

# To Code or Not to Code Across Time: Space-Time Coding with Feedback

Che Lin, Vasanthan Raghavan, and Venugopal V. Veeravalli

**Abstract**—Space-time codes leverage the availability of multiple antennas to enhance the reliability of communication over wireless channels. While space-time codes have initially been designed with a focus on open-loop systems, recent technological advances have enabled the possibility of low-rate feedback from the receiver to the transmitter. The focus of this paper is on the implications of this feedback in a single-user multi-antenna system with a general model for spatial correlation. We assume a limited feedback model, that is, a coherent receiver and statistical knowledge at both the ends, along with  $B$  bits of error-free quantized channel information at the transmitter. We study space-time coding with a family of *linear dispersion (LD) codes* that meet an additional orthogonality constraint so as to ensure low-complexity decoding. Our results show that, when the number of bits of feedback ( $B$ ) is small, a space-time coding scheme that is equivalent to beamforming and *does not code across time* is optimal in a weak sense in that it maximizes the average received SNR. As  $B$  increases, this weak optimality transitions to optimality in a strong sense that is characterized by the maximization of average mutual information. Thus, from a system designer's perspective, our work suggests that beamforming may not only be attractive from a low-complexity viewpoint, but also from an information-theoretic viewpoint.

**Index Terms**—Adaptive coding, diversity methods, fading channels, feedback communications, MIMO systems, multiplexing, quantization.

## I. INTRODUCTION

THE LOW-COMPLEXITY beamforming scheme has attracted significant theoretical attention in multi-input multi-output (MIMO) applications and has been studied with perfect channel state information (CSI) at the transmitter [1], as well as with partial channel knowledge at the transmitter [2]–[5]. The broad family of *space-time codes* generalize beamforming to modulation across space and time. Initial works on space-time codes assume no CSI at the transmitter and study the reliability of information transmission with uncoded inputs, that is, with the input symbols being independent from one coherence block to another. Reliability can be improved by using certain *delay diversity techniques* [6] and these schemes can be extended to the more general framework of *space-time trellis codes* [7]. Although space-time trellis codes are near-optimal in the MIMO setting, they suffer from

decoding complexity that is exponential in the rate or the number of transmit antennas. To overcome these difficulties, *orthogonal space-time block codes* (OSTBC) [8], [9] that achieve the full diversity order<sup>1</sup> of multi-antenna systems and offer the additional advantage of low decoding complexity have been proposed.

Maximizing the diversity order, while being a useful design criterion, is only applicable for uncoded transmissions. In practice, a space-time code is used as an *inner code* in concatenation with an error correction code that is used as an *outer code*, and is designed<sup>2</sup> to achieve maximum possible rate for a given SNR. A good inner code should be such that it does not destroy any degrees of freedom that the outer code can exploit to obtain diversity. In this setting, since the outer code helps the concatenated transmitter in approaching capacity, the mutual information is a meaningful design metric for the inner space-time block codes if soft decisions are allowed at the space-time decoder.

In the no CSI case, it has been established that [11], [12] OSTBC are also optimal from a mutual information viewpoint. The assumption of no CSI at the transmitter is too pessimistic and does not capture reality, where either statistical or partial channel knowledge [13] at the transmitter may be available. In this context, when only the statistics are available at the transmitter, it is known that OSTBC are no longer optimal within the class of linear space-time codes [14], [15]. In practice, in addition to the statistical information, there is usually a viable low rate feedback link from the receiver to the transmitter. Thus recent attention, in both theory and practice, has shifted towards understanding the implications of partial CSI at the transmitter (most notably, in the form of limited or quantized feedback [13]) on the performance of communication systems. In the more practical limited feedback case, there have been some recent works [14], [16]–[20] on the design of space-time codes with CSI feedback. However, much of this body of work ignores spatial correlation and focuses on weighted OSTBC, which implicitly imposes an additional assumption on the maximum rank of the inner space-time codes. As witnessed in the statistical case [14], [15], spatial correlation can potentially lead to significant changes in code design criterion and optimal signaling. Thus, our goal in this work is to address optimal signaling with partial channel knowledge at the transmitter and more generally, whether coding across time is necessary in partial CSI systems.

<sup>1</sup>The diversity order of a code is defined as the exponent of the rate at which error probability decays with SNR.

<sup>2</sup>For example, recent work on low-density parity check codes (see, e.g., [10]) have shown that it is possible to construct outer codes that come close to achieving the mutual information between the input and the output of the inner space-time code.

Manuscript received November 2, 2007; revised April 16, 2008. This research was supported in part by the NSF award CCF 0431088 through the University of Illinois, and by a Vodafone Foundation Graduate Fellowship. This paper was presented in part at the IEEE Global Communications Conference, Washington, DC, 2007 and the Conference for Information Systems and Sciences, Princeton, NJ, 2008.

The authors are with the Coordinated Science Laboratory and the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: {clin, vvv}@uiuc.edu, vasanthan\_raghavan@ieee.org).

Digital Object Identifier 10.1109/JSAC.2008.081024.

