directions where the noise power is high, and scales up the channel output in directions where the noise power is low, while maintaining an uncorrelated noise output. In order to minimize signal attenuation, the group $\bar{G}$ should be formed, such that the noise power in the direction of the desired signal $\mathbf{h}_i$ is minimized. In the channel observation space, the noise power from an interfering symbol s_j, $j \neq i$ projected onto the subspace of s_i is given by

$$\varepsilon_j^2 = \left|\mathbf{h}_i^T\mathbf{h}_j\right|^2 \omega_j^2. \tag{14}$$

The value of ε_j^2 is the projected noise variance caused by s_j if it were considered in isolation. Although the total noise power in the subspace of s_i is not simply the sum of ε_j^2's, it still remains that symbols having a high ε_j^2 will lead to high noise enhancement if they are placed in the interfering group. Accordingly, $\bar{G}$ is formed by the $N_{\bar{G}}$ smallest ε_j^2's, and G is formed by the N_G largest ε_j^2's.

We now consider the MAP detector and the MMSE detector in [2] and [3] as special cases of the RDMAP detector. Consider the following lemmas.

Lemma 1: The RDMAP detector and MAP detector are equivalent for $N_G = N'$.[1]

Lemma 2: The RDMAP detector and MMSE detector in [2] and [3] are equivalent for $N_G = 1$.

B. Group MAP Detector

The bulk of the the RDMAP detector complexity lies in evaluating $\mathbf{r}_G = \mathbf{F}\mathbf{H}_G\mathbf{s}_G$ for all $\mathbf{s}_G \in S_G$ in (13). The RDMAP detector uses each $\mathbf{r}_G$ only once to produce a single bit decision. A bit decision could, however, be produced for each element of $\mathbf{s}_G$ that would lead to an approximate complexity reduction of $1/N_G$. This is the motivation behind the GMAP detector.

For a group G, the GMAP detector produces a log-domain APP according to (13) for each bit b_n, $i = 1, \ldots, N_G N_B$ corresponding to the elements of $\mathbf{s}_G$. With no loss of generality, we assume that the symbol vector $\mathbf{s}$ can be divided into N_Ψ disjoint

groups of equal size, denoted $\Psi = \{G_1, \ldots, G_{N_\Psi}\}$, such that $a \neq b \forall a \in G_i, b \in G_j, i \neq j$. At this point, it can be noted that the RDMAP and GMAP detectors are equivalent in the limiting cases, when $N_G = 1$ and $N_G = N'$. Based on *Lemmas 1* and *2*, it can be concluded that the GMAP detector is also a generalization of both the MAP and MMSE detectors in [2] and [3].

What remains in the development of the GMAP detector is an algorithm to choose the groups in Ψ. Like the RDMAP detector, we propose a correlation-based grouping algorithm, but require a modified weighting criterion. Consider, for example, a situation in which there is a small angle between the vectors $\mathbf{h}_i$ and $\mathbf{h}_j$, but the magnitude of $\mathbf{h}_j$ and/or ω_j^2 is small. The RDMAP detector would likely include s_i in the MAP decision group for b_j, but not include s_j in the MAP decision group for b_i. In a joint detection structure, we require a cost metric that is symmetric to whether a decision on b_i is made with interference from s_j or *vice versa*. To this end, let $\mathbf{R}$ be the normalized correlation matrix whose entry in row i and column j is given by

$$r_{ij} = \frac{\left|\mathbf{h}_i^T\mathbf{h}_j\right|}{\sqrt{\|\mathbf{h}_i\|^2\|\mathbf{h}_i\|^2}}. \tag{15}$$

The element r_{ij} is the normalized correlation between the symbols s_i and s_j at the channel output. We note that if the ω_j^2 in (14) were used as weights, they would be cancelled in the normalization in (15).

Given a normalized correlation matrix $\mathbf{R}$, consider forming Ψ using the following objective function:

$$\Theta = \arg\max_{G_1 \ldots G_{N_\Psi}} \sum_{k=1}^{N_\Psi} \arg\max_{j \in G_k} r_{ij} \tag{16}$$

which is equivalent to maximizing the maximum pairwise correlation among the members of each group, averaged over all groups. We use the maximum pairwise correlation criteria, instead of, for example, the average pairwise correlation, since the former relates to the minimum distance between columns of $\mathbf{H}$. At a high SNR, the BER is usually largely determined by the minimum distance between constellation points, instead of average distances. An exact solution to (16) requires an exhaustive search over all possible groupings, and is computationally infeasible for large N. An approximate solution to (16) can be found using the following greedy algorithm.

```
1   R ← R − I;
2   Ψ ← {G_1, …, G_{N'}}, G_i = {i};
3   while |Ψ| > N_Ψ
4     if ∑_{i=1}^{|Ψ|} I(|G_i| > 1) < N_Ψ
5       δ_ij = I(|G_i| + |G_j| ≤ N_G);
6     else
7       δ_ij = I(2 < |G_i| + |G_j| ≤ N_G);
8     end if
9     r_kl = arg max_{ij} r_ij δ_ij;
10    G_k ← G_k ∪ G_l;
11    Ψ ← Ψ \ {G_l};
12    for n = 1 … |Ψ|, n ≠ k;
13      r_nk = r_kn = max{r_nk, r_nl};
14    end for
15    Remove kth row and column from R
16  end while
```

---
[1] The proofs for all lemmas can be found in the Appendix.

where $I(x)$ is an indicator function that evaluates to 1 if x is true, and 0, otherwise. The preceding algorithm starts with N' groups with one element in each, and finishes with N_Ψ groups with N_G elements in each group. At each iteration, two groups are merged. The list of possible groups mergers is determined in lines 4–8, and a greedy decision is made to merge the two groups having maximum pairwise correlation. The preceding algorithm is greedy, as it attempts to form highly correlated groups first. Although this may not optimally satisfy the objective function in (16), it is advantageous to form highly correlated groups first, since the BER is more strongly influenced by groups with a high pairwise correlation, instead of by the average pairwise correlation across all groups.

IV. BER ANALYSIS

This section presents an analysis of the uncoded BER for the RDMAP and GMAP detectors in a Rayleigh fading environment. We make the assumption that the channel coefficients $\{\tilde{h}_{mn}\}$ have an independent Rayleigh fading distribution with unity variance, $E[\|\tilde{h}_{mn}\|^2] = 1$. Our analysis is based on expressing SER as a Hermitian quadratic form in complex Gaussian random variables. A similar approach has been used in the context of multiuser detection to analyze the performance of the ML detector [20] and the optimal ratio combiner [21].

