

On Some Design Issues of Space-Time Coded Multi-Antenna Systems

Hsuan-Jung Su

*Bell Laboratories, Lucent Technologies, Holmdel, NJ 07733, USA
Email: hjsu@lucent.com*

Evangelos Geraniotis

*Department of Electrical and Computer Engineering, University of Maryland at College Park,
College Park, MD 20742, USA
Email: evangelos@eng.umd.edu*

Received 30 May 2001 and in revised form 30 January 2002

This paper concerns some design issues and tradeoffs of communication systems equipped with multiple transmit and receive antennas. The general space-time coding/modulation structure by Tarokh et al. (1999) is considered. Several design issues are investigated for this structure. The layered space-time architecture by Foschini (1996) is revisited as a special case of the general structure. It is also used to demonstrate the design and complexity tradeoffs of the system. Through intuitive and analytical explanations, as well as simulations, the design considerations for these space-time transmission structures and their contributions to the performance are shown.

Keywords and phrases: space-time codes, array processing, iterative processing.

1. INTRODUCTION

The growing demand of high data rate service has inspired studies on multi-antenna wireless communication systems in recent years (e.g., [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]). From an information theoretic point of view, multiple antennas can increase the capacity of wireless channels [3, 4]. In practice, this increased capacity amounts to improved error performance when the communication data rate is fixed.

When there are not enough antennas at the receiver to resolve the transmission signal spaces, joint detection of the signals transmitted from different antennas may be necessary at the receiver. In this scenario, proper coordination of the signals transmitted on different antennas can help improving the performance without additional cost. The space-time coding techniques in [8, 9, 10, 11], which integrate channel coding, modulation and antenna diversity to allow simultaneous exploitation of coding gain (temporal diversity) and antenna diversity, are examples in this category. The drawback of such signal coordination is intractable joint detection complexity when the number of transmit antennas is large, unless certain throughput-inefficient signal coordinations are introduced [11].

If the number of receive antennas is large enough to resolve the transmission signal spaces, the signals from individ-

ual transmit antennas can be detected separately after proper signal space separation. Signal coordination is not necessary in this case, and the throughput can be kept high. The layered space-time (LST) architecture in [2] is an attempt to realizing the information theoretic result in [3, 4], namely, the multi-antenna channel capacity grows linearly with the smaller of the number of transmit antennas and the number of receive antennas. The LST uses one-dimensional signal per data layer. Its focus is more on the signal space separation at the receiver. While the channel coding method is, contrary to the space-time coding techniques, conventional.

In order to attain a compromise between throughput and performance when the receiver has a certain number of antennas, the authors of [8] later on extended their work to consider grouped space-time transmission (GST) [1]. In this system, the space-time coding advantages are explored by the antennas within the same group, while the system throughput can be kept high by transmitting the groups independently. As in the LST, signal processing at the receiver is required in order to separate the signal groups. A zero-forcing method was employed in [1] for signal space separation.

The GST design possesses the fundamental ingredients of a multi-antenna system: space-time coding, data allocation, and antenna array processing. It therefore can be seen as a general space-time transmission structure, with the LST

being a special case which has one antenna per group. In this paper, we investigate the system properties and design issues of this general structure. Some design considerations and tradeoffs are revealed through intuitive, as well as analytical discussions. The issues discussed are then verified with simulations. Note that the issues considered in this paper pertain to an “open-loop” system, meaning there is no channel information feedback from the receiver to the transmitter.

This paper is organized as follows. In Section 2, the system model of the GST is given. We then discuss the antenna array processing issues in Section 3, which is followed by the power allocation issue in Section 4. In Section 5, spatial interleaving is discussed in the context of diversity advantage. Section 6 revisits the LST in terms of diversity and coding gain. In Section 7, we describe the application of the iterative minimum-mean-squared-error (MMSE) multiuser detection algorithm [13, 14] to the GST. Section 8 then verifies and concludes the design issues discussed with comprehensive simulation results. Finally, some concluding remarks are given in Section 9.

2. SYSTEM MODEL

In the remainder of this paper, the superscripts $*$, T , and H are used to denote complex conjugate, matrix transposition, and Hermitian transposition, respectively. The communication system considered consists of a transmitter which is equipped with n antennas and a receiver which has m antennas. At the transmitter, data is encoded by the channel encoder, the encoded symbols are then divided into n streams which will be transmitted simultaneously using n distinct antennas after modulation.

The signal at each receive antenna is a noisy superposition of the n transmitted signals corrupted by the fading channel which is assumed to be flat, Rayleigh, and spatially independent, meaning fading is statistically independent from one transmit-receive antenna pair to another. Following the convention of [1, 2], we assume that fading remains constant within a data frame and varies from one frame to another. This assumption amounts to the worst case situation where temporal diversity is not available, and can serve as an upper bound on the error probability. For this system, the signal r_t^j received by antenna j ($1 \leq j \leq m$) at time t is given by

$$r_t^j = \sum_{i=1}^n \alpha_{i,j} c_t^i \sqrt{E_i} + \eta_t^j, \quad (1)$$

where i is the transmit antenna index, E_i is the symbol energy of the i th transmit antenna, c_t^i is the symbol transmitted from antenna i at time t , and $\alpha_{i,j}$ is the path gain from the i th transmit antenna to the j th receive antenna. The noise η_t^j is assumed to be additive white Gaussian (AWGN) with variance $N_0/2$ for both real and imaginary components.

For notational simplicity, we redefine $\alpha_{i,j}$ as the product of the channel gain $\alpha_{i,j}$ and the symbol amplitude $\sqrt{E_i}$ in (1).

This change in notation does not affect the zero-forcing array processing, while it allows more concise presentation of our alternative approach. With this new notation, we now rewrite (1) in the vector form

$$\mathbf{r}_t = \Omega \mathbf{c}_t + \boldsymbol{\eta}_t, \quad (2)$$

where

$$\begin{aligned} \mathbf{c}_t &= (c_t^1, c_t^2, \dots, c_t^n)^T, \\ \mathbf{r}_t &= (r_t^1, r_t^2, \dots, r_t^m)^T, \\ \boldsymbol{\eta}_t &= (\eta_t^1, \eta_t^2, \dots, \eta_t^m)^T, \\ \Omega &= \begin{bmatrix} \alpha_{1,1} & \alpha_{2,1} & \cdots & \alpha_{n,1} \\ \alpha_{1,2} & \alpha_{2,2} & \cdots & \alpha_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1,m} & \alpha_{2,m} & \cdots & \alpha_{n,m} \end{bmatrix}. \end{aligned} \quad (3)$$

In the GST, the data to be transmitted are divided into q groups, each space-time coded and transmitted with one group of antennas. The antenna groups are disjoint. The purpose of space-time coding is to ensure maximal spatial diversity, hence the steepest asymptotic slope of the performance curve [8]. Two 2-antenna space-time trellis code examples are shown in Figure 1. They both use the same trellis diagram but different input-output associations. The code to the right (Tarokh et al.) was proposed in [8],¹ while the other (Code A) is a slight alteration of it.