The general framework of *linear dispersion (LD) codes*, introduced in [21], subsumes all linear space-time codes and hence provides a natural framework for studying both beamforming as well as space-time code design in a unified way, given that partial CSI is available at the transmitter. In an LD code, each symbol that is transmitted over the channel is some linear combination of the inputs and their complex conjugates, and the codes are designed to maximize the mutual information between the input and the output of the space-time code. While the generality of the LD framework leads to some complications in code design, recent work [22]–[24] shows that systematic LD code constructions are still possible.

In this work, we impose an additional *Generalized Orthogonal Constraint (GOC)* [9], [15] on the LD codes so that they enjoy the same low-complexity of decoding<sup>3</sup> as OSTBC. That is, we consider the set of *orthogonal LD codes*. The search for the optimal orthogonal LD codes provides significant insights on whether coding across time is necessary or not.

#### Contributions:

- We first show that when there is perfect CSI at the transmitter, the optimal power allocation across the different input symbols is uniform. Furthermore, the rank of the optimal LD code is one and since rank-one LD codes are equivalent to beamforming, the *optimal perfect CSI scheme does not code across time*.
- When only statistical information is available at the transmitter, we establish that uniform symbol power allocation is still optimal. It is also known from [15] that the rank of the optimal linear space-time code is in general dependent on SNR and channel correlation. Thus the *optimal statistical scheme codes across time, in general*.
- In the partial CSI case, we first show the optimality of uniform symbol power allocation, irrespective of the level of channel knowledge at the transmitter. On the question of the rank of the optimal scheme, it is natural to expect a smooth transition as the quality of channel information at the transmitter gets successively refined (that is, as the number of bits of feedback  $B$  increases). Surprisingly, we show that there exists a feedback threshold, above which rank-one schemes enjoy strong optimality properties. That is, when  $B$  is sufficiently large (for example,  $B \gg \log(N_t)$  with  $N_t$  denoting the transmit antenna dimension), we show that rank-one schemes maximize the average mutual information. In contrast, when  $B$  is small, we show a slightly weaker result: Rank-one schemes maximize the average received SNR. While the problem of transition from weak to strong optimality (as a function of  $B$ ) is not addressed here, our study suggests that for most scenarios of interest, the *optimal scheme under the orthogonal LD code framework and quantized feedback corresponds to not coding across time*.
- The optimality of rank-one schemes (beamforming) implies that the low-complexity advantage of scalar coding is justified from an information theoretic sense.

**Notations:** We use  $\mathbf{X}(i, j)$  and  $\mathbf{X}(i)$  to denote the  $i, j$ -th and

$i$ -th diagonal entries of a matrix  $\mathbf{X}$ . The conjugate transpose and regular transpose are denoted by  $(\cdot)^\dagger$  and  $(\cdot)^T$  while  $\mathbf{E}[\cdot]$  and  $\text{Tr}(\cdot)$  stand for the expectation and the trace operators, respectively. We say that a singular value decomposition of a matrix is in its *standard ordering* if the singular values are arranged in non-increasing order. Further, if the matrix is Hermitian,  $\lambda_{\max}(\cdot)$  denotes the largest eigenvalue.

## II. SYSTEM SETUP

We consider a single-user MIMO communication system with  $N_t$  transmit and  $N_r$  receive antennas. The multi-antenna channel matrix experiences fading in time, frequency, and space. In this paper, we assume a narrowband, block fading model for the channel. That is, the channel is frequency flat and remains constant across a block of length  $N_c$  symbols and fades ergodically from block to block. With these simple models for the evolution of the channel across time and frequency, the main focus is on the spatial aspect.

To overcome the impediments of fading, we will consider the design of space-time codes and view the channel across the block length  $N_c$  as corresponding to one *channel use*. The discrete-time, complex baseband model under this setting is given by

$$\mathbf{Y} = \sqrt{\frac{\rho}{N_t}} \mathbf{H} \mathbf{X} + \mathbf{W} \quad (1)$$

where  $\mathbf{X} \in \mathbb{C}^{N_t \times N_c}$  is the transmitted signal matrix,  $\mathbf{Y} \in \mathbb{C}^{N_r \times N_c}$  is the received signal matrix,  $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$  corresponds to the channel matrix, and  $\mathbf{W} \in \mathbb{C}^{N_r \times N_c}$  denotes the complex additive white Gaussian noise with i.i.d. entries,  $\mathbf{W}(i, j) \sim \mathcal{CN}(0, 1)$ . We assume an average power constraint on  $\mathbf{X}$  given by  $\mathbf{E}[\text{Tr}(\mathbf{X}\mathbf{X}^\dagger)] \leq N_t N_c$ , which results in a transmit power constraint  $\rho$  over each symbol duration.

### A. Spatial Correlation

We now describe the spatial fading framework used in this work. It has been well-documented that the assumption of zero-mean Rayleigh fading is an accurate model for  $\mathbf{H}$  in a non line-of-sight setting. Thus the complete channel statistics are described by the second-order moments. Rich scattering environments are accurately modeled by the commonly used i.i.d. model where the channel entries are i.i.d.  $\mathcal{CN}(0, 1)$ . However, the i.i.d. model is not accurate in describing realistic propagation environments. Various statistical models have been proposed to overcome the deficiencies associated with the i.i.d. model.