The RDMAP and GMAP detectors produce an *a posteriori* bit probability in (13), with error probability $P_e = P(\Lambda_1(b_n) < 0|b_n = 1)$, where $\Lambda_1(b_1)$. The summations in the numerator and denominator of (13) are dominated by the term corresponding to the s_G of the closest Euclidean distance to the channel observation in the noise-whitened space. Approximating the summation operators in (13) with max operators yields the ML decision rule, whose decision-error probability is given by

$$P(\Lambda_1(b_n) < 0|b_n = 1)$$

$$\approx P\left(\log \frac{\max_{s \in S^n_{G1}} \exp\left(\frac{-\|y - FH_G s\|^2}{2}\right)}{\max_{s \in S^n_{G0}} \exp\left(\frac{-\|y - FH_G s\|^2}{2}\right)} < 0\right) \quad (17)$$

$$= P\left(\min_{s \in S^n_{G1}} \|y - Gs\|^2 > \min_{s \in S^n_{G0}} \|y - Gs\|^2\right) \quad (18)$$

$$= P\left(s \in S^n_{G0}| \min_{s \in S^n_G} \|y - Gs\|^2\right) \quad (19)$$

where $G = FH_G$ and S_G is the set of all s_G. The ML decision rule chooses the group vector that is closest to the received vector in the noise-whitened space. For a transmitted group vector $s_i \in S^n_{G1}$, an erroneous decision is made if the ML detector chooses an erroneous symbol vector $s_j \in S^n_{G0}$, which will occur if $C_j < C_i$, where C_i is the cost metric associated with a candidate group vector s_i defined by $C_i \triangleq \|y - Gs_i\|^2$. The probability of this pairwise error event is denoted $P(D_{ij} < 0|s_i)$, where $D_{ij} = C_j - C_i$.

The union bound on the probability of a detection error for a transmitted vector s_i is given by the sum of the pairwise error probability over the set of received vectors $s_j \in S^n_{G0}$. Averaging over the set of transmitted vectors yields an upper bound of the error probability that is given by

$$P_e = E_{s_i \in S^n_{G1}} \left[\sum_{s_j \in S^n_{G0}} P(D_{ij} < 0 \mid s_i)\right]. \quad (20)$$

In order to evaluate (20), we will express D_{ij} as a Hermitian quadratic form in complex Gaussian variables. Although it is easier to represent D_{ij} in real-valued Hermitian quadratic form, such an expression does not allow for easy factorization, as will be discussed later. We first require a representation of the group detector derived in Section III in complex variables.

The real channel matrix consists of orthogonal pairs of column vectors h_n, h_{n+N}. Since both the RDMAP and GMAP detectors form the MAP group G, based on the correlation between columns, it is unlikely that h_n and h_{n+N} will both be in the H_G. Under this assumption, each column in H_G is orthogonal to one column in $H_{\bar{G}}$ and approximately independent of the remaining columns of H_G, $H_{\bar{G}}$. The independence would be exact if the MAP group was formed by randomly choosing independent columns of H. If we project the set of interfering sources $\{s_{\beta_n} h_{\beta_n}\}$ onto a column h_{α_i} of H_G, there will be one interfering source with no noise contribution, and a noise contribution from the remaining $N' - N_G - 1$ interfering sources, each of which has a signal subspace defined by $\{h_{\beta_n}\}$, that is independent of h_{α_i}. If we apply the same projection on another column h_{α_j} of H_G, there will be a different interfering source with no noise contribution and a noise contribution from the remaining $N' - N_G - 1$ sources. In order to facilitate analysis, we assume there are $N_I = N' - N_G - 1$ interfering sources whose signal subspace is independent of H_G, leading to an approximate channel model

$$r \approx H_G s_G + H_I s_I + v \quad (21)$$

where $s_I = [s_{I1}, \ldots, s_{IN_I}]^T$ is an interference vector that consists of independent elements of unity variance, and H_I is a $M' \times N_I$ Gaussian matrix that is independent of H_G. Since the grouping criterion for the RDMAP and GMAP detectors is based on the assumption of N_I independent interfering sources, we expect the uncoded BER of both detectors to be quite similar. At this point, we consider the case where N_I is even, letting $N_{Ic} = N_I/2$, and impose a structure on H_I of the form

$$H_I = \begin{bmatrix} H_{Ir} & -H_{Ii} \\ H_{Ii} & H_{Ir} \end{bmatrix} \quad (22)$$

where H_{Ir}, H_{Ii} are $M \times N_{Ic}$ matrices of independent Gaussian entries with variance $1/2$. The structure imposed in (22) is needed to represent the real-valued system in complex variables. One may note that the $N' - 2N_G$ columns of $H_{\bar{G}}$ that are not orthogonal to H_G can be represented in a form similar to (22). Using (22), the real-valued channel model in (21) can be equivalently represented in complex variables as

$$\tilde{r} = H_{Gc} s_G + H_{Ic} s_{Ic} + \tilde{v} \quad (23)$$

where $H_{Gc} = H_{Gr} + iH_{Gi}$, such that $H_G = [H_{Gr}^T, H_{Gi}^T]^T$, $H_{Ic} = H_{Ir} + iH_{Ii}$, and $s_{Ic} = [s_{I1}, \ldots, s_{I\{N_{Ic}\}}]^T + i[s_{I\{N_{Ic}+1\}}, \ldots, s_{IN_I}]^T$. The complex noise-correlation

matrix is $\mathbf{R}_{wc} = \mathbf{H}_{Ic}(2\mathbf{I}_{N_{Ic}})\mathbf{H}_{Ic}^H + \mathbf{I}_M\sigma^2$, and the complex noise-whitening filter is $\mathbf{F_c} = \mathbf{\Sigma}_c^{-1/2}\mathbf{Q}_c^H$, where $\mathbf{\Sigma}_c = \text{diag}[\lambda_1, \ldots, \lambda_M]$ is a real-valued diagonal matrix, and $\mathbf{Q}_c$ is an orthogonal matrix, both obtained from the eigenvalue decomposition of $\mathbf{R}_{wc} = \mathbf{Q}_c\mathbf{\Sigma}_c\mathbf{Q}_c^H$, $\mathbf{Q}_c\mathbf{Q}_c^H = \mathbf{I}$. With no loss of generality, we assume the elements on the diagonal of $\mathbf{\Sigma}_c$ are in ascending order. The noise-whitened output is given by

$$\mathbf{y}_c = \mathbf{\Sigma}_c^{-1/2}\mathbf{Q}_c^H\mathbf{H}_{Gc}\mathbf{s}_G + \mathbf{z}_c \tag{24}$$

where $\mathbf{z}_c$ is a length-M vector of independent elements of unity variance. For the sake of analysis, we further assume that the elements of $\mathbf{z}_c$ have a complex Gaussian distribution. The matrix $\mathbf{H}_{Gc}$ is rotationally invariant, as it has independent Gaussian elements. Accordingly, multiplying $\mathbf{H}_{Gc}$ by the orthonormal matrix $\mathbf{Q}_c^H$ does not change its distribution. The combined channel and whitening-filter response in (24) is a function of the random matrix $\mathbf{H}_{Gc}$ and the random eigenvalue matrix $\mathbf{\Sigma}_c$. We approximate the random eigenvalue matrix $\mathbf{\Sigma}_c$ with its expectation $\hat{\mathbf{\Sigma}}_c = \text{diag}[\text{E}[\lambda_1], \ldots, \text{E}[\lambda_M]]$ in (24), which leads to

$$\mathbf{y}_c \approx \mathbf{G}_c\mathbf{s}_G + \mathbf{z}_c \tag{25}$$

where $\mathbf{G}_c$ is a $M \times N_G$ matrix of zero-mean independent complex Gaussian elements, whose element $\{\mathbf{G}\}_{mn}$ has a variance of $1/\text{E}[\lambda_m]$. The mean eigenvalue matrix $\hat{\mathbf{\Sigma}}_c$ can be obtained in one of several ways, as discussed in Appendix.