At the receiver, the signal groups are successively decoded. Once a group is decoded, its contribution to the received signal is removed before decoding the next group. Let n_j denote the number of transmit antennas in group j , $1 \leq j \leq q$, we have $n_1 + n_2 + \dots + n_q = n$. In decoding a specific group, the uncanceled groups are considered as interferences, therefore antenna array processing is necessary to suppress them. The array processing issues are discussed in Section 3.

3. ARRAY PROCESSING

In [1], a zero-forcing approach was used for array processing. In this section, we recapitulate this approach first. Then we propose an alternative approach which can improve the performance.

3.1. Zero-forcing array processing

Without loss of generality, we consider the decoding of the first group code denoted by \mathcal{C}_1 and assume that the channel gain matrix Ω is known to the receiver. Assuming that $m \geq n - n_1 + 1$, the zero-forcing group interference suppression method proposed in [1] uses the null space of the

¹The Tarokh et al. code in Figure 1 was generated with the encoder polynomials in [8, page 757]. It differs from [8, Figure 6] which has a few typos and results in slightly worse performance.

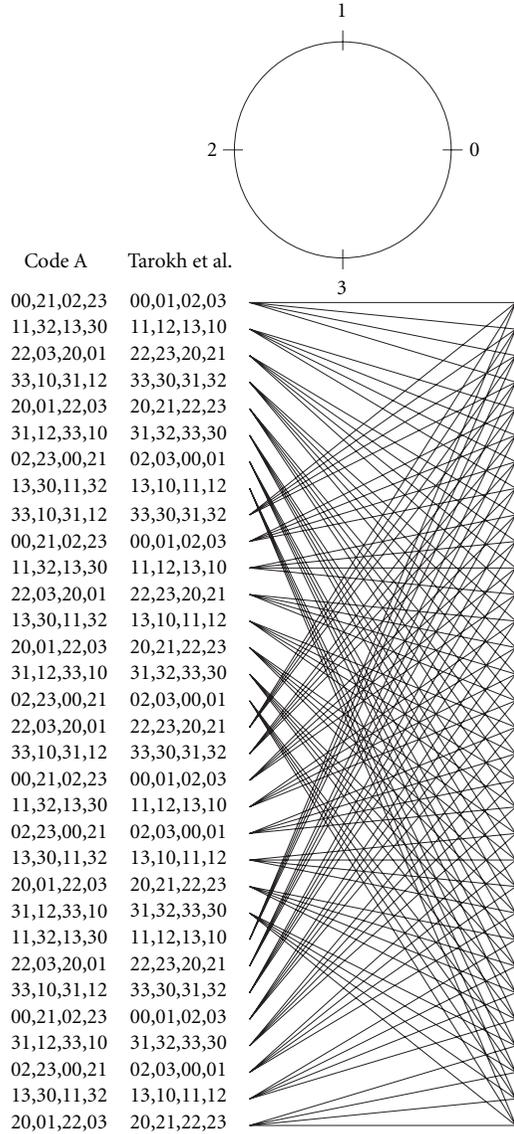


FIGURE 1: Two rate 1/2, 32-state, 4PSK space-time codes.

interference space spanned by the columns of

$$\Lambda(\mathcal{C}_1) = \begin{bmatrix} \alpha_{n_1+1,1} & \alpha_{n_1+2,1} & \cdots & \alpha_{n_1,1} \\ \alpha_{n_1+1,2} & \alpha_{n_1+2,2} & \cdots & \alpha_{n_1,2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{n_1+1,m} & \alpha_{n_1+2,m} & \cdots & \alpha_{n_1,m} \end{bmatrix} \quad (4)$$

as the array processor to remove completely the signals from other groups, however, at the expense of sacrificing some energy of the desired signals. Let

$$\Theta(\mathcal{C}_1) = (\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_{m-n+n_1}) \quad (5)$$

be the array processor, where \mathbf{v}_j , $j = 1, \dots, m - n + n_1$, form the null space of $\Lambda(\mathcal{C}_1)$, the received signal is processed with

$\Theta(\mathcal{C}_1)$. Let

$$\Omega(\mathcal{C}_1) = \begin{bmatrix} \alpha_{1,1} & \alpha_{2,1} & \cdots & \alpha_{n_1,1} \\ \alpha_{1,2} & \alpha_{2,2} & \cdots & \alpha_{n_1,2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1,m} & \alpha_{2,m} & \cdots & \alpha_{n_1,m} \end{bmatrix}, \quad (6)$$

$$\tilde{\mathbf{r}}_t = \Theta^H(\mathcal{C}_1) \mathbf{r}_t, \quad (7)$$

$$\tilde{\Omega} = \Theta^H(\mathcal{C}_1) \Omega(\mathcal{C}_1), \quad (8)$$

$$\tilde{\boldsymbol{\eta}}_t = \Theta^H(\mathcal{C}_1) \boldsymbol{\eta}_t, \quad (9)$$

we arrive at

$$\tilde{\mathbf{r}}_t = \tilde{\Omega} \mathbf{c}_t^1 + \tilde{\boldsymbol{\eta}}_t, \quad (10)$$

where $\mathbf{c}_t^1 = (c_t^1, c_t^2, \dots, c_t^{n_1})^T$. The decoder then decodes \mathcal{C}_1 by choosing the hypothesis $\hat{\mathbf{c}}_t^1$ which minimizes the metric

$$\sum_t |\tilde{\mathbf{r}}_t - \tilde{\Omega} \hat{\mathbf{c}}_t^1|^2. \quad (11)$$

3.2. Maximum signal-to-noise ratio (SNR) array processing

The performance of the array processor can be improved if the balance between interference suppression and noise enhancement can be found. We derive in this section the maximum SNR array processor. For a given time t , since the space-time coded symbols are unknown at the array processing stage, the SNR can only be maximized statistically. For this purpose, the signal and interference covariance matrices are defined, considering decoding the first group code, as

$$\mathbf{R}_s = \Omega(\mathcal{C}_1) E[\mathbf{c}_t^1 (\mathbf{c}_t^1)^H] \Omega^H(\mathcal{C}_1), \quad (12)$$

$$\mathbf{R}_n = \Lambda(\mathcal{C}_1) E[\mathbf{c}_t^o (\mathbf{c}_t^o)^H] \Lambda^H(\mathcal{C}_1) + N_0 \mathbf{I}, \quad (13)$$

respectively, with

$$\begin{aligned} \mathbf{c}_t^1 &= (c_t^1, c_t^2, \dots, c_t^{n_1})^T, \\ \mathbf{c}_t^o &= (c_t^{n_1+1}, c_t^{n_1+2}, \dots, c_t^n)^T. \end{aligned} \quad (14)$$

Using the canonical representation of the MMSE filter in [15], the maximum SNR array processor will require the same number of linear filters as the dimension of the signal space in order to collect all energy of the desired signal. Each of these linear filters can be decomposed into two components: one in the signal space and the other in the orthogonal space. The maximum SNR array processor is obtained using the following theorem.

