V. Conclusion

We have considered the capacity of multiple-antenna systems in Rayleigh flat fading under the assumption that CSI is available at both ends of the system. First, we derived the capacity of such systems in the case when only the transmitter is equipped with multiple antennas. We showed that the capacity of this system is, in fact, the same as a receiver-only diversity system with maximal ratio combining. We also proposed a transmission diversity scheme (maximal gain transmission) that is mathematically equivalent to a receiver-only diversity system with selection combining and evaluated its capacity.

Next, we derived capacity expressions for a general system with multiple antennas at both transmitter and receiver. We showed that the optimal power allocation is given by a matrix water-filling algorithm. We obtained an equation that determines the cutoff value for such systems, which can be evaluated via numerical root-finding, and a corresponding closed-form expression for the capacity with optimal power and rate adaptation. We evaluated this capacity for some representative situations and demonstrated similarities with the capacity of such systems when CSI is available only at the receiver end.

In all these cases, the only step that required numerical techniques in determining the capacity is the evaluation of the cutoff value γ_0. In order to circumvent this problem, we also derived approximations to the cutoff value for all cases considered. Numerical results show that these approximations yield good capacity estimates when the SNR or the number of antennas is sufficiently large.

From these capacity computations for multiple-antenna systems with adaptive transmission techniques we observe that large capacity gains are possible compared to the receiver-CSI-only systems. The tradeoff for these increased capacity values is the outage probability incurred by the adaptive power and rate allocation schemes. We derived simple upper bounds for this outage probability and showed that the channel outage probability may also be decreased by increasing the number of antennas.

References

[1] T. Ericson, "A Gaussian channel with slow fading," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 353–355, May 1970.

[2] L. H. Ozarow, S. Shamai (Shitz), and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359–378, May 1994.

[3] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1986–1992, Nov. 1997.

[4] H. Viswanathan, "Capacity of Markov channels with receiver CSI and delayed feedback," *IEEE Trans. Inform. Theory*, vol. 45, pp. 761–771, May 1999.

[5] I. C. Abou-Faycal, M. D. Trott, and S. Shamai (Shitz), "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1290–1301, May 2001.

[6] J. S. Richters, "Communication over fading dispersive channels," MIT Res. Lab. Electronics, Cambridge, MA, Tech. Rep. 464, Nov. 30, 1967.

[7] S. Shamai (Shitz) and I. Bar-David, "The capacity of average and peak-power-limited quadrature Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1060–1071, July 1995.

[8] J. G. Smith, "On the information capacity of peak and average power constrained Gaussian channels," Ph.D. dissertation, Dept. Elec. Eng., Univ. California, Berkeley, 1969.

[9] J. G. Smith, "The Information capacity of amplitude and variance-constrained scalar Gaussian channels," *Inform. Contr.*, vol. 18, pp. 203–219, 1971.

[10] G. J. Foschini, "Layered space-time architecture for wireless communication in a flat fading environment when using multi-element antennas," *Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[11] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311–335, 1998.

[12] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.

[13] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139–157, Jan. 1999.

[14] M. S. Alouini and A. J. Goldsmith, "Capacity of Rayleigh fading channels under different adaptive transmission and diversity-combining techniques," *IEEE Trans. Veh. Technol.*, vol. 48, pp. 1165–1181, July 1999.

[15] W. C. Jakes, *Microwave Mobile Communications*. Piscataway, NJ: IEEE Press, 1994.

[16] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[17] P. Lancaster and M. Tismenetsky, *The Theory of Matrices with Applications*. Orlando, FL: Academic, 1985.

[18] A. T. James, "Distributions of matric variates and latent roots derived from normal samples," *Ann. Math. Statist.*, vol. 35, pp. 475–501, June 1964.

[19] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[20] A. Leon-Garcia, *Probability and Random Processes for Electrical Enginners*. Reading, MA: Addison-Wesley, 1994.

[21] A. Erdelyi, Ed., *Higher Transcendental Functions*. New York: McGraw-Hill, 1953.

[22] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products*. New York: Academic, 1965.

[23] H. H. Goldstine and J. von Neumann, "Numerical inverting of matrices of high order II," *Proc. Amer. Math. Soc.*, vol. 2, pp. 188–202, 1951.

[24] A. Edelman, "Eigenvalues and condition numbers of random matrices," Ph.D. dissertation, Dept. Math., MIT, Cambridge, MA, 1989.