The most general, mathematically tractable spatial correlation model is based on a decomposition of the channel onto its canonical coordinates: the eigen-bases of the transmit and the receive covariance matrices [25]–[27]. The canonical model assumes that the auto- and the cross-covariance matrices of *all* the columns of  $\mathbf{H}$  have a common eigen-basis (this common eigen-matrix is denoted by  $\mathbf{U}_r$ ). Further, it also assumes that the eigen-matrices of the auto- and the cross-covariance matrices of *all* the rows of  $\mathbf{H}$  are identical (this common eigen-matrix is denoted by  $\mathbf{U}_t$ ). The above redundancies are exploited to decompose  $\mathbf{H}$  as

$$\mathbf{H} = \mathbf{U}_r \mathbf{H}_{\text{ind}} \mathbf{U}_t^\dagger \quad (2)$$

<sup>3</sup>Satisfaction of the GOC ensures that the joint maximum-likelihood (ML) decoding of the vector input reduces to individual ML decoding of the scalar inputs.

where  $\mathbf{H}_{\text{ind}}$  has independent, but not necessarily identically distributed entries. The columns of  $\mathbf{U}_r$  and  $\mathbf{U}_t$  also serve as eigenvectors for the receive and the transmit covariance matrices, which are defined as  $\mathbf{\Sigma}_r = \mathbf{E}[\mathbf{H}\mathbf{H}^\dagger]$  and  $\mathbf{\Sigma}_t = \mathbf{E}[\mathbf{H}^\dagger\mathbf{H}]$ , respectively. It can be checked that [25] the model in (2) reduces to some well-known models such as the separable correlation model or the virtual representation [28]–[30].

### B. Signaling Scheme - Linear Dispersion Codes

As mentioned in the introduction, the coding problem for the MIMO channel  $\mathbf{H}$  can be separated into the design of an inner space-time block code and an outer code. Accordingly, input data  $\mathbf{x}[t]$  is demultiplexed into  $K$  data-streams denoted by  $\mathbf{x}_1[t], \dots, \mathbf{x}_K[t]$  for the space-time encoder at a given symbol time  $t$ . We make the following simplifying assumptions on the input symbols in this work.

#### Assumption 1:

- The data-streams corresponding to  $\mathbf{x}_k[t]$  are i.i.d. zero mean across time for all  $k$  and they are drawn from some *real* constellation with marginal distribution  $p(x_k)$ .
- For any  $t$  and all  $i, j$  such that  $i \neq j$ ,  $\mathbf{x}_i[t]$  and  $\mathbf{x}_j[t]$  are independent.

The second assumption can be justified if  $\mathbf{x}_1[t], \dots, \mathbf{x}_K[t]$  are produced as outputs of independent scalar outer encoders as in the V-BLAST signaling scheme. Applications involving the use of bit-interleaved codes at the outer encoder also justify the second assumption. Furthermore, both assumptions can be justified if the data coming from the encoder is fed through a random interleaver, a very practical assumption. Since  $\mathbf{x}_k[t]$  are i.i.d. across time, we will drop the time index  $t$  in the ensuing discussion.

While arbitrarily structured space-time coding schemes can be considered for signaling, in this work, we will focus on a specific LD code-based signaling [21]. The definition of an LD code involves a set of dispersion matrices  $\{\mathbf{A}_k\}_{k=1}^K \in \mathbb{C}^{N_t \times N_c}$ , such that the space-time code  $\mathbf{X}$  is

$$\mathbf{X} = \sum_{k=1}^K \mathbf{A}_k \mathbf{x}_k \quad (3)$$

where the symbols  $\{\mathbf{x}_k\}_{k=1}^K$  satisfy Assumption 1. That is, at a given symbol time, the outer encoder produces a set of independent symbols  $\{\mathbf{x}_k\}$ , which are then spread across the spatial and temporal dimensions through  $\{\mathbf{A}_k\}$ .

It is important to note that LD codes encompass all possible linear space-time codes. In addition, we assume that the class of LD codes satisfy the *Generalized Orthogonal Constraint (GOC)*, that is,  $\mathbf{A}_k \mathbf{A}_j^\dagger + \mathbf{A}_j \mathbf{A}_k^\dagger = \mathbf{0}$  for all  $k, j, k \neq j$ . It has been shown in [9], [11] that the GOC is equivalent to the condition  $p(\mathbf{x}_1, \dots, \mathbf{x}_K | \mathbf{Y}, \mathbf{H}_{\text{ind}}) = \prod_{k=1}^K p(\mathbf{x}_k | \mathbf{Y}, \mathbf{H}_{\text{ind}})$ . That is, the likelihood function factors and the complexity of the LD decoder is greatly reduced since the joint ML decoding reduces to individual ML decoding for each symbol. In other words, the channel decouples into  $K$  parallel sub-channels. It is important to note that the decoding complexity for this class of LD codes, labeled henceforth as *orthogonal LD codes*, is the same as that necessary for OSTBC.

After normalizing  $\mathbf{x}_k$  such that  $\mathbf{E}[\mathbf{x}_k^2] = 1$ , the power constraint is applied to  $\mathbf{A}_k$ , resulting in  $\sum_{k=1}^K \text{Tr}(\mathbf{A}_k \mathbf{A}_k^\dagger) \leq$

$N_t N_c$ . The power allocated to the  $k$ -th symbol is  $\text{Tr}(\mathbf{A}_k \mathbf{A}_k^\dagger)$ . We say that the symbol power allocation is uniform if  $\text{Tr}(\mathbf{A}_k \mathbf{A}_k^\dagger)$  is the same for all  $k$ .