Theorem 1. *Let m be the number of receive antennas, n_1 be the number of transmit antennas of the space-time code group in consideration; let \mathbf{R}_s and \mathbf{R}_n be the signal and interference covariance matrices, respectively, defined by (12) and (13). The linear maximum SNR array processor for this group, denoted by*

$$\Theta_M(\mathcal{C}_1) = (\mathbf{w}_1 \ \mathbf{w}_2 \ \cdots \ \mathbf{w}_k), \quad (15)$$

consists of a set of k linearly independent eigenvectors corresponding to the nonzero eigenvalues, counting multiplicities, of the generalized eigenvalue problem

$$\mathbf{R}_s \mathbf{w} = \lambda \mathbf{R}_n \mathbf{w}, \quad (16)$$

where $k \leq \min(m, n_1)$ is the rank of \mathbf{R}_s . The filtering outputs of these eigenvectors $\{\mathbf{w}_i^H \mathbf{r}_t\}_{i=1}^k$ are uncorrelated with one another.

Proof. See the appendix. \square

This theorem generalizes the MMSE filtering for multi-dimensional signals. When $n_1 = 1$ (single antenna transmission) or $c_t^1 = c_t^2 = \dots = c_t^{n_1}$ (repetitive transmission), the maximum SNR array processor becomes the conventional (one-dimensional) MMSE filter. Intuitively, the maximum SNR array processor prewhitens the interference and orthogonalizes the desired signal. Maximum ratio combining of the maximum SNR array processor outputs is done at the decoder by substituting $\Theta(\mathcal{C}_1)$ in (7), (8), (9), (10), (11) with $\Theta_M(\mathcal{C}_1)$. Using Gaussian approximation of the array processor outputs [16], the decoding becomes maximum likelihood. Compared to the zero-forcing approach, the decoder now only needs to process at most $\min(m, n_i)$ samples per trellis branch instead of $(m - n + \sum_{j=1}^i n_j)$ [1] when decoding \mathcal{C}_i . Unlike the zero-forcing method, the maximum SNR approach does not impose any requirement on the number of receive antennas. It works even when $(m - n + \sum_{j=1}^i n_j) \leq 0$. The performance under this situation, however, will suffer.

The principal ratio combining (PRC) method proposed in [17] is also a subset of the maximum SNR method. The PRC filter is the eigenvector of (16) corresponding to the largest eigenvalue. In practice, when the number of filters is limited by the equipment cost, the best possible performance can be achieved by using the eigenvectors of (16) corresponding to the largest eigenvalues (SNRs).

Given similar optimization procedures and the existence of local optima, the adaptive algorithm in [15] can be applied to find the maximum SNR filters when only the channel gains of the desired signal (i.e., $\Omega(\mathcal{C}_1)$) are known.

4. POWER ALLOCATION

As shown in [1], the diversity gains of the zero-forcing approach at different decoding stages can be expressed as an increasing sequence $\{n_i(m - n + \sum_{j=1}^i n_j)\}_{i=1}^q$. The authors of [1] thus proposed to allocate power among the antenna groups according to a geometrically decreasing sequence.

In the case of maximum SNR filtering, the optimization of power allocation is extremely difficult due to the fact that the filter outputs contain interference from other groups, and the contribution of this interference depends on the operating point (in terms of channel gains and AWGN power spectral density) of the filter. By fixing the overall power consumption, we nevertheless can see that, giving one group more power by depriving the power of the others increases the SNR two-fold: the signal power is increased and the interference power is decreased. We therefore argue that the power allocation pattern does not need to decrease so rapidly.

This conclusion can also be reached from a different perspective. From the discussion in Section 3, we can see that the desired signal energy is completely collected by the maximum SNR array processor, while this is not true in the zero-forcing case. The receive diversity gain of the maximum SNR method is therefore larger than that of the zero-forcing method. This diversity advantage, however, decreases as the transmission groups are decoded and canceled. For the last decoded group, both zero-forcing and maximum SNR methods give the same diversity. With this observation, we again can conclude that the power allocation pattern of the maximum SNR approach does not need to decrease so rapidly as the zero-forcing approach. In the simulation, we will compare geometric power allocation with arithmetic power allocation.

5. SPATIAL INTERLEAVING

The focus so far has been on maximizing the transmit and receive diversities within a group. When the number of transmit antennas is larger than required for one group (i.e., more than one group is allowed), it is possible to increase the transmit diversity beyond what is provided by space-time coding, if spatial interleaving is used. In other words, it might be beneficiary if the association between data groups and transmit antenna groups varies with time. One must bear in mind that, although the concept of improving the diversity gain by spatial interleaving is very intuitive and straightforward, inappropriate spatial interleaving can decrease instead of increasing the diversity.

We consider the case where the total number of transmit antennas is divisible by the number of code streams per group. A "group-based" spatial interleaving is realized by dividing the transmit antennas into q disjoint groups, then at every symbol duration, each space-time code group is transmitted on one distinct antenna group. The mapping from code groups to antenna groups may vary with time, but no change in the code stream order is made within a group. At the receiver, the array processing methods discussed previously still apply, except that the matrix indices need to be reordered accordingly.

Lemma 1. Let $\mathbf{B} = (\mathbf{B}_1 \ \mathbf{B}_2 \ \dots \ \mathbf{B}_q)$, where each matrix \mathbf{B}_k , $1 \leq k \leq q$, has dimension $n' \times l_k$. Let r, r_1, r_2, \dots, r_q be the ranks of $\mathbf{B}, \mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_q$, respectively. Then $\sum_{k=1}^q r_k \geq r$.

Proof. Using column operations can leave \mathbf{B}_k , $1 \leq k \leq q$, with r_k nonzero columns, respectively. Therefore, the rank of \mathbf{B} must be smaller than $\sum_{k=1}^q r_k$. \square

Since our focus is on the transmit diversity, and the receive diversity only depends on the number of receive antennas and the decoding order, we can consider, without loss of generality, the case where the code group of interest is transmitted alone. Assume that after the group-based spatial interleaving the space-time coded symbols are transmitted from the k th antenna group if $t \in \mathcal{T}_k$, $1 \leq k \leq q$, where the \mathcal{T}_k 's represent disjoint sets of time instants with their sizes l_k 's

satisfying $\sum_{k=1}^q l_k = l$. The elements in a set are not necessarily consecutive. The pairwise error probability that the decoder will prefer an alternate codeword

$$\mathbf{e} = \begin{bmatrix} e_1^1 & e_2^1 & \cdots & e_l^1 \\ e_1^2 & e_2^2 & \cdots & e_l^2 \\ \vdots & \vdots & \ddots & \vdots \\ e_1^{n'} & e_2^{n'} & \cdots & e_l^{n'} \end{bmatrix} \quad (17)$$

when

$$\mathbf{c} = \begin{bmatrix} c_1^1 & c_2^1 & \cdots & c_l^1 \\ c_1^2 & c_2^2 & \cdots & c_l^2 \\ \vdots & \vdots & \ddots & \vdots \\ c_1^{n'} & c_2^{n'} & \cdots & c_l^{n'} \end{bmatrix} \quad (18)$$

is transmitted, is upper bounded by [8]

$$P(\mathbf{c} \rightarrow \mathbf{e} \mid \alpha_{i,j}^k, i = 1, 2, \dots, n', j = 1, 2, \dots, m', k = 1, 2, \dots, q) \leq \exp\left(-d^2(\mathbf{c}, \mathbf{e}) \frac{E_s}{4N_0}\right), \quad (19)$$

where E_s is the symbol energy, n' is the number of transmit antennas per group, m' is the number of receive antennas (or the receive diversity level after array processing in the presence of the other groups [1]), $\alpha_{i,j}^k$ is the path gain from the i th transmit antenna in group k to the j th receive antenna, and

$$d^2(\mathbf{c}, \mathbf{e}) = \sum_{k=1}^q \sum_{t \in \mathcal{T}_k} \sum_{j=1}^{m'} \left| \sum_{i=1}^{n'} \alpha_{i,j}^k (c_t^i - e_t^i) \right|^2. \quad (20)$$