Thus far, we have developed the complex noise-whitened channel output in (25) under the assumption that the N_I is even. The case where N_I is odd corresponds to having purely real noise with variance $1/2$ for one element of $\mathbf{s}_{ic}$. Although the mean eigenvalue matrix $\hat{\mathbf{\Sigma}}_c$ can be evaluated for $N_{Ic} = N_I/2$ with an odd N_I, this may not correspond to having real noise with variance $1/2$ in one dimension. The appropriate value of N_{Ic} is, however, somewhere between $N_{Ic} = [\lfloor N_I/2 \rfloor, \lceil N_I/2 \rceil]$. By considering many different values of M, N, and N_G, it was found that

$$N_{Ic} = \begin{cases} \frac{N_I}{2}, & N_I \text{ even} \\ \lfloor \frac{N_I}{2} \rfloor + 0.2, & N_I \text{ odd} \end{cases} \tag{26}$$

produced a BER curve that had approximately the same slope as the simulated BER at high SNR.

In terms of the complex noise-whitened channel output in (25), the ML detector cost metric associated with a group vector $\mathbf{s}_i \in S_G$ can be expressed as $C_i = \|\mathbf{y}_c - \mathbf{G}_c\mathbf{s}_i\|^2$. The pairwise error metric D_{ij} can be expressed as the sum of M Hermitian quadratic forms

$$D_{ij} = \sum_{m=1}^{M} \mathbf{x}_m^H\mathbf{F}_{ij}\mathbf{x}_m \tag{27}$$

where $\mathbf{x}_m$ is a length $N_G + 1$ Gaussian vector defined by $\mathbf{x}_m = [y_m, g_{m1}, \ldots, g_{mN_G}]^T$, $\mathbf{F}_{ij}$ is a Hermitian matrix defined by $\mathbf{F}_{ij} = \mathbf{u}_j\mathbf{u}_j^T - \mathbf{u}_i\mathbf{u}_i^T$, and $\mathbf{u}_i = [1, -\mathbf{s}_i^T]^T$. Each term $d_{ijm} = \mathbf{x}_m^H\mathbf{F}_{ij}\mathbf{x}_m$ in (27) is a Hermitian quadratic form in $N_G + 1$ zero-mean complex Gaussian random variates, whose pdf has a two-sided Laplace transform given by [22], [23, App. B]

$$\phi_{d_{ijm}}(s) = \frac{1}{\det(\mathbf{I} + s\mathbf{R}_m\mathbf{F}_{ij})} = \frac{1}{\prod_{k=1}^{M+1}(1 + s\gamma_{mk})} \tag{28}$$

where $\mathbf{R_m} = \text{E}[\mathbf{x}_m\mathbf{x}_m^H \mid \mathbf{s}_i]$ is the covariance matrix given by

$$\mathbf{R}_m = \begin{bmatrix} 1 + \frac{N_G}{\lambda_m} & \frac{s_{i1}}{\lambda_m} & \frac{s_{i2}}{\lambda_m} & \cdots & \frac{s_{iN_G}}{\lambda_m} \\ \frac{s_{i1}}{\lambda_m} & \frac{1}{\lambda_m} & 0 & \cdots & 0 \\ \frac{s_{i2}}{\lambda_m} & 0 & \frac{1}{\lambda_m} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{s_{iN_G}}{\lambda_m} & 0 & 0 & \cdots & \frac{1}{\lambda_m} \end{bmatrix} \tag{29}$$

and γ_{mk} is the kth eigenvalue of $\mathbf{R}_m\mathbf{F}_{ij}$. As discussed in [20], $\mathbf{R}_m$ is, in general, full rank, and $\mathbf{F}_{ij}$ has at most two nonzero eigenvalues, thus $\mathbf{R}_m\mathbf{F}_{ij}$ has, in general, two nonzero eigenvalues denoted by γ_{m1}, γ_{m2}. Since the Gaussian vectors $\mathbf{x}_m$ are independent, the Laplace transform of the pdf of D_{ij} can be expressed as the product of $\phi_{d_{ijm}}(s)$ given by

$$\Phi_{D_{ij}}(s) = \prod_{m=1}^{M} \phi_{d_{ijm}}(s) = \frac{1}{\prod_{m=1}^{M}(1 + s\gamma_{m1})(1 + s\gamma_{m2})} \tag{30}$$

with the region of convergence being the vertical strip enclosing the $j\omega$ axis bounded by the closest pole on either side. The function $\Phi_{D_{ij}}(s)$ has up to $2M$ distinct poles at points $p = -1/\gamma_{mk}$. Let p_k, $k = 1, \ldots, N_p$ denote the set of N_p distinct poles of $\Phi_{D_{ij}}(s)$, and let n_k be the multiplicity of pole p_k. Through a partial fraction expansion, $\Phi_{D_{ij}}(s)$ can be expressed as

$$\Phi_{D_{ij}}(s) = \sum_{k=1}^{N_p}\sum_{l=1}^{n_k} \frac{a_{kl}}{(s - p_k)^l} \tag{31}$$

where

$$a_{kl} = \frac{1}{l!}\lim_{s \longrightarrow p_k}\left(\frac{d}{ds}\right)^k \left(\Phi_{D_{ij}}(s)(s - p_k)^{n_k}\right). \tag{32}$$

At this point, it is worth justifying the need for a complex noise-whitened channel output. If a real output were used, the $(1 + s\gamma_{mk})$ factors in the denominator of (30) would be replaced by factors of the form $\sqrt{1 + s\gamma_{mk}}$ [22], which would not allow for the expansion in (31). From tables of inverse Laplace transforms, the pdf of D_{ij} is given by

$$P(x) = \sum_{k=1}^{N_p}\sum_{l=1}^{n_k} a_{kl}\text{sgn}(-p_k)\frac{x^{l-1}}{(l-1)!}\exp(p_kx)u(\text{sgn}(-p_k)t) \tag{33}$$

where $\text{sgn}(x)$ is 1 if $x \geq 0$, and -1 otherwise, and $u(x)$ is 1 if $x \geq 0$, and 0, otherwise. Integrating (33) from $-\infty$ to 0 yields

$$P(D_{ij} < 0 \mid \mathbf{s}_i) = \sum_{k=1}^{N_p}\sum_{l=1}^{n_k} \frac{a_{kl}}{(-p_k)^l(l-1)!}\Gamma(l)I(p_k > 0) \tag{34}$$

which can be substituted back into (20) to evaluate the BER.