On the Separability of Demodulation and Decoding for Communications Over Multiple-Antenna Block-Fading Channels

Yibo Jiang, *Student Member, IEEE*, Ralf Koetter, *Member, IEEE*, and Andrew C. Singer, *Member, IEEE*

***Abstract*—We study the separability of demodulation and decoding for communications over multiple-antenna block-fading channels when bit-linear linear dispersion (BL-LD) codes are used. We assume the channel is known to the receiver only, and find necessary and sufficient conditions on the dispersion matrices for the separation of demodulation and decoding at the receiver without loss of optimality.**

***Index Terms*—Multiple-antenna systems, space–time codes.**

I. Introduction

We consider a multiple-input–multiple-output (MIMO) communication setup with t transmit antennas and r receive antennas. Information-theoretic results by Foschini and Gans [1] and Telatar [2] have sparked tremendous interest and effort in the design of practical channel codes for communications over multiple-antenna channels (i.e., MIMO chan-

Manuscript received October 29, 2002; revised April 29, 2003. This work was supported in part by the National Science Foundation under Grants CCR-9979381, ITR 00-85929, and CCR-9984515, and in part by the Motorola Center for Communications. The material in this correspondence was presented at the IEEE International Symposium on Information theory, Yokohama, Japan, June/July 2003.

The authors are with the Coordinated Science Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: yjiang@uiuc.edu; koetter@uiuc.edu; acsinger@uiuc.edu).

Communicated by B. M. Hochwald, Guest Editor.

Digital Object Identifier 10.1109/TIT.2003.817478

nels). Some prominent schemes include Bell Labs layered space–time (BLAST) [3], space–time trellis codes [4], space–time block codes (STBCs) from orthogonal designs [5], [6], and linear dispersion codes (LD codes) [7].

In this correspondence, we study bit-linear linear dispersion (BL-LD) codes, a class of LD codes whose codeword matrices are linear in each information bit. In a multiple-antenna communication system, if a BL-LD code is used as the modulation code and there is some channel coding preceding it, joint maximum-likelihood demodulation and decoding is optimal in the sense of minimizing the error probability. Here, we treat the decoding of BL-LD codes as demodulation. We find necessary and sufficient conditions on the dispersion matrices of the BL-LD code such that the probability distribution of the information bit vector (at the input of the BL-LD encoder) conditioned on the channel output and the channel state information (CSI) is a product distribution. Therefore, under those conditions, demodulation and decoding can be separated without loss of optimality. The demodulator outputs a log-likelihood ratio (LLR) vector.

Our notation will be as follows: capital letters denote matrices; underscores denote vectors; boldfaced letters denote random objects; daggers denote complex conjugate transpose; asterisks denote complex conjugation.

This correspondence is organized as follows. In Section II, we describe the signal model used in this correspondence. In Section III, we will present our main result, the necessary and sufficient conditions for the separation of demodulation and decoding without loss of optimality. We will show that the probability distribution of the LLR vector conditioned on the CSI is a product distribution. Section IV contains our conclusions.

II. Signal Model

We assume that the MIMO channel contains t transmit antennas and r receive antennas. The coefficient $\boldsymbol{h}_{i,j}$ denotes the complex channel gain between transmit antenna j and receive antenna i. We assume that $\boldsymbol{h}_{i,j}$ experiences independent and identically distributed (i.i.d.) frequency flat fading for all i and j. The CSI matrix $\boldsymbol{H} = [\boldsymbol{h}_{i,j}] \in \mathbb{C}^{r\times t}$ is known to the receiver only and does not change within a block of T symbols. From block to block, $\boldsymbol{H}$ changes independently (block fading). Let $\boldsymbol{X} \in \mathbb{C}^{t\times T}$ indicate the transmitted signal matrix, $\boldsymbol{Y} \in \mathbb{C}^{r\times T}$ be the received signal matrix such that the discrete-time baseband equivalent channel model is

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{N} \tag{1}$$

where $\boldsymbol{N} \in \mathbb{C}^{r\times T}$ is an additive noise matrix with i.i.d. entries $\boldsymbol{n}_{i,j} \sim \mathcal{CN}(0, 2\sigma^2)$.

We focus on the study of a class of LD codes which is defined as follows.

Definition 1: A BL-LD code $\mathcal{C}$ is an LD code whose codeword matrices are linear in the information bits

$$\mathcal{C} = \left\{X\colon X = \sum_{k=1}^{K} b_k A_k,\ b_k \in \{-1, +1\}, \forall\, k\right\} \tag{2}$$

where $A_k \in \mathbb{C}^{t\times T}$, $1 \le k \le K$ are the dispersion matrices.

It is easy to verify that with pulse amplitude modulation (PAM) or quadrature amplitude modulation (QAM) (with each dimension as an independent PAM), the STBCs [6] and LD codes [7] are BL-LD codes.

We assume that the information bits are independent and equiprobable.