By taking expectation over the complex Gaussian variables $\alpha_{i,j}^k$, inequality (19) becomes

$$P(\mathbf{c} \rightarrow \mathbf{e}) \leq \prod_{k=1}^q \left(\frac{1}{\prod_{i=1}^{n'} (1 + \lambda_i^k E_s / 4N_0)} \right)^{m'}, \quad (21)$$

where λ_i^k , $i = 1, 2, \dots, n'$ are the eigenvalues of matrix \mathbf{A}_k defined element-wise by

$$A_{k,ij} = \sum_{t \in \mathcal{T}_k} (c_t^i - e_t^i)(c_t^j - e_t^j)^*. \quad (22)$$

Let r_k denote the rank of the matrix \mathbf{A}_k , and reorder the notations so that λ_i^k , $i = 1, 2, \dots, r_k$, are the nonzero eigenvalues of \mathbf{A}_k , it follows that

$$P(\mathbf{c} \rightarrow \mathbf{e}) \leq \left(\prod_{k=1}^q \prod_{i=1}^{r_k} \lambda_i^k \right)^{-m'} \left(\frac{E_s}{4N_0} \right)^{-\sum_{k=1}^q r_k m'}. \quad (23)$$

Thus a diversity advantage of $\sum_{k=1}^q r_k m'$ and a coding advantage of $(\prod_{k=1}^q \prod_{i=1}^{r_k} \lambda_i^k)^{1/\sum_{k=1}^q r_k}$ are achieved. When the number of transmit antennas is equal to the number of space-time

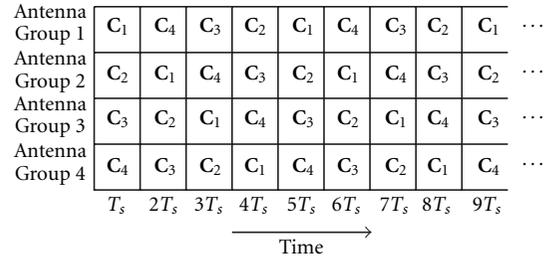


FIGURE 2: Spatial interleaving ("rotation"), T_s : symbol duration.

code streams per group, the antenna group index in (23) can be dropped to arrive at the original bound in [8]

$$P(\mathbf{c} \rightarrow \mathbf{e}) \leq \left(\prod_{i=1}^r \lambda_i \right)^{-m'} \left(\frac{E_s}{4N_0} \right)^{-rm'} \quad (24)$$

with diversity advantage rm' .

For notational simplicity, we give new, consecutive indices to the elements in \mathcal{T}_k and denote them with a "prime." By construction, the matrix

$$\mathbf{B}_k = \begin{bmatrix} e_{1'}^1 - c_{1'}^1 & e_{2'}^1 - c_{2'}^1 & \cdots & e_{l_k'}^1 - c_{l_k'}^1 \\ e_{1'}^2 - c_{1'}^2 & e_{2'}^2 - c_{2'}^2 & \cdots & e_{l_k'}^2 - c_{l_k'}^2 \\ \vdots & \vdots & \ddots & \vdots \\ e_{1'}^{n'} - c_{1'}^{n'} & e_{2'}^{n'} - c_{2'}^{n'} & \cdots & e_{l_k'}^{n'} - c_{l_k'}^{n'} \end{bmatrix} \quad (25)$$

is a square root of \mathbf{A}_k , and the ranks of \mathbf{A}_k and \mathbf{B}_k are the same. Based on (23), (24), and (25), the following proposition is an immediate result of Lemma 1.

Proposition 1. *When the number of transmit antennas is divisible by the number of space-time code streams per group, the diversity gain after the group-based spatial interleaving is no less than that provided by the space-time coding.*

Since errors in trellis decoding usually appear in clusters, letting consecutive code symbols experience independent fadings might give larger transmit diversity ($\sum_{k=1}^q r_k$). We consider in this paper a cycled spatial interleaving termed "rotation" (see Figure 2). In this interleaving scheme, the k th antenna group is devoted to the i th space-time code group when $t \in \mathcal{T}_k^i$, where $\mathcal{T}_k^i = \{k - i + 1 + hq \mid 1 \leq k - i + 1 + hq \leq l, h \in \mathbb{Z}\}$.

When the number of transmit antennas is large, the diversity advantage with rotation is dominated by the minimum free distance of the code.

6. LAYERED SPACE-TIME ARCHITECTURE

The LST proposed in [2] can be seen as a special case of the GST with one transmission stream per group. On the other hand, it is different from the space-time coded GST in that the coded symbols from the same group are multiplexed in time instead of simultaneously transmitted. With the same throughput (code rate $1/n'$), the LST transmits the codeword

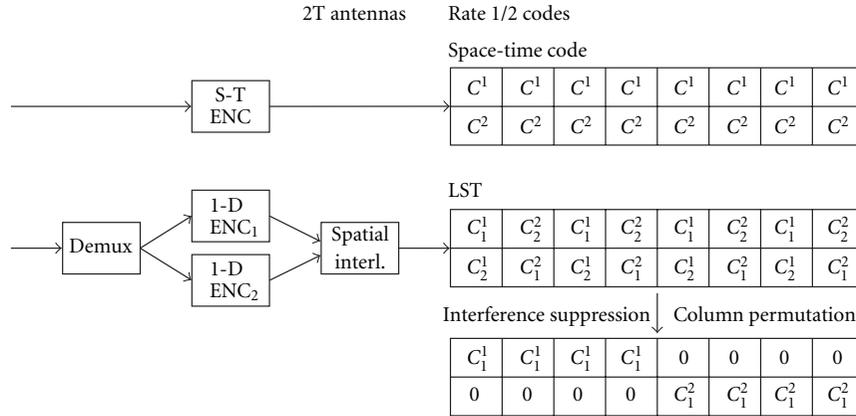


FIGURE 3: An example illustrating the design simplicity of the LST.

(18) column by column and one symbol at a time as

$$\mathbf{c} = c_1^1 c_2^2 \cdots c_{n'}^{n'} c_{n'+1}^1 c_{n'+2}^2 \cdots c_{2n'}^{n'} \cdots c_{(l-1)n'+1}^1 c_{(l-1)n'+2}^2 \cdots c_{ln'}^{n'}, \quad (26)$$

where the time indices have been modified to reflect the actual transmission order. As mentioned in [2], the LST uses only one-dimensional coding without signal coordination among antennas, so the design is much simpler. We elaborate in the following, using specific arguments on diversity and coding gain not found in [2], the LST design advantages.