V. SIMULATED RESULTS

This section analyzes the coded and uncoded BER performance of the RDMAP and GMAP detectors, along with the complexity of both detectors. The analysis in the previous section of the uncoded BER is also examined. The MAP and MMSE detectors in [2] and [3] are used as references for comparison, in terms of complexity and performance.

A complexity analysis for the RDMAP and GMAP detectors is shown in Table I, along with the complexity of the MAP and MMSE [2], [3] detectors for reference. All operations

TABLE I
APPROXIMATE COMPLEXITY OF MAP, MMSE, RDMAP, AND GMAP DETECTORS FOR EACH CODED BIT DECISION

	MAP	MMSE	RDMAP	GMAP
mult	$2^{N'N_B}(M'+N')$	$M'^2 N'$	$M'^{N_{\bar G}} + 2^{N_G N_B} M' N_G$	$M'^2 \frac{N_{\bar G}}{N_G} + 2^{N_G N_B} M'$
add	$2^{N'}M'$	$M'^2 N'$	$M'^2 N_{\bar G} + 2^{N_G N_B} M' N_G$	$M'^2 \frac{N_{\bar G}}{N_G} + 2^{N_G N_B} M'$
inv		$O(M'^3)$		
eig			$O(M'^3)$	$O(M'^3)/N_G$

are shown for a single bit decision. The number of multiplications (mult) and additions (add) is approximate, since only the highest polynomial term of M', N', N_G, etc., is shown for clarity. The number of elementary operations involved in a matrix inverse (inv) and eigenvalue decomposition (eig) is difficult to evaluate, thus, the complexity is expressed in $O(n)$ notation. It was assumed that the noise-free channel outputs $\mathbf{r} = \mathbf{Hs}$ were precomputed for the MAP detector, but produced online for the RDMAP and GMAP detectors. The computational load of the MAP detector grows exponentially with N', and this detector is rendered infeasible for systems with a moderate-to-large number of transmitter antennas. The MMSE detector has polynomial complexity with respect to N' and M', with the matrix inversion typically being the most computationally intensive task. The complexity of the proposed RDMAP and GMAP detectors grows exponentially with N_G, and thus, it is important to choose a moderately small N_G. For small N_G, the complexity of the RDMAP and GMAP detectors is polynomial with respect to N' and M', with the eigenvalue decomposition requiring a large part of the computational effort. In all operations, the GMAP detector has approximately $1/N_G$ the complexity of the RDMAP detector.

Simulations were performed for both QPSK modulation and 16QAM using an independent Rayleigh fading model with perfect channel knowledge at the receiver. For simulations of the coded BER, a rate-1/2 recursive systematic convolutional code with generating polynomial (7,5) was used with a random interleaver and deinterleaver. For the QPSK simulations, bursts of 200 symbols were used, final bit decisions were produced after 10 turbo iterations, and the prescaling constant was set to $\kappa = 1$. For the 16QAM simulations, bursts of 400 symbols were used, with 15 turbo iterations and $\kappa = 0.65$. The 16QAM system was more sensitive to correlated intrinsic input, and thus required a longer symbol burst and a nonunity κ.

The effect of the group size N_G on the the uncoded BER performance of the RDMAP and GMAP detectors is shown in Fig. 3, for a system with $N = 6$ transmitting antennas and $M = 6$ receiver antennas using QPSK modulation. From *Lemma 1*, the MMSE detector in [2] and [3] corresponds to the case of $N_G = 1$ for either detector. The uncoded BER performance of the both the RDMAP and GMAP detectors tends to improve with larger N_G, with the RDMAP detector performing slightly better than the GMAP detector, especially for larger N_G. The RDMAP and GMAP detectors hence have a significant uncoded BER improvement over the MMSE [2], [3]. Looking at BER produced by the analysis in the previous section, the analysis is almost exact in the $N_G = 1$ case. For larger N_G, the BER produced by analysis
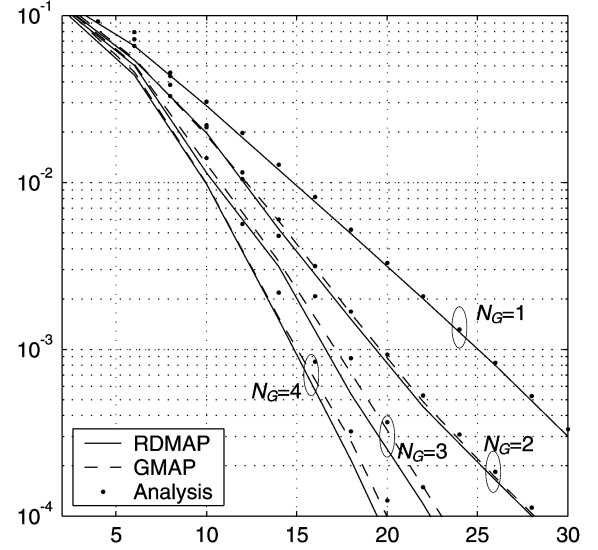

Fig. 3.   Uncoded BER performance of RDMAP and GMAP detectors for the $N = 6$, $M = 6$ system using QPSK modulation with $N_G = 1, 2, 3, 4$.

tends to approach the simulated BER of the RDMAP and GMAP detectors from above with increasing SNR. In the low-SNR region, the gap between the analysis and the simulated performance is largely the result of the union-bound approximation, and this gap is more pronounced for large N_G.

The effect of the number of receiver antennas on the uncoded BER performance is shown in Fig. 4 for a system with $N = 6$ transmitter antennas, a group size of $N_G = 3$, and different numbers of receiver antennas. As expected, BER performance improves with a larger number of receiver antennas, and the RDMAP detector performed slightly better than the GMAP detector. A BER floor at 4×10^{-2} exists in the $N = 6$, $M = 4$ system, having fewer receiver antennas that transmitter antennas. The BER produced by analysis tended to provide an approximate upper bound on the BER, with relative tightness for the $N = 6, M = 6$ and $N = 6, M = 8$ systems. The analysis in the $N = 6, M = 4$ system tends to significantly overestimate the BER because of the error floor and the poorness of the union-bound approximation when the BER is high. In the uncoded system having fewer receiver than transmitter antennas, both the RDMAP and GMAP detectors reach an error floor.