## III. OPTIMAL SIGNALING SCHEMES

In this section, we study the problem of optimal LD code construction under different assumptions on the available CSI at the transmitter. There are four relevant cases of CSI at the transmitter: 1) An extreme case of no CSI at the transmitter, 2) The other extreme of perfect CSI, 3) A setting where only statistical information is available, and 4) The case where partial CSI in the form of limited feedback is available at the transmitter. The first three cases are the subject of this section; the last case is addressed in more detail in Sec. IV.

### A. No CSI at the Transmitter

When no channel information is available at the transmitter, it is well-known that the optimal scheme is to assume that the channel is i.i.d. Thus, any space-time code tailored to the i.i.d. case can be used. In particular, Hassibi and Hochwald [21] have applied the mutual information criterion to design optimal codes (within the class of LD codes) with i.i.d. Gaussian inputs. Jiang [11] has studied the design of optimal LD codes for i.i.d. channels with binary inputs and conjectured that the optimal code is the generalized orthogonal design introduced in [9]. Bresler and Hajek [12] proved the above conjecture and extended the work to arbitrary real inputs. The following sections demonstrate how channel information can help improve performance.

### B. Statistical Information at the Transmitter

We now study the structure of the optimal LD codes when only statistical information is available at the transmitter. We build on the recent work in [15] where optimal LD codes are constructed by maximizing the average mutual information between the input and the output of the inner code. If channel correlation is modeled with the canonical framework as in (2), we obtain the following equivalent channel model:

$$\tilde{\mathbf{Y}} = \sqrt{\frac{\rho}{N_t}} \sum_{k=1}^K \mathbf{H}_{\text{ind}} \tilde{\mathbf{A}}_k \mathbf{x}_k + \tilde{\mathbf{W}} \quad (4)$$

where  $\tilde{\mathbf{A}}_k = \mathbf{U}_t^\dagger \mathbf{A}_k$ ,  $\tilde{\mathbf{Y}} = \mathbf{U}_r^\dagger \mathbf{Y}$ , and  $\tilde{\mathbf{W}} = \mathbf{U}_r^\dagger \mathbf{W}$ . The GOC is equivalent to  $\tilde{\mathbf{A}}_k \tilde{\mathbf{A}}_j^\dagger + \tilde{\mathbf{A}}_j \tilde{\mathbf{A}}_k^\dagger = \mathbf{0}$  for all  $k, j, k \neq j$  and the original power constraint is equivalent to  $\sum_{k=1}^K \text{Tr}(\tilde{\mathbf{A}}_k \tilde{\mathbf{A}}_k^\dagger) \leq N_t N_c$ . We have the following theorem characterizing the structure of optimal LD codes.

*Theorem 1:* Let  $\tilde{\mathbf{X}} = \mathbf{U}_t^\dagger \mathbf{X}$  be an LD code as in (3) with  $K$  symbols and let the corresponding dispersion matrices be  $\{\tilde{\mathbf{A}}_k, k = 1, \dots, K\}$ . Also, let the input symbols  $\mathbf{x}_1, \dots, \mathbf{x}_K$  satisfy Assumption 1 and the dispersion matrices satisfy the GOC. Suppose there exists an LD code satisfying the *spatial power allocation*:  $\tilde{\mathbf{A}}_k \tilde{\mathbf{A}}_k^\dagger = \mathbf{\Lambda}_{\text{stat}}$  for all  $k$ , then such a code maximizes the average mutual information  $\mathbf{E}[I(\mathbf{X}; \mathbf{Y} | \mathbf{H} = \mathbf{H})]$  and achieves the ergodic capacity. Here,

$\mathbf{\Lambda}_{\text{stat}}$  is a positive semidefinite diagonal matrix with

$$\mathbf{\Lambda}_{\text{stat}} = \arg \max_{\mathbf{\Lambda}} \mathbf{E} \left[ \varphi \left( \frac{\rho}{N_t} \text{Tr} \left( \mathbf{H}_{\text{ind}} \mathbf{\Lambda} \mathbf{H}_{\text{ind}}^\dagger \right) \right) \right] \quad \text{s.t. } \text{Tr}(\mathbf{\Lambda}) = \frac{N_t N_c}{K}, \quad (5)$$

where  $\varphi(a) = h(\sqrt{a}\mathbf{x} + \mathbf{n} | \mathbf{H} = \mathbf{H})$ ,  $h(\cdot)$  denotes the differential entropy, and  $\mathbf{n}$  is a real zero-mean Gaussian of variance 1/2.