6.1. Diversity

We, again, consider the transmission of a specific group in the absence of the other groups. Assume that the number of transmit antennas is divisible by n' and “rotation” is used. For each of the n' code streams (rows in (18)), a matrix as in Lemma 1 can be constructed with time shuffling

$$\mathbf{B}^i = (\mathbf{B}_1^i \ \mathbf{B}_2^i \ \cdots \ \mathbf{B}_q^i), \quad 1 \leq i \leq n', \quad (27)$$

where the constituent matrices $\mathbf{B}_k^i, 1 \leq k \leq q$, are given, with new consecutive time indices as in (25), by

$$\mathbf{B}_k^i = (e_{1'}^i - c_{1'}^i, \ e_{2'}^i - c_{2'}^i, \ \cdots, \ e_{l_k'}^i - c_{l_k'}^i), \quad (28)$$

and are subject to independent fadings. According to the discussion in Section 5, the diversity of this code stream is no less than the rank of \mathbf{B}^i , which is always one provided that the code considered is useful, meaning no two distinct input sequences share the same codeword on this code stream.

With the number of transmit antennas being divisible by n' and each code stream using a distinct set of antennas, we can construct for the entire transmission sequence of this group a matrix

$$\mathbf{B} = (\mathbf{B}^1 \ \mathbf{B}^2 \ \cdots \ \mathbf{B}^{n'}) \quad (29)$$

with the constituent matrices subject to independent fadings. It is now straightforward that the transmit diversity of a group is no less than n' as long as a useful code is employed.

To illustrate such a design simplicity of the LST, an example where $n' = 2$ and the number of transmit antennas equals n' is given in Figure 3. In this figure, the superscript is the code stream index, while the subscript is the group index which is not used by the space-time coding case as it has only one group. In order to achieve full (two) transmit diversity, the space-time coded system has to be carefully designed subject to the rank criterion in [8]. The LST with rotation, on the contrary, can provide full transmit diversity for both groups with any useful codes, assuming that the receiver has enough antennas to perform interference suppression. This much relaxed code design criterion facilitates the use of a larger class of good codes.

Similar interpretation of the LST is also given in [18], with detailed diversity derivations.

6.2. Coding gain

In this section, we consider the coding gain when the system complexity is constrained. From Theorem 1, we can see that, if there are enough receive antennas, the array processing complexity depends only on the number of transmit antennas, no matter how the transmit antennas are grouped. In other words, antenna grouping and coding may only affect the system complexity through decoding complexity, which is defined by the number of trellis branch computations per information bit. For the LST, the decoding of each group only involves one-dimensional signal, so the in-phase and quadrature components can be decoded separately. This is not true for a GST with more than one antenna per group, because the channel induced phase shifts are not the same for every dimension. Using the two-dimensional QPSK space-time codes in Figure 1 as examples, their decoding complexity is 64 trellis branch computations per information bit. A complexity-wise equivalent LST can allow 32 states for both in-phase and quadrature components, if they are binary coded. As a result, while the two-dimensional QPSK space-time code has constraint length 4, a complexity-wise equivalent LST can have constraint length 6 for both in-phase and quadrature components. This increased constraint length not only gives better coding gain, it also increases the

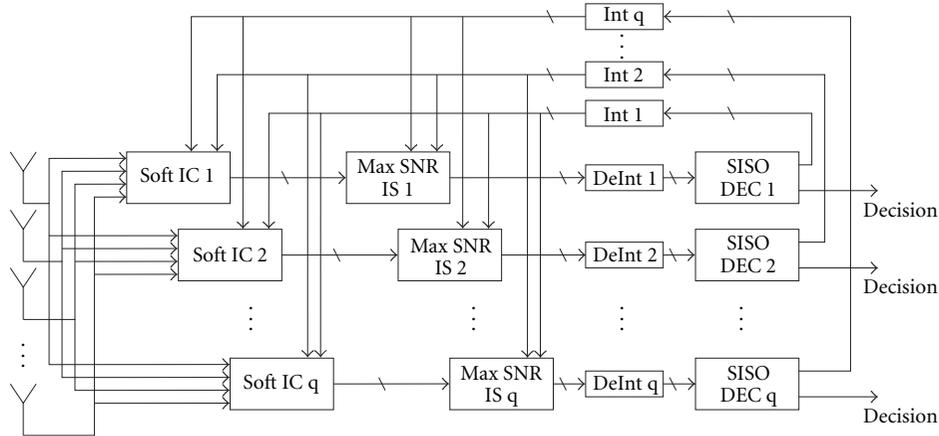


FIGURE 4: Iterative maximum SNR array processing and decoding of GST.

transmit diversity when the number of transmit antennas is large and rotation is used.

7. ITERATIVE PROCESSING

With a maximum SNR based front-end filter followed by the group decoders at the receiver, the GST finds itself an immediate application of the MMSE based iterative algorithm proposed in [13, 14]. In the GST application, multigroup is equivalent to multiuser in [13, 14]. The CDMA chip sampling is replaced by the spatial (antenna array) sampling, and the spreading sequences are replaced by the random channel gains seen by individual antennas. This algorithm has been implemented recently for the LST in [18, 19].

The iterative maximum SNR array processing and decoding algorithm is depicted in Figure 4. The basic idea of this algorithm is to concatenate the maximum SNR based front-end filter with the group decoder, then apply iterative processing [20] by properly exchanging soft information between them. The front-end filter provides interference-suppressed inputs to the group decoder. While the group decoder, being soft-input soft-output (SISO), computes the a posteriori probabilities of the coded symbols needed in the (soft) interference cancellation. The goal of the interference cancellation is to minimize the residual interference power after cancellation. It can be shown that the reconstructed interfering signals used in the cancellation should be the MMSE estimates of the original signals [21]. These MMSE estimates can be computed by taking expectations using the a posteriori probabilities. After the interference cancellation is done, a new maximum SNR filter is computed for the next iteration based on the residual interference powers. Intuitively, one can see that if the decoding result of a symbol (from other groups) is less reliable, its MMSE estimate will be farther away from the original, and its residual power after cancellation will be higher. When updating the maximum SNR filter, the filter weights will be adjusted to deemphasize the signal space of this symbol.

As the iterations go on, one can expect that the decoding

results, as well as the cancellation, becomes better. In the end, the cancellation might be good enough that the maximum SNR filters become matched filters which maximum ratio combine full receive diversity for all groups. When this is the case, the optimal power allocation is to assign equal power to every group.

To accelerate the convergence of the iterative process, it is desirable to (temporally) interleave the groups independently before rotation so the neighboring symbols in the two estimation stages (front-end filter and decoder) are as different as possible. For the GST with space-time coding, group-based temporal interleaving can be used to maintain the same transmit diversity after interleaving. For the LST, the same effect can be achieved by separately interleaving the code streams within a group.