III. The Separability of Demodulation and Decoding

Given the channel output, it is well known that joint maximum-likelihood demodulation and decoding is optimal in the sense of minimizing the error probability. However, if the probability distribution of the modulator input $\boldsymbol{b}_1, \dots, \boldsymbol{b}_K$ conditioned on the channel output $\boldsymbol{Y} = Y$ and the CSI $\boldsymbol{H} = H$ is a product distribution, i.e.,

$$p(\boldsymbol{b}_1 = b_1, \dots, \boldsymbol{b}_K = b_K | Y, H) = \prod_{k=1}^{K} p(\boldsymbol{b}_k = b_k | Y, H) \tag{3}$$

then demodulation and decoding can be separated without loss of optimality. Define the LLR for bit $\boldsymbol{b}_k$ as

$$L_e(\boldsymbol{b}_k) \triangleq \ln \frac{p(\boldsymbol{b}_k = +1 | Y, H)}{p(\boldsymbol{b}_k = -1 | Y, H)}. \tag{4}$$

If (3) is satisfied, the demodulator outputs a K-dimensional LLR vector $(L_e(\boldsymbol{b}_1), \dots, L_e(\boldsymbol{b}_K))$ which contributes a sufficient statistic for decoding. Here we treat the decoding of BL-LD codes as demodulation. The following theorem states necessary and sufficient conditions on the dispersion matrices $\{A_1, \dots, A_K\}$ for the separation of demodulation and decoding without loss of optimality.

Theorem 1: Let $\mathcal{C}$ be a BL-LD code with dispersion matrices $A_k \in \mathbb{C}^{t\times T}$, $1 \le k \le K$. The probability distribution of $\boldsymbol{b}_1, \dots, \boldsymbol{b}_K$ conditioned on the channel output $\boldsymbol{Y} = Y$ and the CSI $\boldsymbol{H} = H$ is a product distribution for all Y and H, i.e.,

$$p(\boldsymbol{b}_1 = b_1, \dots, \boldsymbol{b}_K = b_K | Y, H) = \prod_{k=1}^{K} p(\boldsymbol{b}_k = b_k | Y, H)$$

if and only if $A_i A_j^\dagger + A_j A_i^\dagger = \mathbf{0}$ for all i, j, and $i \ne j$.

Proof: See the Appendix. □

If the number of information bits K is large, it may be computationally inconvenient to verify the conditions $A_i A_j^\dagger + A_j A_i^\dagger = \mathbf{0}$ for $1 \le i \le K$, $1 \le j \le K$, and $i \ne j$. Next, we derive necessary and sufficient conditions on the codeword matrices of the BL-LD codes for the requirement $A_i A_j^\dagger + A_j A_i^\dagger = \mathbf{0}$, $\forall\, i, \forall j, i \ne j$ to hold. In some cases, the conditions on the codeword matrices are easier to verify than the conditions on the dispersion matrices.

Lemma 1: Let $\mathcal{C}$ be a BL-LD code

$$\mathcal{C} = \left\{X(\underline{b})\colon X(\underline{b}) = \sum_{k=1}^{K} b_k A_k,\ \underline{b} = (b_1, \dots, b_K) \in \{-1, +1\}^K\right\}$$

with dispersion matrices $A_k \in \mathbb{C}^{t\times T}$, $1 \le k \le K$. Matrices A_i and A_j satisfy the condition $A_i A_j^\dagger + A_j A_i^\dagger = \mathbf{0}$, $\forall\, i, \forall\, j$, $i \ne j$ if and only if the codeword matrix $X(\underline{b})$ satisfies $X(\underline{b}) X^\dagger(\underline{b}) = C_0$, $\forall\, \underline{b}$. The constant matrix C_0 must be equal to $\sum_{k=1}^{K} A_k A_k^\dagger$.

Proof: We first prove the only if part of the claim.

If $A_i A_j^\dagger + A_j A_i^\dagger = \mathbf{0}$, we have

$$\begin{aligned} X(\underline{b}) X(\underline{b})^\dagger &= \left(\sum_{k=1}^{K} b_k A_k\right)\left(\sum_{l=1}^{K} b_l A_l\right)^\dagger \\ &= \sum_{k=1}^{K} A_k A_k^\dagger \triangleq C_0. \end{aligned} \tag{5}$$

We now prove the direct part of the claim.

Define $\hat{\underline{b}}$ to be the vector consisting of the elements of $\underline{b}$ except for b_i and b_j, and define $\hat{X}(\hat{\underline{b}}) = X(\underline{b}) - b_i A_i - b_j A_j$, then we have

$$\left[\hat{X}(\hat{\underline{b}}) + b_i A_i + b_j A_j\right]\left[\hat{X}(\hat{\underline{b}}) + b_i A_i + b_j A_j\right]^\dagger = C_0. \tag{6}$$

After some manipulation, we obtain

$$\begin{aligned} b_i &\left[\hat{X}(\hat{\underline{b}}) A_i^\dagger + A_i \hat{X}(\hat{\underline{b}})^\dagger\right] + b_j \left[\hat{X}(\hat{\underline{b}}) A_j^\dagger + A_j \hat{X}(\hat{\underline{b}})^\dagger\right] \\ &+ b_i b_j (A_i A_j^\dagger + A_j A_i^\dagger) \\ &= C_0 - \hat{X}(\hat{\underline{b}})(\hat{X}(\hat{\underline{b}}))^\dagger - A_i A_i^\dagger - A_j A_j^\dagger. \end{aligned} \tag{7}$$