*Proof:* See Appendix A.  $\blacksquare$

Theorem 1 states that uniform symbol power allocation across the data-streams is optimal from an average mutual information viewpoint, i.e.,  $\tilde{\mathbf{A}}_k \tilde{\mathbf{A}}_k^\dagger$  is the same for all  $k$ . The optimal spatial power allocation is given by  $\mathbf{\Lambda}_{\text{stat}}$ , and in general,  $\mathbf{\Lambda}_{\text{stat}}$  excites the multiple modes non-uniformly. We now elaborate on the structure of  $\{\tilde{\mathbf{A}}_k\}$ . Given that  $r$  denotes the number of spatial modes excited by the optimal statistical scheme, it can be seen that

$$\tilde{\mathbf{A}}_k = \begin{cases} [\sqrt{\mathbf{\Lambda}_{\text{stat}}} \mathbf{0}_{N_t \times N_c - N_t}] \mathbf{Y}_k^\dagger & \text{if } N_c \geq N_t \\ [\sqrt{\mathbf{\Lambda}_{\text{stat}}}]_{\text{prin}} \mathbf{Y}_k^\dagger & \text{if } r \leq N_c < N_t \end{cases} \quad (6)$$

where  $\tilde{\mathbf{Y}}_k$  is an arbitrary  $N_c \times N_c$  unitary matrix and  $[\sqrt{\mathbf{\Lambda}_{\text{stat}}}]_{\text{prin}}$  is the  $N_t \times N_c$  principal sub-matrix of  $\sqrt{\mathbf{\Lambda}_{\text{stat}}}$ . With the above structure for  $\tilde{\mathbf{A}}_k$  and with  $\tilde{\mathbf{Y}}_k$  denoting the  $N_c \times r$  principal sub-matrix of  $\tilde{\mathbf{Y}}_k$ , we need

$$\tilde{\mathbf{Y}}_k^\dagger \tilde{\mathbf{Y}}_j + \tilde{\mathbf{Y}}_j^\dagger \tilde{\mathbf{Y}}_k = \mathbf{0} \quad \text{for all } k \neq j \quad (7)$$

to meet the GOC. If  $rK \leq N_c$ , (7) can be met by letting  $\{\tilde{\mathbf{Y}}_k\}$  to be a set of  $r$  distinct columns of a random  $N_c \times N_c$  unitary matrix. In fact, this choice leads to a stronger condition where  $\tilde{\mathbf{Y}}_k^\dagger \tilde{\mathbf{Y}}_j = \mathbf{0}$  for any  $k \neq j$ . Initial studies suggest that  $rK \leq 2N_c$  is both necessary and sufficient for a feasible construction that meets (7). These results and their connections to constructions via generalized orthogonal designs [9] will be reported elsewhere.

### C. Perfect CSI at the Transmitter

Recall that the system equation can be written as

$$\mathbf{Y} = \sqrt{\frac{\rho}{N_t}} \sum_{k=1}^K \mathbf{H} \mathbf{A}_k \mathbf{x}_k + \mathbf{W} \quad (8)$$

with power constraint  $\sum_{k=1}^K \text{Tr}(\mathbf{A}_k \mathbf{A}_k^\dagger) \leq N_t N_c$ . Following [11], [15], it can be shown that we have the following upper bound for the mutual information when perfect CSI is available at both the ends:

$$\begin{aligned} I(\mathbf{X}; \mathbf{Y} | \mathbf{H} = \mathbf{H}) &= I(\mathbf{x}_1, \dots, \mathbf{x}_K; \mathbf{Y} | \mathbf{H} = \mathbf{H}) \\ &\leq \sum_{k=1}^K \underbrace{I\left(\mathbf{x}_k; \sqrt{\frac{\rho}{N_t}} \mathbf{H} \mathbf{A}_k \mathbf{x}_k + \mathbf{W} \mid \mathbf{H} = \mathbf{H}\right)}_{I_k}. \end{aligned}$$

Equality in (9) holds if and only if the GOC is satisfied. From [11], [15], we also have

$$I_k = \varphi \left( \frac{\rho}{N_t} \text{Tr}(\mathbf{H} \mathbf{Q}_k \mathbf{H}^\dagger) \right) - h(\mathbf{n}) \quad (9)$$

where  $\mathbf{Q}_k = \mathbf{A}_k \mathbf{A}_k^\dagger$ . The structure of the optimal LD code is as follows.

*Theorem 2:* Let  $\mathbf{X}$  be an LD code as in (3) with  $K$  symbols and let the corresponding dispersion matrices be  $\{\mathbf{A}_k\}_{k=1}^K$ . Also, let the input symbols  $\mathbf{x}_1, \dots, \mathbf{x}_K$  satisfy Assumption 1. The instantaneous mutual information can be upper bounded as

$$I(\mathbf{X}; \mathbf{Y} | \mathbf{H} = \mathbf{H}) \leq K \left[ \varphi \left( \frac{\rho N_c}{K} \lambda_{\max}(\mathbf{H}^\dagger \mathbf{H}) \right) - h(\mathbf{n}) \right] \quad (10)$$

with equality if and only if  $\{\mathbf{A}_k\}$  satisfy the GOC and  $\mathbf{A}_k \mathbf{A}_k^\dagger = \mathbf{Q}_k = \mathbf{Q}$  for all  $k$  where  $\mathbf{Q} = \mathbf{U}_H \mathbf{\Lambda}_H \mathbf{U}_H^\dagger$ . The matrix  $\mathbf{U}_H$  is an eigenvector matrix of  $\mathbf{H}^\dagger \mathbf{H}$  (in the standard order) and the only non-zero entry in  $\mathbf{\Lambda}_H$  is the leading diagonal element whose value is  $\frac{N_t N_c}{K}$ .