We now consider the coded BER performance of these detectors in a turbo processing receiver in systems with fewer receiver antennas than transmitter antennas. For a group size $N_G = 4$, the BER performance of the RDMAP and GMAP detector are shown in Fig. 5 for a $N = 6, M = 5$ system and
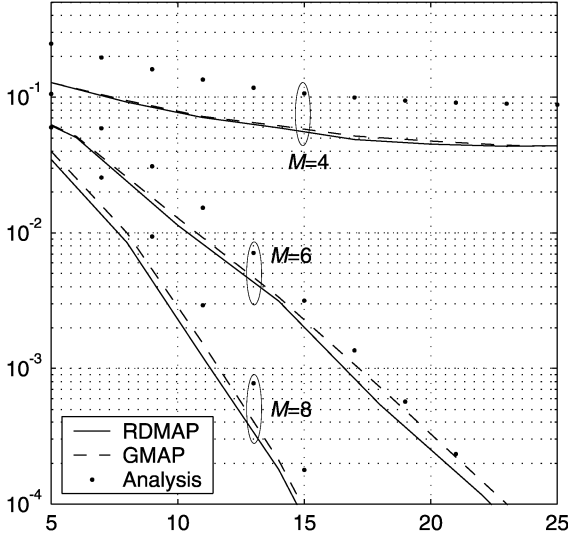
Fig. 4. Uncoded BER performance of RDMAP and GMAP detectors for QPSK system with $N = 6$, $N_G = 3$, and different M.
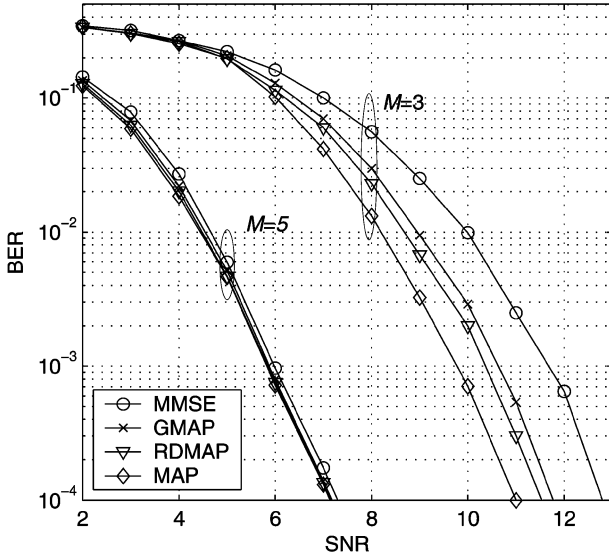

Fig. 5. Coded BER comparison of RDMAP, GMAP, MMSE, and MAP detectors using QPSK for the $N = 6$, $M = 3$ system, and the $N = 6$, $M = 5$ system. Group size $N_G = 4$ for the RDMAP and GMAP detectors.

Fig. 6. Coded BER as a function of the number of turbo iterations for RDMAP and GMAP detectors using QPSK for the $N = 6$, $M = 3$ system with $N_G = 4$.

a $N = 6$, $M = 3$ system. The BER performance of the MMSE detector from [2] and [3] and MAP detector are also shown for reference. Unlike the uncoded case, there is no error floor, and the turbo processing receiver is able to equalize both systems that have fewer receiver than transmitter antennas. In both the $N = 6$, $M = 3$ and $N = 6$, $M = 5$ systems, the MAP detector performed the best, the MMSE detector performed the worst, and the RDMAP detector performed better than the GMAP detector. In the $N = 6$, $M = 5$ system, the performance gap between the MMSE and MAP detector is small, suggesting the MMSE has near-optimal performance. In the $N = 6$, $M = 3$ system, however, there is a 2-dB performance gap between the MMSE and MAP detectors, and the RDMAP and GMAP detectors have a performance improvement of approximately 1.3 and 1 dB, respectively, over the MMSE detector.

The coded BER as a function of the number of turbo iteration is shown in Fig. 6 for the RDMAP and GMAP detectors
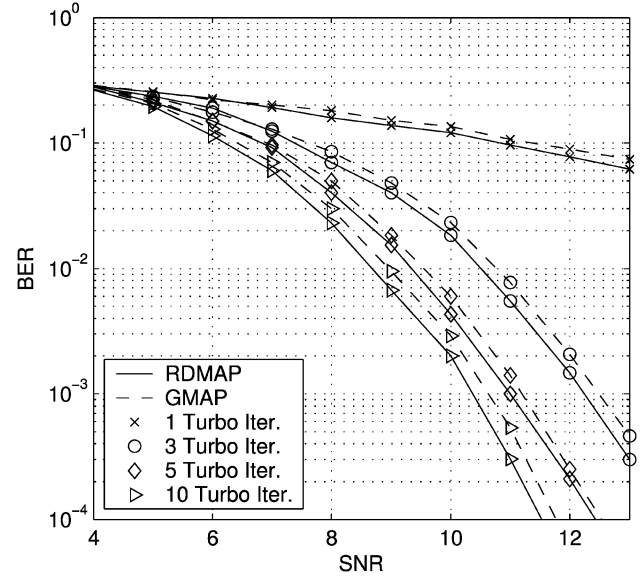
with $N_G = 4$ in a $N = 6$, $M = 3$ system using QPSK. The coded BER is produced after 1, 3, 5, and 10 turbo iterations. For both detectors, there is a significant BER improvement after the first few turbo iterations, and the incremental improvement is less with each iteration. There is only a 0.3-dB performance improvement between the fifth and tenth iteration.

The effect of group size on the coded BER performance of the RDMAP and GMAP detector is shown in Fig. 7 as a function of the group size $N_G = 1, 2, 4, 5$ for a system with $N = 10$ transmitter antennas and $M = 4$ receiver antennas using QPSK. The MAP detector is omitted, as it is computationally infeasible for this system, and the MMSE detector [2], [3] is equivalent to the case of $N_G = 1$. The BER performance of the RDMAP and GMAP detectors improves with larger N_G, with the RDMAP performing better than the GMAP detector for a given N_G. For the case of $N_G = 5$, the RDMAP and GMAP detectors have a performance gain of 3 and 2.5 dB, respectively, over the MMSE detector at a nominal BER of 10^{-3}. The better performance of the RDMAP detector over the GMAP detector stems from the dynamic group selection, which incorporates the intrinsic symbol probabilities in the grouping criteria. As a result, the signal power at the whitening filter output is generally higher for the RDMAP detector.