If we fix $\hat{\underline{b}}$, the above matrix equation should hold for arbitrary $(b_i, b_j) \in \{-1, 1\}^2$. Before proceeding, we first consider the following one-dimensional problem. If the equation

$$\alpha_1 b_i + \alpha_2 b_j + \alpha_3 b_i b_j = \alpha_4$$

holds for all $(b_i, b_j) \in \{-1, 1\}^2$, we obtain the following four equations by considering the possible values of (b_i, b_j):

$$\alpha_1 + \alpha_2 + \alpha_3 = \alpha_4, \qquad \alpha_1 - \alpha_2 - \alpha_3 = \alpha_4$$
$$-\alpha_1 + \alpha_2 - \alpha_3 = \alpha_4, \quad -\alpha_1 - \alpha_2 + \alpha_3 = \alpha_4.$$

It follows that the constant coefficients must satisfy $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$. Applying this result, we obtain

$$\begin{aligned} A_i A_j^\dagger + A_j A_i^\dagger &= \mathbf{0} \\ \hat{X}(\hat{\underline{b}}) A_i^\dagger + A_i \hat{X}(\hat{\underline{b}})^\dagger &= \mathbf{0} \\ \hat{X}(\hat{\underline{b}}) A_j^\dagger + A_j \hat{X}(\hat{\underline{b}})^\dagger &= \mathbf{0} \\ C_0 - \hat{X}(\hat{\underline{b}})(\hat{X}(\hat{\underline{b}}))^\dagger - A_i A_i^\dagger - A_j A_j^\dagger &= \mathbf{0}. \end{aligned}$$

It then follows that $C_0 = \sum_{k=1}^K A_k A_k^\dagger$. □

Now we obtain two different forms of the necessary and sufficient conditions for the separation of demodulation and decoding without loss of optimality.

Corollary 1: Let $\mathcal{C}$ be a BL-LD code

$$\mathcal{C} = \left\{ X(\underline{b}) \colon X(\underline{b}) = \sum_{k=1}^K b_k A_k,\ \underline{b} = (b_1, \dots, b_K) \in \{-1, +1\}^K \right\}$$

with dispersion matrices $A_k \in \mathbb{C}^{t \times T}$, $1 \le k \le K$. The following three statements are equivalent:

1) $p(\boldsymbol{b}_1 = b_1, \dots, \boldsymbol{b}_K = b_K | Y, H) = \prod_{k=1}^K p(\boldsymbol{b}_k = b_k | Y, H)$;
2) $A_i A_j^\dagger + A_j A_i^\dagger = \mathbf{0}$, $\forall i, \forall j, i \ne j$;
3) $X(\underline{b}) X^\dagger(\underline{b}) = C_0$, $\forall \underline{b}$.

Proof: Follows from Theorem 1 and Lemma 1. □

Next, we give a BL-LD code example which satisfies Corollary 1.

The STBCs from orthogonal designs can be described in the following way:

$$\mathcal{C} = \left\{ X \colon X = \sum_{k=1}^M (S_k B_k + S_k^* \tilde{B}_k) \right\} \tag{8}$$

where S_k, $1 \le k \le M$ are the modulated symbols; B_k, $1 \le k \le M$ and $\tilde{B}_k$, $1 \le k \le M$, are complex coefficient matrices. If binary phase-shift keying (BPSK) or quaternary phase-shift keying (QPSK) is used, as mentioned in Section II, the STBCs from orthogonal designs are BL-LD codes. Furthermore, the codeword matrix X satisfies $XX^\dagger = (\sum_{k=1}^M |S_k|^2) I_t = M |S_1|^2 I_t$. Applying Corollary 1, we conclude that for the STBCs from orthogonal designs (with BPSK or QPSK modulation), the probability distribution of $\boldsymbol{b}_1, \dots, \boldsymbol{b}_K$ conditioned on the channel output $\boldsymbol{Y} = Y$ and the CSI $\boldsymbol{H} = H$ is a product distribution.

The question arises as to whether or not there is such a BL-LD code that it is not from an orthogonal design but satisfying Corollary 1. The answer is Yes. The following BL-LD code is an example:

$$X = \begin{pmatrix} b_1 & b_2 \\ b_1 & b_2 \end{pmatrix}.$$

In [8], for several STBCs from orthogonal designs, simple decoding algorithms were provided and it was shown that maximum-likelihood decoding of the symbols $S_1, \dots, S_M$ can be decoupled. It is equivalent to say the probability distribution of the symbols $S_1, \dots, S_M$ conditioned on the channel output and the CSI is a product distribution, i.e.,

$$p(\boldsymbol{S}_1 = S_1, \dots, \boldsymbol{S}_M = S_M | Y, H) = \prod_{k=1}^M p(\boldsymbol{S}_k = S_k | Y, H).$$

In this case, if each symbol is modulated by w bits, the STBC decoder outputs an $M2^w$-dimensional probability vector.