*Proof:* See Appendix B.  $\blacksquare$

The above result shows that the optimal choice of  $\{\mathbf{Q}_k\}$  is independent of  $k$ , that is, uniform symbol power allocation is *still* optimal. Furthermore, this scheme excites only the dominant spatial mode. A generic singular value decomposition for  $\mathbf{A}_k$  shows that it has to satisfy  $\mathbf{A}_k = \sqrt{\frac{N_t N_c}{K}} \mathbf{U}_H[1] \mathbf{v}_k$ , where  $\mathbf{U}_H[1]$  is  $N_t \times 1$  and is the first column of  $\mathbf{U}_H$ , and  $\mathbf{v}_k$  is an  $1 \times N_c$  vector of unit norm. With this structure, it can also be checked that the GOC can be met if and only if the  $K \times N_c$  matrix  $\mathbf{V}$  defined as

$$\mathbf{V} = [\mathbf{v}_1^T \quad \mathbf{v}_2^T \quad \dots \quad \mathbf{v}_K^T]^T \quad (11)$$

satisfies  $\mathbf{V} \mathbf{V}^\dagger = \mathbf{I}_K + i\mathbf{X}$  where  $\mathbf{X}$  is real skew-symmetric<sup>4</sup>. Based on this decomposition, we can completely characterize the structure of the dispersion matrices  $\{\mathbf{A}_k\}$ , and can establish the existence of an LD code if and only if  $K \leq 2N_c$ . The following proposition states this result.

*Proposition 1:* There exists a  $K \times N_c$  matrix  $\mathbf{V}$  such that  $\mathbf{V} \mathbf{V}^\dagger = \mathbf{I}_K + i\mathbf{X}$  where  $\mathbf{X}$  is real skew-symmetric if and only if  $K \leq 2N_c$ .  $\blacksquare$

Due to space constraints, the proof of the claim and the explicit construction of the dispersion matrices are not reported here. However, we provide simple illustrations of these constructions now. When  $K \leq N_c$ , any set of  $K$  rows of an arbitrary  $N_c \times N_c$  unitary matrix works for  $\mathbf{V}$ . Since  $\{\mathbf{v}_k\}$  are pairwise orthogonal, the GOC is naturally met. In fact, it is to be noted that a stronger condition (than the GOC) holds:  $\mathbf{A}_k \mathbf{A}_j^\dagger = \mathbf{0}$  for all  $k \neq j$  in this case. Under the same conditions as above, further explicit constructions to meet the GOC are proposed in [9]. However, these conditions are only sufficient, but not necessary as the statement of Prop. 1 illustrates.

The surprising<sup>5</sup> claim of Prop. 1 is that the GOC can be met and the data-streams separated temporally, as long as  $K \leq 2N_c$ . The two-fold gain in the maximum possible choice of  $K$  (which is  $2N_c$ ) over the ‘naturally’ expected limit of  $K = N_c$  stems primarily from the weaker condition that  $2\text{Re}(\mathbf{A}_k \mathbf{A}_j^\dagger) = \mathbf{A}_k \mathbf{A}_j^\dagger + \mathbf{A}_j \mathbf{A}_k^\dagger = \mathbf{0}$  for all  $k \neq j$ , instead of the more stringent condition that  $\mathbf{A}_k \mathbf{A}_j^\dagger = \mathbf{0}$  for all  $k \neq j$ .

<sup>4</sup>An  $n \times n$  matrix  $\mathbf{X}$  is said to be real skew-symmetric if it has real entries and satisfies  $\mathbf{X}^T = -\mathbf{X}$ .

<sup>5</sup>In retrospect, this claim is not all that surprising since the use of real input symbols means that  $R$  bits can be transmitted per complex-dimension if a  $2^R$ -ary constellation is used for signaling.

For example, when  $K = 2N_c$ , the condition in Prop. 1 can be satisfied by choosing

$$\mathbf{v}_k = \begin{cases} \begin{bmatrix} 0 & \cdots & 0 & \underbrace{1}_{k_1} & 0 & \cdots & 0 \end{bmatrix} & \text{if } k \in \mathcal{A} \\ \begin{bmatrix} 0 & \cdots & 0 & \underbrace{j}_{k_2} & 0 & \cdots & 0 \end{bmatrix} & \text{if } k \in \mathcal{B} \end{cases} \quad (12)$$

The set  $\mathcal{A}$  corresponds to  $k$  is odd,  $k_1 = \frac{k+1}{2}$  and the set  $\mathcal{B}$  corresponds to  $k$  is even,  $k_2 = \frac{k}{2}$ .

We are now prepared to state the main result of this section.

*Theorem 3:* Consider the family of orthogonal LD codes whose decoding complexity is comparable with OSTBC. The mutual information achievable with such codes is a non-decreasing function of  $K$  which implies that  $K = 2N_c$  is necessary for optimal signaling. Furthermore, the optimal signaling scheme reduces to beamforming along  $\mathbf{U}_H[1]$ , the dominant right singular vector of  $\mathbf{H}$ . More simply stated, not coding across time is optimal from an information theoretic perspective.