8. NUMERICAL RESULTS

The design issues discussed in this paper are verified through simulation of an 8-transmit 8-recv multi-antenna system over a flat, Rayleigh, quasi-static (constant within a frame), and spatially independent channel. For the GST, we use similar simulation parameters as in [1]. That is, every transmit group contains two antennas; and a rate 1/2, 32-state, 4PSK space-time trellis code (Tarokh et al. in Figure 1) is used. The space-time codes for all groups are the same. Each frame consists of a total of 131 transmissions (128 data + 3 tail) from each transmit antenna. Using Figure 1 and the assumption of equally distributed source sequence, we have diagonal $E[\mathbf{c}_i^t (\mathbf{c}_i^t)^H]$. The extreme case with one antenna per group, namely, the LST, encodes its in-phase and quadrature components separately using rate 1/2 binary convolutional codes with maximal minimum free distances. Again, all groups use the same code. Each frame still consists of 128 data transmissions from each transmit antenna. The number of tail transmissions, however, depends on the allowable constraint lengths of individual systems.

We first demonstrate in Figure 5 the array processing and power allocation issues. The case in [1] with zero-forcing

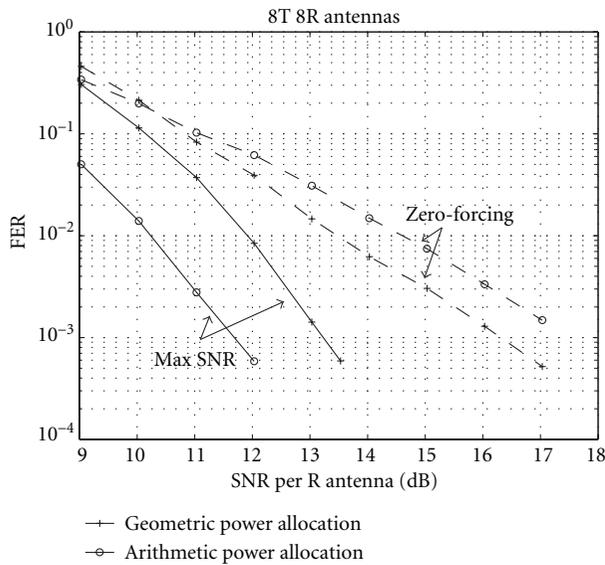


FIGURE 5: Performance comparison between zero-forcing and maximum SNR methods.

array processing and geometrically decreasing power allocation (8 : 4 : 2 : 1) among groups is reproduced. We then apply maximum SNR array processing and/or an arithmetically decreasing power allocation (4 : 3 : 2 : 1) to the same system. In this figure, it is shown clearly that the maximum SNR approach outperforms the zero-forcing approach. When the same (geometric) power allocation is used, the maximum SNR approach is 1.5 dB better at 10^{-2} frame error rate (FER), and its advantage increases as the signal power increases due to the steeper FER slope. The steeper slope is a result of higher receive diversity, as the maximum SNR array processor does not preclude the received signal components lying in the interference space. The divergence of the asymptotic performances of the zero-forcing and the maximum SNR approaches might seem contradictory to the common understanding. It is, however, necessary to clarify that the common understanding was built on the model which has a deterministic channel gain. With the quasi-static, random channel gains in our simulation, there is always a possibility that some of the channel gains are very small and the AWGN cannot be ignored, no matter how high the transmitted power is. It is these worst cases which limit the average error performance.

When arithmetic power allocation is applied, the maximum SNR approach performs even better. Its gain over the zero-forcing method is 3 dB. As arithmetic power allocation is not matched to the zero-forcing diversity gains, it worsens the performance 1 dB at 10^{-2} FER.

Figure 6 gives a comprehensive illustration of the design issues discussed. First, three curves from Figure 5 are shown again to demonstrate the advantages of the maximum SNR algorithm and the arithmetic power allocation, respectively. Then the performance when rotation is applied in addition to these two design considerations is shown (dia-

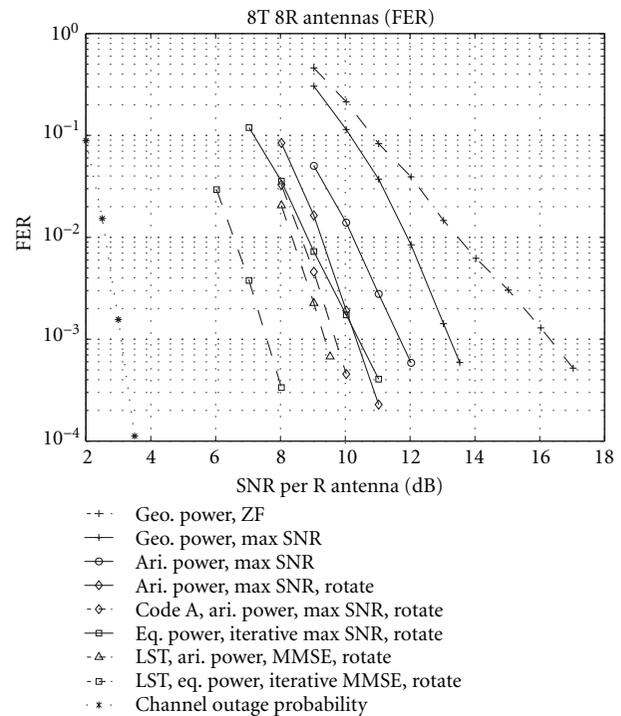


FIGURE 6: Performance when different design issues are considered.

mond marks). At 10^{-2} FER, rotation gives about 1 dB gain. The steeper slope of this configuration implies that rotation does increase the transmit diversity. A slight alteration of the Tarokh et al. code is also simulated. This code, referred to as Code A in Figure 1, has encoder polynomials (defined in [8, page 757]): (2, 2), (3, 3), (2, 0), (2, 2), (1, 1), (0, 2), (2, 1). It slightly outperforms the Tarokh et al. code.

The solid curve with square marks uses rotation and the iterative maximum SNR array processing and decoding described in Section 7. It has equal transmission powers for all groups. Due to the necessity of SISO decoding, the soft output Viterbi algorithm (SOVA) [22] is modified to generate soft outputs of the coded symbols. The SOVA decoding complexity is approximately twice the Viterbi decoder. Therefore, to maintain similar complexity, the space-time code used is the best 4-state code found in [23], and the number of array processing and decoding iterations is four. Group-based random temporal interleaving is applied to accelerate the decoding convergence. Due to its extremely small constraint length, this system only shows a diversity gain similar to the non-iterative system without rotation. Nevertheless, iterative processing does improve the decoding performance when SNR is low.