Now assume the demodulation and decoding can be separated without loss of optimality. The demodulator outputs a K-dimensional LLR vector $(L_e(\boldsymbol{b}_1), \dots, L_e(\boldsymbol{b}_K))$ which contributes a sufficient statistic for decoding. The following lemma states the properties of the probability distribution of the LLR vector.

Lemma 2: Let $\mathcal{C}$ be a BL-LD code

$$\mathcal{C} = \left\{ X(\underline{b}) \colon X(\underline{b}) = \sum_{k=1}^K b_k A_k,\ \underline{b} = (b_1, \dots, b_K) \in \{-1, +1\}^K \right\}$$

with dispersion matrices $A_k \in \mathbb{C}^{t \times T}$, $1 \le k \le K$. If $A_i A_j^\dagger + A_j A_i^\dagger = \mathbf{0}$, $\forall i, \forall j, i \ne j$, the following two properties hold:

1) $$\begin{aligned} p(\boldsymbol{L_e}(\boldsymbol{b}_1) = L_1, \dots, \boldsymbol{L_e}(\boldsymbol{b}_K) = L_K | H, b_1, \dots, b_K) \\ = \prod_{k=1}^K p(\boldsymbol{L_e}(\boldsymbol{b}_k) = L_k | H, b_k); \end{aligned}$$

2) $$\begin{aligned} p(\boldsymbol{L_e}(\boldsymbol{b}_1) = L_1, \dots, \boldsymbol{L_e}(\boldsymbol{b}_K) = L_K | H) \\ = \prod_{k=1}^K p(\boldsymbol{L_e}(\boldsymbol{b}_k) = L_k | H). \end{aligned}$$

Proof: See the Appendix. □

IV. Conclusion

We have studied several properties of BL-LD codes, including necessary and sufficient conditions on the dispersion matrices for the *a posteriori* distribution of the information bit vector to be a product distribution, i.e., the separation of demodulation and decoding at the receiver without loss of optimality. These properties are also useful in the derivation of a variety of mutual information inequalities with respect to $I(\boldsymbol{b}_1, \dots, \boldsymbol{b}_K; \boldsymbol{Y} | \boldsymbol{H})$, which can be used to aid in the design of the dispersion matrices for BL-LD codes. This is an area of ongoing research.

Appendix I
Proof of Theorem 1

Here we prove that necessary and sufficient conditions for

$$p(\boldsymbol{b}_1 = b_1, \dots, \boldsymbol{b}_K = b_K | Y, H)$$

to be a product distribution are $A_i A_j^\dagger + A_j A_i^\dagger = \mathbf{0}$ for all i, j, and $i \ne j$.

Proof: Define the vector $\underline{b}$ as $\underline{b} = (b_1, \dots, b_K)$. Since

$$\boldsymbol{Y} = \sum_{k=1}^K \boldsymbol{b}_k \boldsymbol{H} A_k + \boldsymbol{N}$$

we have

$$\begin{aligned} &p(\underline{\boldsymbol{b}} = \underline{b}, \boldsymbol{Y} = Y, \boldsymbol{H} = H) \\ &= p(\underline{b}) p(H) p(Y | H, \underline{b}) \\ &= 2^{-K} p(H) (\sqrt{2\pi}\sigma)^{-2rT} \\ &\quad \cdot \exp\left\{ -\frac{\operatorname{tr}\left\{ \left(Y - \sum_{k=1}^K b_k H A_k\right)\left(Y - \sum_{k=1}^K b_k H A_k\right)^\dagger \right\}}{2\sigma^2} \right\} \\ &= 2^{-K} p(H) \left(\sqrt{2\pi}\sigma\right)^{-2rT} \\ &\quad \cdot \exp\left\{ -\frac{\operatorname{tr}\left\{ YY^\dagger + \sum_{k=1}^K H A_k A_k^\dagger H^\dagger \right\}}{2\sigma^2} \right\} \end{aligned}$$

$$\cdot \exp\left\{\frac{\mathrm{tr}\left\{\sum_{k=1}^{K} b_k(YA_k^\dagger H^\dagger + HA_kY^\dagger)\right\}}{2\sigma^2}\right\}$$
$$\cdot \exp\left\{\frac{\mathrm{tr}\left\{-\sum_{i=1}^{K-1}\sum_{j=i+1}^{K} b_ib_jH(A_iA_j^\dagger + A_jA_i^\dagger)H^\dagger\right\}}{2\sigma^2}\right\}$$
$$= \alpha(H, Y) \cdot \exp\left\{\sum_{k=1}^{K} b_kL_k(Y, H) - \sum_{i=1}^{K-1}\sum_{j=i+1}^{K} b_ib_jG_{i,j}(H)\right\} \tag{9}$$