The coded BER performance of the MMSE, RDMAP, and GMAP detectors for 16QAM modulation is shown in Fig. 8 for a $N = 4$, $M = 2$ system for group sizes $N_G = 2, 4$. The MMSE detector was not able to equalize the channel having an error floor at 10^{-1}. The RDMAP and GMAP detectors could equalize the channel, although the GMAP detector with $N_G = 2$ reaches an error floor just above 10^{-3}. At a nominal BER of 10^{-3}, the RDMAP detector with $N_G = 4$ had the best performance, with the GMAP detector $N_G = 4$ and RDMAP detector $N_G = 2$ having a 2 and 5 dB performance loss, respectively. For systems having significantly fewer receiver antennas than transmitter antennas, the RDMAP and GMAP detectors can provide a significant coded BER performance improvement over the MMSE detector at a computational cost that is significantly smaller than
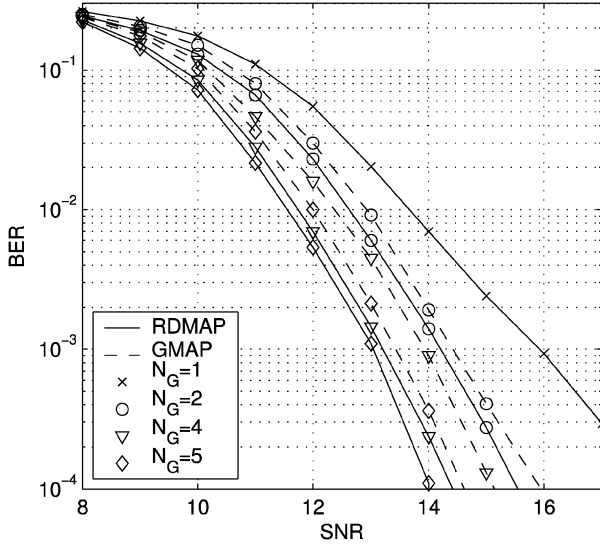
Fig. 7.   Coded BER comparison of RDMAP and GMAP detectors for the $N = 10$, $M = 4$ QPSK system with $N_G = 1, 2, 4, 5$.
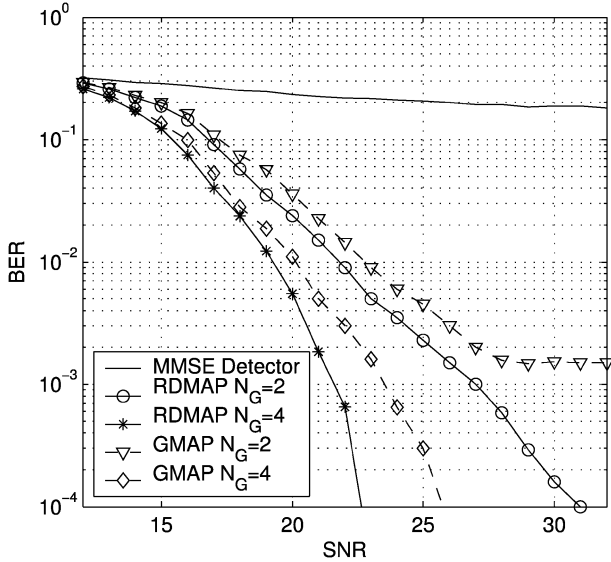

Fig. 8.   Coded BER comparison of MMSE, RDMAP, and GMAP detectors for the $N = 4$, $M = 2$ system using 16QAM modulation with $N_G = 2, 4$.

the MAP detector. The performance improvement is more significant with high-order modulations, where the MMSE detector may be ineffective.

VI. SUMMARY AND CONCLUSIONS

In this paper, we proposed two novel group detectors that operate within a turbo-processing BLAST receiver. These detectors allow a tradeoff between complexity and performance through the MAP group size, and include as special cases both the MAP and MMSE detectors in [2] and [3]. A novel grouping algorithm is developed for the GMAP detector. The uncoded BER performance of the RDMAP and GMAP detectors has been analyzed, and the analysis provided an approximate upper bound on the BER. In the uncoded case, the RDMAP and GMAP detectors have a significant performance improvement over the MMSE detector for a moderately chosen MAP group

size. In the coded case, the RDMAP and GMAP detectors can have a large performance improvement over the MMSE detector in systems having fewer receiver antennas than transmitter antennas.

APPENDIX

In this appendix, we prove *Lemmas 1* and *2*, and thus show that the RDMAP/GMAP detector is a generalization of both the MAP and MMSE detectors in [2] and [3].

Lemma 1: The RDMAP detector and MAP detector are equivalent for $N_G = N'$.

Proof: For $N_G = N'$, $\mathbf{s}_G = \mathbf{Ps}$ and $\mathbf{H}_G = \mathbf{HP}^T$, where $\mathbf{P}$ is a permutation matrix. Also, $\mathbf{w} = \mathbf{v}$ is a zero-mean Gaussian scaler that has a whitening filter vector $\mathbf{F} = 1/\sigma^2$. The RDMAP decision in (13) for a bit b_i correspond in a symbol $s_i = s_{Gj}$ can be expressed as

$$\Lambda_1[b_i]$$

$$= \log \frac{\sum\limits_{\mathbf{s_G} \in S_G, s_{Gj}=+1} P(\mathbf{s}_G) \exp\left(\frac{-\|\sigma^{-1}\mathbf{r} - \sigma^{-1}\mathbf{HP}^T\mathbf{Ps}\|^2}{2}\right)}{\sum\limits_{\mathbf{s_G} \in S_G, s_{Gj}=-1} P(\mathbf{s}_G) \exp\left(\frac{-\|\sigma^{-1}\mathbf{r} - \sigma^{-1}\mathbf{HP}^T\mathbf{Ps}\|^2}{2}\right)}$$

(35)

$$= \log \frac{\sum\limits_{\mathbf{s} \in S, s_i=+1} P(\mathbf{s}) \exp\left(\frac{-\|\mathbf{r} - \mathbf{Hs}\|^2}{2\sigma^2}\right)}{\sum\limits_{\mathbf{s} \in S, s_i=-1} P(\mathbf{s}) \exp\left(\frac{-\|\mathbf{r} - \mathbf{Hs}\|^2}{2\sigma^2}\right)}$$

(36)

which is the MAP decision rule in (8).   ∎

Lemma 2: The RDMAP detector and MMSE detector in [2] and [3] are equivalent for $N_G = 1$.

Proof: The MMSE detector in [2] and [3] uses a combination of soft interference cancellation and an MMSE residual suppression filter. To make a decision for a bit b_i, the channel observation after interference cancellation is given by

$$\tilde{\mathbf{r}}_i = \mathbf{r} - \mathbf{H}\hat{\mathbf{s}}_i \tag{37}$$

where $\hat{\mathbf{s}}_i = [\hat{s}_1, \ldots, \hat{s}_{i-1}, 0, \hat{s}_{i+1}, \ldots, \hat{s}_N]^T$, $\hat{s}_j = \tanh(\lambda_2^p[b_j]/2)$, and QPSK is assumed. The output $\tilde{\mathbf{r}}_i$ is filtered with a residual MMSE filter given by $\mathbf{w}_i^T = 2\mathbf{h}_i^T \mathbf{R}_i^{-1}\tilde{\mathbf{r}}$, with $\mathbf{R}_i = [\mathbf{H}\boldsymbol{\Omega}_i\mathbf{H}^T + \sigma^2\mathbf{I}]$, $\boldsymbol{\Omega}_i = \mathrm{diag}(\omega_1, \ldots, \omega_{i-1}, 1, \omega_{i+1}, \omega_N)$ and $\omega_j = \mathrm{E}[|s_i - \hat{s}_i|^2] = 1 - \hat{s}_i^2$. The filter output is approximated as a Gaussian distribution to produce an extrinsic LLR given by

$$\lambda_1[b_i] = \frac{2\mathbf{h}_i^T \mathbf{R}_i^{-1}\tilde{\mathbf{r}}_i}{1 - \mathbf{h}_i^T \mathbf{R}_i^{-1}\mathbf{h}_i}. \tag{38}$$