The LST performances are also given. The decoding complexities of these examples are kept the same as the GST configurations. According to the discussion in Section 6.2, when there is only one antenna per group, the in-phase and quadrature components can be separately encoded to increase the coding gain. The curve with triangular marks uses a rate 1/2 binary convolutional code with polynomials

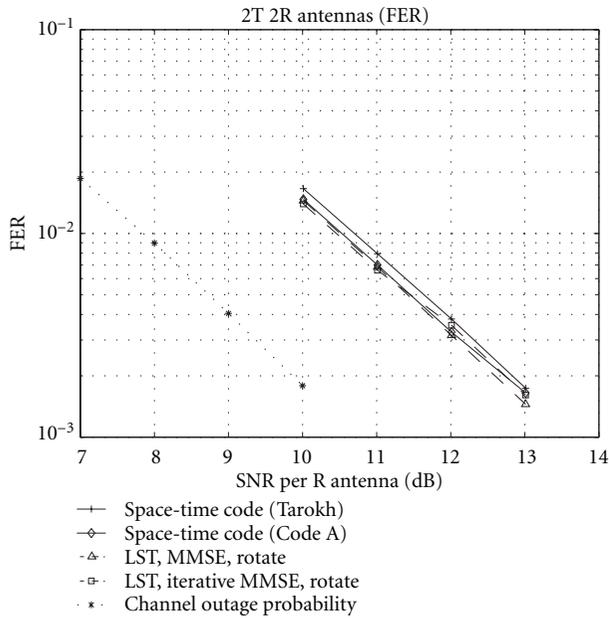


FIGURE 7: 2-transmit, 2-receive performance.

(65₈, 57₈) (32 states). For fair comparison with the corresponding GST cases, the power allocation of this system is the same (4 : 4 : 3 : 3 : 2 : 2 : 1 : 1). Its array processing method is MMSE. From Figure 6, we can see that the increased coding gain improves the performance. The slope of this curve, on the other hand, is very similar to the GST case with Code A. This is an implication that both codes (QPSK space-time code with constraint length 4 and binary convolutional code with constraint length 6) have the same degree of diversity after rotation.

The LST with iterative MMSE algorithm and equal group powers uses a (5₈, 7₈) (4 states) convolutional code. The number of decoding iterations is four. Random temporal interleaving is applied, separately and independently, to each of the four code streams (two in-phase, two quadrature) of every group. The performance of this system is about 2 dB better than the LST case with single-sweep decoding and hard decision feedback cancellation. As to the diversity gain, although this system has higher receive diversity due to better interference cancellation, its transmit diversity suffers from the shortened minimum free distance. The resultant performance slope of this system is very similar to that of the single-sweep LST case which has larger minimum free distance. At 10⁻² FER, this system is about 4 dB away from the channel outage capacity [3, 4]. Figure 6 also shows that, due to either imperfect signal (code) design or sub-optimal decoding, all these practical systems do not achieve the performance slope given by information theory.

Similar experiment was conducted for the 2-transmit 2-receive case where the space-time coding (GST with one group) system uses optimal joint detection, while the LST suffers from inefficient interference suppression due to the small number of receive antennas (see Figure 7). With space-time coding being able to achieve maximum diversity, the

LST is not superior anymore. At 10⁻² FER, all these systems are about 2.5 dB away from the channel outage capacity. They also seem to be able to achieve the information theoretic performance slope.

9. CONCLUSION

In this paper, we discussed some design issues of the open-loop space-time coded multi-antenna system. Through intuitive, as well as analytical explanations, we revealed one by one the advantages of the design considerations. To summarize, the zero-forcing array processing method in [1] can be improved with a maximum SNR approach. This new array processing approach then necessitates a different power allocation among the transmission groups. To increase the transmit diversity when there are more than one group, group-based spatial interleaving (“rotation”) can be applied. We also considered the LST as a special case of the GST which has one antenna per group. When the number of receive antennas is large enough to resolve the transmission signal spaces, the LST allows easier code design and separate in-phase and quadrature encoding/decoding. These advantages can increase the coding gain and possibly the transmit diversity when the decoding complexity is constrained. The GST with maximum SNR array processing also finds itself a direct application of the iterative algorithm proposed in [13, 14]. In the 8-antenna LST example we gave, the iterative algorithm improves the performance about 2 dB.

APPENDIX

Proof of Theorem 1. We consider the linear filtering in two steps. First, a linear filter bank $\Theta_M(\mathcal{C}_1)$ is used to filter the receive antenna outputs. The outputs of this filter bank are then combined to achieve maximum SNR. Without loss of generality, we assume that the interference components of the filter bank outputs are uncorrelated with one another. In other words, $\Theta_M^H(\mathcal{C}_1)\mathbf{R}_n\Theta_M(\mathcal{C}_1)$ is diagonal. This assumption is reasonable since for any filter bank consisting of k linear filters, we can find, using the singular-value decomposition (SVD) [24] on $\Theta_M^H(\mathcal{C}_1)\mathbf{R}_n\Theta_M(\mathcal{C}_1)$, a nonsingular $k \times k$ matrix to transform it and diagonalize the interference covariance matrix of its outputs. This transformation is reversible and does not destroy the information contained in the filter bank outputs.

As one of our goals is to minimize the number of linear filters required, we further assume that the linear filters of this filter bank are linearly independent of one another. If this is not true, we can always combine some of the filters to form a linearly independent filter bank with fewer filters.

With the above assumptions and the fact that \mathbf{R}_n is non-singular, the filter bank outputs will have nonzero and uncorrelated interference components. Under this condition, the maximum SNR combining of the filter bank outputs is maximum ratio combining. Given independent group transmission so that the signal and interference codeword (\mathbf{c}_t^1 and \mathbf{c}_t^0) expectations can be taken separately, the maximum SNR

filter bank satisfies

$$\max_{\mathbf{w}_1, \dots, \mathbf{w}_k} \sum_j \mathbf{w}_j^H \mathbf{R}_s \mathbf{w}_j \quad \text{subject to} \quad \mathbf{w}_i^H \mathbf{R}_n \mathbf{w}_i = \nu, \quad \forall i, \quad (\text{A.1})$$

where ν is a constant. With the quadratic form, this condition is equivalent to the following Lagrange multipliers

$$\frac{\partial}{\partial \mathbf{w}_i^*} \left(\sum_j \mathbf{w}_j^H \mathbf{R}_s \mathbf{w}_j - \sum_j \lambda_j \mathbf{w}_j^H \mathbf{R}_n \mathbf{w}_j \right) = \mathbf{0}, \quad \forall i. \quad (\text{A.2})$$

The solutions to the Lagrange multipliers satisfy

$$(\mathbf{R}_s - \lambda_i \mathbf{R}_n) \mathbf{w}_i = \mathbf{0}, \quad \forall i, \quad (\text{A.3})$$

with

$$\lambda_i = \frac{\mathbf{w}_i^H \mathbf{R}_s \mathbf{w}_i}{\mathbf{w}_i^H \mathbf{R}_n \mathbf{w}_i} \quad (\text{A.4})$$

being the SNRs of the maximum SNR filter outputs.