where

$$\alpha(H, Y) \triangleq 2^{-K}p(H)\left(\sqrt{2\pi}\sigma\right)^{-2rT} \cdot \exp\left\{-\frac{\mathrm{tr}\left\{YY^\dagger + \sum_{k=1}^{K} HA_kA_k^\dagger H^\dagger\right\}}{2\sigma^2}\right\}$$
$$L_k(Y, H) \triangleq \frac{\mathrm{tr}\{YA_k^\dagger H^\dagger + HA_kY^\dagger\}}{2\sigma^2}$$

and

$$G_{i,j}(H) \triangleq \frac{\mathrm{tr}\{H(A_iA_j^\dagger + A_jA_i^\dagger)H^\dagger\}}{2\sigma^2}. \tag{10}$$

We first prove the direct part of the claim.

If $A_iA_j^\dagger + A_jA_i^\dagger = \mathbf{0}$ for all i, j, and $i \neq j$, we have $G_{i,j}(H) = 0$ and from (9), we obtain that

$$p(\underline{b}, Y, H) = \alpha(H, Y)\exp\left\{\sum_{i=1}^{K} b_iL_i(Y, H)\right\}. \tag{11}$$

Then, we have

$$\frac{p(\boldsymbol{b}_k = 1|Y, H)}{p(\boldsymbol{b}_k = -1|Y, H)} = \frac{p(\boldsymbol{b}_k = 1, Y, H)}{p(\boldsymbol{b}_k = -1, Y, H)} = \frac{\sum_{\underline{b}:\, b_k = 1} p(\underline{b}, Y, H)}{\sum_{\underline{b}:\, b_k = -1} p(\underline{b}, Y, H)} = e^{2L_k(Y, H)}. \tag{12}$$

It follows that

$$\prod_{k=1}^{K} p(\boldsymbol{b}_k = b_k|Y, H) = \prod_{k=1}^{K}\frac{e^{b_kL_k(Y, H)}}{e^{L_k(Y, H)} + e^{-L_k(Y, H)}}. \tag{13}$$

On the other hand, after some manipulation, we obtain

$$p(\underline{\boldsymbol{b}} = \underline{b}|Y, H) = \frac{p(\underline{b}, Y, H)}{p(Y, H)} = \frac{\exp\left\{\sum_{k=1}^{K} b_kL_k(Y, H)\right\}}{\sum_{\tilde{b}_1\in\{-1,+1\}}\cdots\sum_{\tilde{b}_K\in\{-1,+1\}}\exp\left\{\sum_{k=1}^{K}\tilde{b}_kL_k(Y, H)\right\}}$$
$$= \frac{\exp\left\{\sum_{k=1}^{K} b_kL_k(Y, H)\right\}}{\prod_{k=1}^{K}\sum_{\tilde{b_k}\in\{-1,+1\}}\exp\{\tilde{b}_kL_k(Y, H)\}} = \prod_{k=1}^{K}\frac{e^{b_kL_k(Y,H)}}{e^{L_k(Y,H)} + e^{-L_k(Y,H)}}. \tag{14}$$

Comparing (13) and (14), we arrive at the product distribution.

We now prove the only if part of the claim. According to (9), we have

$$\frac{p(\underline{\boldsymbol{b}} = (b_1, \ldots, b_{i-1}, +1, b_{i+1}, \ldots, b_K)|Y, H)}{p(\underline{\boldsymbol{b}} = (b_1, \ldots, b_{i-1}, -1, b_{i+1}, \ldots, b_K)|Y, H)} = \exp\left\{2L_i(Y, H) - 2\sum_{j=1}^{i-1} b_jG_{j,i}(H) - 2\sum_{j=i+1}^{K} b_jG_{i,j}(H)\right\}.$$

If $p(\boldsymbol{b}_1 = b_1, \ldots, \boldsymbol{b}_K = b_K|Y, H)$ is a product distribution, we obtain

$$\exp\left\{2L_i(Y, H) - 2\sum_{j=1}^{i-1} b_jG_{j,i}(H) - 2\sum_{j=i+1}^{K} b_jG_{i,j}(H)\right\} = \frac{p(\boldsymbol{b}_i = +1|Y, H)}{p(\boldsymbol{b}_i = -1|Y, H)}. \tag{15}$$

Since the right-hand side (RHS) of (15) is invariant to $b_j \in \{-1, +1\}$ for all $j \neq i$ and must hold for all H, it follows $G_{j,i}(H) = 0$ if $1 \le j \le i-1$ and $G_{i,j}(H) = 0$ if $i+1 \le j \le K$. According to the definition in (10), it follows that $A_iA_j^\dagger + A_jA_i^\dagger = \mathbf{0}$ for $1 \le i \le K$, $1 \le j \le K$, and $i \neq j$ since H is arbitrary. □

Appendix II
Proof of Lemma 2

Here we prove two properties of the probability distribution of the LLR vector in Lemma 2.