For the case where $N_G = 1$, the RDMAP decision in (13) can be simplified as

$$\Lambda[b_i] = \log \frac{P(s_i = +1)}{P(s_i = -1)} - \frac{\|\mathbf{y} - \mathbf{Fh}_i\|^2}{2} + \frac{\|\mathbf{y} + \mathbf{Fh}_i\|^2}{2}$$

(39)

$$= \lambda_2^p[b_i] + 2\mathbf{h}_i^T \mathbf{F}^T\mathbf{y} \tag{40}$$

where the second term in (40) is the extrinsic output that can be further simplified as

$$\lambda_1[b_i] = 2\mathbf{h}_i^T \mathbf{F}^T\mathbf{F}(\mathbf{r} - \bar{\mathbf{w}}) = 2\mathbf{h}_i^T \mathbf{R}_w^{-1}(\mathbf{r} - \bar{\mathbf{w}}). \tag{41}$$

We now require to show that (41) and (38) are equivalent. We first note that $\tilde{\mathbf{r}}_i = \mathbf{r} - \bar{\mathbf{w}}$, as it can be readily seen that $\mathbf{H}\hat{\mathbf{s}}_i = \mathbf{H}_{\bar{G}}\mathbf{s}_{\bar{G}}$. What remains is to show that

$$\mathbf{R}_w^{-1} \equiv \mathbf{R}_i^{-1}\left(1 - \mathbf{h}_i^T\mathbf{R}_i^{-1}\mathbf{h}_i\right)^{-1}. \tag{42}$$

Substituting $\mathbf{R}_w = \mathbf{R}_i + \mathbf{h}_i\mathbf{h}_i^T$ on the left-hand side of (42) and using the matrix inversion lemma $(\mathbf{A} + \mathbf{X}\mathbf{B}\mathbf{X}^T)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{X}\left(\mathbf{B}^{-1} + \mathbf{X}^T\mathbf{A}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{A}^{-1}$ with $\mathbf{A} = \mathbf{R}_i, \mathbf{B} = -1$, and $\mathbf{X} = \mathbf{h}_i$ gives the expression on the right-hand side (RHS) of (42) after some simplification. ∎

A. Evaluation of Noise Correlation Eigenvalues

This section discusses several methods to evaluate the expectation of the ordered eigenvalues of the noise-correlation matrix $\hat{\boldsymbol{\Sigma}}_c$. A direct method to determine $\hat{\boldsymbol{\Sigma}}_c$ is through Monte Carlo simulations. Such a method can, however, be computationally intensive. Alternatively, the eigenvalues of $\hat{\boldsymbol{\Sigma}}_c$ can be expressed as

$$\text{eig}[\hat{\boldsymbol{\Sigma}}_c] = 2\text{E}[\mathbf{W}(M, N_{Ic})] + \text{diag}[\mathbf{I}\sigma^2] \tag{43}$$

where $\mathbf{W}(m, n)$ is a complex Wishart matrix of the form $\mathbf{W}(m, n) = \mathbf{H}(m, n)\mathbf{H}(m, n)^H$, and $\mathbf{H}(m, n)$ is a $m \times n$ matrix of independent complex zero-mean Gaussian entries with unity variance. The pdf of the ordered eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_m$ of $\mathbf{W}(m, n)$ is given by [24]

$$P(\lambda_1, \lambda_2, \ldots, \lambda_m) = \frac{\pi^{m(m-1)}}{\Gamma_m(n)\Gamma_m(m)}$$
$$\times \exp\left[\sum_{i=1}^{m}\lambda_i\right]\prod_{i=1}^{m}\lambda_i^{n-m}\prod_{i<j}(\lambda_i - \lambda_j)^2 \tag{44}$$

where $\Gamma_m(a) = \pi^{m(m-1)/2}\prod_{i=1}^{m}\Gamma(a - i + 1)$ is the multivariate gamma function. The expectation of the Wishart matrix eigenvalues can be directly evaluated by marginalizing the joint pdf in (44). The number of factors on the RHS of (44) grows with $m!$, and quickly becomes computationally infeasible for moderately large m. A computationally more efficient method to find the expected eigenvalues can be adapted from [21, App. B]. This method exploits the intrinsic symmetry in eigenvalue pdf and converts the m nested integrals (required to find the expectation) into m separate integrals.

The method used in this paper to find the expected eigenvalues of the Wishart matrix is simpler than the methods previously discussed, and it is based on the the asymptotic eigenvalue distribution as $n, m \to \infty$. Letting m, n increase without bound, such that $m/n \to \beta$ asymptotically, the empirical distribution of $F(x) = P(\lambda < x)$ of $\mathbf{W}(m, n)/n$ converges to the nonrandom limit defined by [24]

$$dF(x) = \frac{\sqrt{(x - a(\beta))(b(\beta) - x)}}{2\pi\beta x} + \delta(x)I(\beta > 0)\left(1 - \frac{1}{\beta}\right) \tag{45}$$

for $\beta \neq 1$, where $a(\beta) = (1 - \sqrt{(\beta)})^2, b(\beta) = (1 + \sqrt{(\beta)})^2$. In the case of $\beta = 1$, we have $dF(x) = (2\pi)^{-1}((4 - x)/x)^{1/2}$ for $0 \leq x \leq 4$. The expectation of the eigenvalues of $\mathbf{W}(m, n)$ can be found from (45), by dividing the domain of $dF(x)$ into m

adjacent regions, such that each region has the same area under $dF(x)$. Each region can be represented by its center of mass under $dF(x)$, denoted $x_i, i = 1, \ldots, m$ for the ith region. The x_i's can be easily found by numerically integrating (45). The expected eigenvalues of $\mathbf{W}(m, n)$ can then be approximated as $\text{E}[\text{eig}[\mathbf{W}(m, n)]] \approx n\text{diag}[x_1, \ldots, x_m]$. The use of (45) to approximate the eigenvalues of $\mathbf{W}(m, n)$ tends to be accurate, even for small m, especially when $n \geq m$. For the simulated examples in Section V, there was negligible difference in the estimated BER, whether the mean eigenvalue matrix $\hat{\boldsymbol{\Sigma}}_c$ was determined through Monte Carlo simulations or by using the asymptotic distribution in (45).

REFERENCES

[1] G. Foschini and M. Gans, "On limits of wireless communication in a fading enviornment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 311–335, 1998.

[2] M. Sellathurai and S. Haykin, "Turbo-BLAST for wireless communications: Theory and experiments," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2538–2546, Oct. 2002.