*Proof:* First, note that the GOC has to be met for the case of orthogonal LD codes. From Prop. 1, we see that  $K = 2N_c$  is the largest value of  $K$  such that this is possible. With  $K = 2N_c$  and  $\mathbf{v}_k$  as in (12), the average mutual information can be expressed as

$$I(\mathbf{x}_1, \dots, \mathbf{x}_K; \mathbf{Y} | \mathbf{H} = \mathbf{H}) = K\varphi\left(\frac{Z}{K}\right) - Kh(\mathbf{n}) \quad (13)$$

where  $Z = \rho N_c \lambda_{\max}(\mathbf{H}^\dagger \mathbf{H})$ . Letting  $K$  to be a continuous parameter in the previous expression, we observe that the derivative of the mutual information with respect to  $K$  is positive. For this, note that for any  $Z \geq 0$ , we have

$$\varphi\left(\frac{Z}{K}\right) - h(\mathbf{n}) = \int_0^{\frac{Z}{K}} \varphi'(y) dy \geq \frac{Z}{K} \varphi'\left(\frac{Z}{K}\right) \quad (14)$$

since  $\varphi(\cdot)$  is a differentiable function with  $\varphi(0) - h(\mathbf{n}) = 0$  and  $\frac{d\varphi(a)}{da} = \frac{1}{2} \text{mse}(a)$ , and hence monotonically decreasing in  $a$ ; see App. B for details. In this setting, we have

$$\mathbf{Y} = \sqrt{\frac{\rho N_c}{K}} \mathbf{H} \mathbf{U}_H[1] \mathbf{x} + \mathbf{W} \quad (15)$$

$$= \mathbf{H} \mathbf{U}_H[1] \mathbf{x}_{\text{trans}} + \mathbf{W}, \mathbf{x}_{\text{trans}} = \sqrt{\frac{\rho}{2}} \mathbf{x}, \quad (16)$$

$$\mathbf{x} = [\mathbf{x}_1 + i\mathbf{x}_2, \mathbf{x}_3 + i\mathbf{x}_4, \dots, \mathbf{x}_{2N_c-1} + i\mathbf{x}_{2N_c}]$$

for the system equation. In other words, the optimal signaling scheme reduces to *beamforming the complex symbol*  $(\mathbf{x}_{2k-1} + i\mathbf{x}_{2k})/\sqrt{2}$  along the fixed transmit direction  $\mathbf{U}_H[1]$  in the  $k$ -th symbol period of the coherence block with a transmit energy of  $\rho$ . This completes our proof. ■

#### IV. QUANTIZED CSI AT THE TRANSMITTER

We now study the quantized CSI case where an error-free, negligible-delay link, that allows  $B$  bits of channel information to the transmitter, is available. In addition, we assume that the statistics of the channel are known perfectly at the transmitter. From the previous section, we see that rank-one signaling (beamforming) is optimal in the perfect CSI case, while in the statistical case, the rank of the optimal scheme could be greater than one, in general. It is natural to expect a smooth transition

in the rank as the quality of CSI gets refined with increasing  $B$ . In this section, we show that a rank-one scheme satisfies certain strong optimality properties. This observation is based on the following two results: 1) When  $B$  is sufficiently large (to be characterized more precisely soon), a rank-one scheme maximizes the average mutual information, 2) In the small  $B$  regime, we can show that a rank-one scheme maximizes the average received SNR.

We first make precise the notion of a  $B$ -bit limited feedback scheme in the context of LD codes. We assume the knowledge of a codebook (of  $2^B$  codewords) at the transmitter and the receiver, where each codeword is a set of  $K$  dispersion matrices satisfying a total power constraint. That is, the codebook  $\mathcal{C}$  is

$$\mathcal{C} = \left\{ \mathcal{C}^\ell = (\mathbf{A}_1^\ell, \dots, \mathbf{A}_K^\ell) : \sum_{k=1}^K \text{Tr}(\mathbf{A}_k^\ell \mathbf{A}_k^{\ell\dagger}) \leq N_t N_c, \right. \\ \left. \ell = 1, \dots, 2^B. \right. \quad (17)$$

If the  $\ell$ -th codeword is used in signaling, the system model is described by

$$\mathbf{Y} = \sqrt{\frac{\rho}{N_t}} \sum_{k=1}^K \mathbf{H} \mathbf{A}_k^\ell \mathbf{x}_k + \mathbf{W}. \quad (18)$$

Recall from the previous section that

$$I(\mathbf{x}_1, \dots, \mathbf{x}_K; \mathbf{Y} | \mathbf{H} = \mathbf{H}) \\ \leq \sum_{k=1}^K I\left(\mathbf{x}_k; \sqrt{\frac{\rho}{N_t}} \mathbf{H} \mathbf{A}_k^\ell \mathbf{x}_k + \mathbf{W} | \mathbf{H} = \mathbf{H}\right) \\ = \sum_{k=1}^K \varphi\left(\frac{\rho}{N_t} \text{Tr}(\mathbf{H} \mathbf{Q}_k^\ell \mathbf{H}^\dagger)\right) - Kh(\mathbf{n})$$

where the upper bound is met if  $\{\mathbf{A}_k^\ell\}_{k=1}^K$  satisfies the GOC,  $\mathbf{Q}_k^\ell = \mathbf{A}_k^\ell \mathbf{A}_k^{\ell\dagger}$ , and  $h(\mathbf{n})$  is defined in Theorem 1. Thus, the mutual information is completely characterized by the set of covariance matrices  $(\mathbf{Q}_1^\ell, \dots, \mathbf{Q}_K^\ell)$ . Over each coherence block, the receiver feeds back  $\ell^*$ , the index of the optimal codeword that maximizes the instantaneous mutual information to the transmitter. The transmitter communicates over the remaining symbols in the coherence block according to (3) with dispersion matrices  $\{\mathbf{A}_k^{\ell^*}\}$ . We now show that uniform symbol power allocation is optimal even in the partial CSI case.