Proof: We have

$$Y = HX + N = \sum_{k=1}^{K} b_kHA_k + N.$$

If $A_iA_j^\dagger + A_jA_i^\dagger = \mathbf{0}$, $i \neq j$, from the definition of LLR in (4), and along with (12), we have

$$L_e(\boldsymbol{b}_k) = \frac{\mathrm{tr}\{YA_k^\dagger H^\dagger + HA_kY^\dagger\}}{\sigma^2} = \frac{\mathrm{tr}\left\{\sum_{i=1}^{K} b_iH(A_iA_k^\dagger + A_kA_i^\dagger)H^\dagger + NA_k^\dagger H^\dagger + HA_kN^\dagger\right\}}{\sigma^2}$$
$$= \frac{\mathrm{tr}\{2HA_kA_k^\dagger H^\dagger\}}{\sigma^2}b_k + \frac{\mathrm{tr}\{NA_k^\dagger H^\dagger + HA_kN^\dagger\}}{\sigma^2}.$$

Conditioned on $\boldsymbol{H} = H$, we have

$$\frac{\mathrm{tr}\{\boldsymbol{N}(HA_k)^\dagger + HA_k\boldsymbol{N}^\dagger\}}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{r}\sum_{j=1}^{T}[\boldsymbol{n}_{i,j}(HA_k)^*_{i,j} + (HA_k)_{i,j}\boldsymbol{n}^*_{i,j}]. \tag{16}$$

Thus, the left-hand side of (16) is a zero mean Gaussian random variable since $\boldsymbol{n}_{i,j}$, $\forall i, j$ are i.i.d. zero mean complex Gaussian random variables with variance $2\sigma^2$. Hence, conditioned on H and b_k, $\boldsymbol{L_e}(\boldsymbol{b}_k)$ is a Gaussian random variable. Furthermore, conditioned on H and $b_1, \ldots, b_K$, $(\boldsymbol{L_e}(\boldsymbol{b}_1), \ldots, \boldsymbol{L_e}(\boldsymbol{b}_K))$ is a Gaussian random vector. In order to prove the first property in the lemma, we need to show that $\boldsymbol{L_e}(\boldsymbol{b}_i)$ and $\boldsymbol{L_e}(\boldsymbol{b}_j)$ are conditionally uncorrelated. We have

$$\mathrm{E}[\mathrm{tr}\{\boldsymbol{N}(HA_i)^\dagger + HA_i\boldsymbol{N}^\dagger\}\mathrm{tr}\{\boldsymbol{N}(HA_j)^\dagger + HA_j\boldsymbol{N}^\dagger\}]$$
$$= \mathrm{E}\left[\left(\sum_{k=1}^{r}\sum_{l=1}^{T}[\boldsymbol{n}_{k,l}(HA_i)^*_{k,l} + (HA_i)_{k,l}\boldsymbol{n}^*_{k,l}]\right) \cdot \left(\sum_{m=1}^{r}\sum_{q=1}^{T}[\boldsymbol{n}_{m,q}(HA_j)^*_{m,q} + (HA_j)_{m,q}\boldsymbol{n}^*_{m,q}]\right)\right]$$
$$= 2\sigma^2\sum_{k=1}^{r}\sum_{l=1}^{T}[(HA_i)^*_{k,l}(HA_j)_{k,l} + (HA_i)_{k,l}(HA_j)^*_{k,l}]$$
$$= 2\sigma^2[\mathrm{tr}\{(HA_j)(HA_i)^\dagger\} + \mathrm{tr}\{(HA_i)(HA_j)^\dagger\}]$$
$$= 2\sigma^2\mathrm{tr}\{H(A_jA_i^\dagger + A_iA_j^\dagger)H^\dagger\} = 0. \tag{17}$$

Hence, conditioned on H and $b_1, \ldots, b_K$, $\boldsymbol{L_e}(\boldsymbol{b}_i)$ and $\boldsymbol{L_e}(\boldsymbol{b}_j)$ are uncorrelated and thus independent. It follows that conditioned on H and $b_1, \ldots, b_K$, the LLR vector $(\boldsymbol{L_e}(\boldsymbol{b}_1), \ldots, \boldsymbol{L_e}(\boldsymbol{b}_K))$ has a product distribution as

$$\begin{aligned} p(\boldsymbol{L_e}(\boldsymbol{b}_1) &= L_1, \ldots, \boldsymbol{L_e}(\boldsymbol{b}_K) = L_K | H, b_1, \ldots, b_K) \\ &= \prod_{k=1}^{K} p(\boldsymbol{L_e}(\boldsymbol{b}_k) = L_k | H, b_1, \ldots, b_K) \\ &= \prod_{k=1}^{K} p(\boldsymbol{L_e}(\boldsymbol{b}_k) = L_k | H, b_k). \end{aligned} \tag{18}$$