[3] T. Abe and T. Matsumoto, "Space–time turbo equalization and symbol detection in frequency selective MIMO channels," in *Proc. Veh. Technol. Conf.*, vol. 2, Oct. 2001, pp. 1230–1234.

[4] P. Wolniansky *et al.*, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. IEEE ISSSE Conf.*, Sep. 1998, pp. 295–300.

[5] G. Foschini *et al.*, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 11, pp. 1841–1852, Nov. 1999.

[6] O. Damen, A. Chkeif, and J. Belfiore, "Lattice code decoder for space–time codes," *IEEE Commun. Lett.*, vol. 4, no. 5, pp. 161–163, May 2000.

[7] B. Hassibi and H. Vikalo, "On the expected complexity of integer least-squares problems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, 2002, pp. 1497–1500.

[8] A. Bhargave, R. de Figueiredo, and T. Eltoft, "A detection algorithm for the V-BLAST system," in *Proc. Veh. Technol. Conf.*, vol. 2, Oct. 2001, pp. 494–498.

[9] X. Li *et al.*, "Reduced-complexity detection algorithms for systems using multi-element arrays," in *Proc. Global Telecommun. Conf.*, vol. 2, Nov. 2000, pp. 1072–1076.

[10] W. Choi, R. Negi, and J. Cioffi, "Combined ML and DFE decoding for the V-BLAST system," in *Proc. IEEE Int. Conf. Commun.*, vol. 3, 2000, pp. 1243–1248.

[11] A. M. Tonello, "Space–time bit-interleaved coded modulation with an iterative decoding strategy," in *Proc. Veh. Technol. Conf.*, Sep. 2000, pp. 473–478.

[12] S. L. Ariyavisitakul, "Turbo space–time processing to improve wireless channel capacity," *IEEE Trans. Commun.*, vol. 48, no. 8, pp. 1347–1359, Aug. 2000.

[13] I. Lee, A. Chan, and C. Sundberg, "Space–time bit-interleaved coded modulation for OFDM systems," *IEEE Trans. Signal Process.*, vol. 52, no. 3, pp. 820–825, Mar. 2004.

[14] X. Wang and H. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 9, pp. 1046–1061, Sep. 1999.

[15] H. El Gamal and E. Geraniotis, "Iterative multiuser detection for coded CDMA signals in AWGN and fading channels," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 1, pp. 30–41, Jan. 2000.

[16] I. Lee, A. Chan, and C. Sundberg, "Design of simplified receivers for space-time bit-interleaved coded modulation systems," in *Proc. Veh. Technol. Conf.*, Apr. 2003, pp. 22–25.

[17] H. El Gamal and A. Hammons, "A new approach to layered space–time coding and signal processing," *IEEE Trans. Inf. Theory*, vol. 47, no. 9, pp. 2321–2334, Sep. 2001.

[18] H. El Gamal and M. Damen, "Universal space–time coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1097–1119, May 2003.

[19] S. Benedetto *et al.*, "A soft-input soft-output APP module for iterative decoding of concatenated codes," *IEEE Commun. Lett.*, vol. 1, no. 1, pp. 22–24, Jan. 1997.

[20] S. Grant and J. Cavers, "Performance enhancement through joint detection of cochannel signals using diversity arrays," *IEEE Trans. Commun.*, vol. 46, no. 8, pp. 1038–1049, Aug. 1998.

[21] D. Lao and A. Haimovich, "Exact closed-form performance analysis of optimum combining with multiple cochannel interferers and Rayleigh fading," *IEEE Trans. Commun.*, vol. 51, no. 6, pp. 995–1003, Jun. 2003.

[22] G. L. Turin, "Characteristic function of Hermitian quadratic forms in complex normal variables," *Biometrika*, vol. 47, pp. 199–201, 1960.

[23] M. Schwartz, W. Bennett, and S. Stein, *Communication Systems and Techniques*.   New York: McGraw-Hill, 1966.

[24] A. Edelman, "Eigenvalues and condition numbers of random matricies," *SIAM J. Matrix Anal. Appl.*, vol. 9, no. 4, pp. 543–560, Oct. 1988.

Akrum Elkhazin was born in Luton, U.K., in 1977. He received the B.A.Sc. degree in systems design from the University of Waterloo, Waterloo, ON, Canada, in 2000 and the M.A.Sc. degree in 2002 in electrical engineering from the University of Toronto, Toronto, ON, Canada, where he is currently working toward the Ph.D. degree in electrical engineering.

His research interests include signal processing, equalization and detection, space–time coding, and turbo processing receivers.

Konstantinos (Kostas) N. Plataniotis received the B.Eng. degree in computer engineering and informatics from University of Patras, Patras, Greece, in 1988, and the M.S. and Ph.D. degrees in electrical engineering from Florida Institute of Technology, Melbourne, FL, in 1992 and 1994 respectively.

He is currently an Associate Professor with Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada. His research interests include signal and image processing, communications systems, biometrics, stochastic estimation, and pattern recognition.

Dr. Plataniotis is a Registered Professional Engineer in the Province of Ontario, and a member of the Technical Chamber of Greece. He is an Associate Editor for the IEEE TRANSACTIONS ON NEURAL NETWORKS, the Technical Program Co-Chair for the IEEE International Conference on Multimedia and Expo 2006, and the Vice-Chair of the IEEE Intelligent Transportation Systems Conference 2006. He is the 2005 recipient of IEEE Canada's Outstanding Engineering Educator Award "for contributions to engineering education and inspirational guidance of graduate students".

Subbarayan Pasupathy (M'73–SM'81–F'91) was born in Chennai (Madras), India, on September 21, 1940. He received the B.E. degree in telecommunications from the University of Madras, Madras, India, in 1963, the M.Tech. degree in electrical engineering from the Indian Institute of Technology, Madras, India, in 1966, and the M.Phil. and Ph.D. degrees in engineering and applied science from Yale University, New Haven, CT, in 1970 and 1972, respectively.

He joined the faculty of the University of Toronto, Toronto, ON, Canada, in 1973 and became a Professor of Electrical Engineering in 1983. He has served as Chairman of the Communications Group and as the Associate Chairman of the Department of Electrical Engineering at the University of Toronto. His research interests are in the areas of communications theory, digital communications, and statistical signal processing.

Dr. Pasupathy is a Registered Professional Engineer in the Province of Ontario. He was awarded the Canadian Award in Telecommunications in 2003 by the Canadian Society of Information Theory, and was elected as a Fellow of the Engineering Society of Canada in 2004. He has served as a Technical Associate Editor for the *IEEE Communications Magazine* (1979–1982) and as an Associate Editor for the *Canadian Electrical Engineering Journal* (1980–1983). During 1982–1989, he was an Area Editor for Data Communications and Modulation for the IEEE TRANSACTIONS ON COMMUNICATIONS. From 1984 to 1998, he wrote a regular column entitled "Light Traffic" for the *IEEE Communications Magazine*.