Clearly, we need only filters giving nonzero SNRs. To be consistent with the assumption of uncorrelated interference components, we can use Cholesky decomposition $\mathbf{R}_n = \mathbf{L}\mathbf{L}^H$ [24], as \mathbf{R}_n is Hermitian and positive definite. Premultiplying (16) by \mathbf{L}^{-1} , we get

$$\mathbf{C}(\mathbf{L}^H \mathbf{w}) = \lambda(\mathbf{L}^H \mathbf{w}), \quad (\text{A.5})$$

where

$$\mathbf{C} = \mathbf{L}^{-1} \mathbf{R}_s (\mathbf{L}^{-1})^H \quad (\text{A.6})$$

is Hermitian and nonnegative definite with rank k . According to (12), k can be no larger than the rank of $\Omega(\mathcal{C}_1)$, so $k \leq \min(m, n_1)$. The eigenvalues of \mathbf{C} are the same as those of the original problem (16). There are k nonzero eigenvalues, and a set of orthogonal eigenvectors $\{\mathbf{L}^H \mathbf{w}_i\}_{i=1}^k$ can be found by using SVD. It is then straightforward that

$$\mathbf{w}_i^H \mathbf{R}_n \mathbf{w}_j = \begin{cases} \nu, & i = j, \\ 0, & i \neq j, \end{cases} \quad \mathbf{w}_i^H \mathbf{R}_s \mathbf{w}_j = \begin{cases} \lambda_i \nu, & i = j, \\ 0, & i \neq j. \end{cases} \quad (\text{A.7})$$

Since

$$E[\mathbf{w}_i^H \mathbf{r}_i \mathbf{r}_i^H \mathbf{w}_j] = \mathbf{w}_i^H (\mathbf{R}_s + \mathbf{R}_n) \mathbf{w}_j, \quad (\text{A.8})$$

then the filter bank outputs are uncorrelated with one another. \square

ACKNOWLEDGMENTS

This work was presented at the IEEE International Conference on Third Generation Wireless Communications, San Francisco, California, USA, June 2000.

This work was supported in part by NASA through contract NCC3528 and in part by the Army Federated Lab through CTA contract DAAD19-01-2-0011.

REFERENCES

- [1] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Combined array processing and space-time coding," *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1121–1128, 1999.
- [2] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.
- [3] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.
- [4] E. Teletar, "Capacity of multi-antenna Gaussian channels," Tech. Rep., Internal Tech. Memo., AT&T-Bell Labs, June 1995.
- [5] N. Seshadri and J. H. Winters, "Two signaling schemes for improving the error performance of frequency-division-duplex (FDD) transmission systems using transmitter antenna diversity," *International Journal of Wireless Information Networks*, vol. 1, no. 1, pp. 49–60, 1994.
- [6] A. Wittneben, "Base station modulation diversity for digital SIMULCAST," in *Proc. IEEE Vehicular Technology Conference*, pp. 505–511, Secaucus, NJ, USA, May 1993.
- [7] A. Wittneben, "A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation," in *Proc. IEEE International Conference on Communications*, pp. 1630–1634, Geneva, Switzerland, June 1993.
- [8] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criteria and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, 1998.
- [9] J.-C. Guey, M. P. Fitz, M. R. Bell, and W.-Y. Kuo, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," *IEEE Trans. Communications*, vol. 47, no. 4, pp. 527–537, 1999.
- [10] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998.
- [11] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456–1467, 1999.
- [12] A. R. Hammons, Jr. and H. El Gamal, "On the theory of space-time codes for PSK modulation," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 524–542, 2000.
- [13] X. Wang and H. V. Poor, "Iterative (Turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Communications*, vol. 47, no. 7, pp. 1046–1061, 1999.
- [14] H. El Gamal and E. Geraniotis, "Iterative multi-user detection for coded CDMA signals in AWGN and fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 1, pp. 30–41, 2000.
- [15] M. L. Honig, U. Madhow, and S. Verdú, "Blind adaptive multiuser detection," *IEEE Transactions on Information Theory*, vol. 41, no. 4, pp. 944–960, 1995.
- [16] H. V. Poor and S. Verdú, "Probability of error in MMSE multiuser detection," *IEEE Transactions on Information Theory*, vol. 43, no. 3, pp. 858–871, 1997.
- [17] V. Tarokh and T. K. Y. Lo, "Principal ratio combining for fixed wireless applications when transmitter diversity is employed," *IEEE Communications Letters*, vol. 2, no. 8, pp. 223–225, 1998.
- [18] H. El Gamal and A. R. Hammons Jr., "A new approach to layered space-time coding and signal processing," *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2321–2334, 2001.
- [19] D. Shiu and J. M. Kahn, "Scalable layered space-time codes

- for wireless communications: performance analysis and design criteria,” in *Proc. IEEE Wireless Communications and Networking Conference*, pp. 159–163, New Orleans, La, USA, September 1999.
- [20] J. Hagenauer, E. Offer, and L. Papke, “Iterative decoding of binary block and convolutional codes,” *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 429–445, 1996.
- [21] F. Tarköy, “MMSE-optimal feedback and its applications,” in *Proc. IEEE International Symposium on Information Theory*, p. 334, Whistler, Canada, September 1995.
- [22] J. Hagenauer and P. Hoeher, “A Viterbi algorithm with soft-decision outputs and its applications,” in *Proc. GLOBECOM '89*, pp. 1680–1686, Dallas, Tex, USA, November 1989.
- [23] S. Bärö, G. Bauch, and A. Hansmann, “Improved codes for space-time trellis-coded modulation,” *IEEE Communications Letters*, vol. 4, no. 1, pp. 20–22, 2000.
- [24] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Englewood Cliffs, NJ, USA, 3rd edition, 1996.

Hsuan-Jung Su received his B.S. degree in electronic engineering from the National Chiao-Tung University, Taiwan, in 1992, and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland, College Park, in 1996 and 1999, respectively. From 1999 to 2000, he was a Postdoctoral Research Associate with the Institute for Systems Research, University of Maryland. Since November 2000, he has been with Bell Laboratories, Lucent Technologies, Holmdel, New Jersey, where he has been involved in the design and performance evaluation of adaptive coding/modulation, fast Hybrid-ARQ, scheduling, and Radio Link Control protocol for 3G wireless networks. His research interests cover coding, modulation, signal processing, power control and synchronization of narrowband and wideband communication systems.



Evangelos Geraniotis obtained his Ph.D. in electrical engineering from the University of Illinois at Urbana-Champaign in 1982. He has been with the University of Maryland since 1985 where he is now Professor of Electrical Engineering and a member of the Institute for Systems Research and the Center of Satellite and Hybrid Communication Networks. Dr. Geraniotis's research has been in Communication Systems and Networks. In the communication systems area, his earlier research has focused on spread-spectrum and anti-jam communications; receiver design for fading channels; and schemes for interception, feature-detection, and classification of signals. His recent work pertains to several design issues of DS/CDMA, FH/SSMA, and OFDM wireless communications, including power control, advanced modulation, FEC coding, array processing, and interference cancellation techniques, as well as retransmission schemes, MAC protocols, handoff, and switching schemes. His research on communication networks has encompassed channel and traffic modeling, performance evaluation, and design of multi-access protocols for mobile, satellite, cellular, and PCS networks; and multi-media integration schemes for wireless networks, high-speed ATM networks, and hybrid satellite/terrestrial networks. He is the author of over 200 technical papers in journals and conference proceedings on the above areas. He has been consulting regularly for Government and Industry in the above areas. Dr. Geraniotis was an Editor for Spread-Spectrum of the *IEEE Transactions on Communications* from 1989 to 1992.