Next, since the information bit vectors are equally probable, we have

$$\begin{aligned} p(\boldsymbol{L_e}(\boldsymbol{b}_1) &= L_1, \ldots, \boldsymbol{L_e}(\boldsymbol{b}_K) = L_K | H) \\ &= 2^{-K} \sum_{b_1} \cdots \sum_{b_K} \\ &\quad p(\boldsymbol{L_e}(\boldsymbol{b}_1) = L_1, \ldots, \boldsymbol{L_e}(\boldsymbol{b}_K) = L_K | H, b_1, \ldots, b_K) \\ &= 2^{-K} \sum_{b_1} \cdots \sum_{b_K} \prod_{k=1}^{K} p(\boldsymbol{L_e}(\boldsymbol{b}_k) = L_k | H, b_k) \\ &= 2^{-K} \prod_{k=1}^{K} \left(\sum_{b_k} p(\boldsymbol{L_e}(\boldsymbol{b}_k) = L_k | H, b_k) \right) \\ &= \prod_{k=1}^{K} p(\boldsymbol{L_e}(\boldsymbol{b}_k) = L_k | H). \end{aligned} \tag{19}$$

This concludes the proof. □

References

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

[2] E. Telatar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Labs Tech. Rep., June 1995.

[3] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *AT&T Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[4] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[5] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[6] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456–1467, July 1999.

[7] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1804–1824, July 2002.

[8] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: Performance results," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 451–460, Mar. 1999.

Unitary Subgroup Space–Time Codes Using Bruhat Decomposition and Weyl Groups

Tatsumi Konishi, *Member, IEEE*

***Abstract*—We propose a new class of unitary subgroups which is suitable to make differential unitary space–time codes for fast fading channels. These subgroups are derived from a double coset decomposition of unitary groups called Bruhat decomposition. We give some examples of the proposed unitary subgroup space–time codes for phase-shift keying (PSK), and display the performance for Rayleigh-fading channels. Finally, we describe how the new space–time codes are combined with coded modulation systems.**

***Index Terms*—Bruhat decomposition, differential modulation, fading channels, space–time coding, unitary group codes.**

I. Introduction

Differential unitary space–time modulation is useful for multiple-antenna wireless communications where neither the transmitter nor the receiver knows the fading coefficients [1]–[3]. The design criterion is the following [1]. Let M be the number of transmitter antennas and R the desired transmission rate. Construct a set $\mathcal{V}$ of $L = 2^{RM}$ unitary $M \times M$ matrices such that for any two distinct elements A and B in $\mathcal{V}$, the quantity $|\det(A - B)|$ is as large as possible. Any set $\mathcal{V}$ such that

$$|\det(A - B)| > 0$$

for all distinct $A, B \in \mathcal{V}$ is said to have full diversity.

In [4], Hassibi and Khorrami have studied whether infinite fully diverse unitary groups exist, and they have revealed that there are only two types: $U(1)$ and $SU(2)$. We are very interested in the result. Therefore, we attempt to make fully diverse unitary *subgroups* by cutting redundant parts from infinite unitary *groups*. In order to make the subgroups, we use a double coset decomposition of unitary groups that is called Bruhat decomposition.

This correspondence is organized as follows: the next section explains differential unitary space–time modulation. Section III introduces Bruhat decomposition. Section IV gives fully diverse unitary subgroups and some practical examples of space–time codes using the subgroups. In Section V, we show that the proposed method is not useful for an odd number of transmitter antennas systems, and discuss how to combine the coded modulation into the proposed space–time codes. Finally, Section VI concludes this correspondence.

II. Differential Unitary Space–Time Modulation

We consider a multiple-antenna communication system with M transmitter and N receiver antennas. Assuming that the channel is constant over the M channel uses. Then, the $M \times N$ received signal matrix X_τ is

$$X_\tau = \sqrt{\rho} S_\tau H + W_\tau \tag{1}$$

where H is the $M \times N$ matrix of Rayleigh-fading coefficients, ρ is the signal-to-noise ratio (SNR) at each receiver antenna, S_τ is the $M \times M$

Manuscript received October 30, 2002; revised June 24, 2003. The material in this correspondence was presented in part at the International Symposium on Information Theory and its Applications, Xi'an, China, October 2002.

The author is with Aichi Institute of Technology, Toyota City, Aichi 470-0392, Japan (email: konishi@in.aitech.ac.jp).

Communicated by B. Hassibi, Associate Editor for Communications.
Digital Object Identifier 10.1109/TIT.2003.817462