*Proposition 2:* Let the  $B$ -bit quantized feedback system be described as in (18). For any choice of  $B$ , the average mutual information is maximized by a codebook that allocates uniform power to input symbols. In fact, for any codeword index  $\ell$ , we have  $\mathbf{Q}_k^\ell = \mathbf{Q}^\ell$  for all  $k$ .

*Proof:* See Appendix C. ■

Thus, from the above theorem, we only need to quantize  $\mathbf{Q}^\ell$ . The most natural quantization for a covariance matrix is based on an eigen-decomposition of  $\mathbf{Q}^\ell$ . For this, we let  $N_1$  and  $N_2$  be such that  $N_1 N_2 = 2^B$ , and quantize  $\mathbf{Q}^\ell$  as

$$\mathbf{Q}^\ell = \mathbf{Q}^{(i-1)N_2+j} = \mathbf{U}_i \mathbf{\Lambda}_j \mathbf{U}_i^\dagger \triangleq \mathbf{Q}^{i,j}, \\ i = 1, \dots, N_1 \text{ and } j = 1, \dots, N_2. \quad (19)$$

Here,  $\{\mathbf{U}_i\}_{i=1}^{N_1}$  are unitary and  $\{\mathbf{\Lambda}_j\}_{j=1}^{N_2}$  are positive semi-definite diagonal with  $\text{Tr}(\mathbf{\Lambda}_j) \leq N_t N_c / K$  for all  $j$ . While the

above quantization seems natural, it is unclear how to allocate  $B$  into  $N_1$  and  $N_2$  *optimally*. In the ensuing discussion, we consider two cases and study the optimality of rank-one codebooks.

#### A. Strong Optimality of Rank-One Codebooks when $N_2 \geq N_t$

Since the rank of  $\mathbf{Q}^{i,j}$  is that of  $\mathbf{\Lambda}_j$ , we fix  $\{\mathbf{U}_i\}_{i=1, \dots, N_1}$  to be a known family and optimize over all possible  $\{\mathbf{\Lambda}_j\}$  to find the rank of the optimal codebook. This results in the following optimization:

$$\mathbf{\Lambda}_{j^*} = \max_{\{\mathbf{\Lambda}_j : \text{Tr}(\mathbf{\Lambda}_j) \leq \frac{N_t N_c}{K}\}} \mathbf{E} \left[ \max_{i,j} \varphi \left( \frac{\rho}{N_t} \text{Tr}(\mathbf{H} \mathbf{U}_i \mathbf{\Lambda}_j \mathbf{U}_i^\dagger \mathbf{H}^\dagger) \right) \right].$$

We say that rank-one codebooks are *strongly optimal* if a codebook of rank-one power allocations is sufficient to maximize the average mutual information for any choice of  $\{\mathbf{U}_i\}$ .

*Theorem 4:* If  $N_2 \geq N_t$ , a rank-one codebook is *strongly optimal*.

*Proof:* For each realization of the channel  $\mathbf{H} = \mathbf{H}$ , we seek to maximize the instantaneous mutual information by choosing the optimal codeword at the receiver and feed back the index  $(i^*, j^*)$  to the transmitter through the feedback link. For this, note that for any fixed  $\{\mathbf{U}_i\}$ ,

$$\begin{aligned} \max_{i,j} \varphi \left( \frac{\rho}{N_t} \text{Tr}(\mathbf{H} \mathbf{Q}^{i,j} \mathbf{H}^\dagger) \right) &\stackrel{(a)}{=} \max_{i,j} \varphi \left( \frac{\rho}{N_t} \text{Tr}(\mathbf{S}_i \mathbf{\Lambda}_j \mathbf{S}_i^\dagger) \right) \\ &\stackrel{(b)}{=} \max_{i,j} \varphi \left( \frac{\rho}{N_t} \sum_{m=1}^{N_t} \mathbf{\Lambda}_j(m) \|s_{im}\|^2 \right) \\ &\stackrel{(c)}{=} \max_{i,j} \varphi \left( \frac{\rho N_c}{K} \sum_{m=1}^{N_t} \alpha_{jm} s_{im} \right) \stackrel{(d)}{\leq} \varphi \left( \frac{\rho N_c}{K} s_{i^* m^*} \right) \end{aligned}$$

where (a) follows by defining  $\mathbf{S}_i = \mathbf{\Lambda}_H^{1/2} \mathbf{U}_H^\dagger \mathbf{U}_i$  and  $\mathbf{H}^\dagger \mathbf{H} = \mathbf{U}_H \mathbf{\Lambda}_H \mathbf{U}_H^\dagger$ , (b) and (c) follow by denoting the  $m$ -th column of  $\mathbf{S}_i$  by  $s_{im}$ , its norm by  $\|s_{im}\|$ ,  $\alpha_{jm} = \frac{\mathbf{\Lambda}_j(m) K}{N_t N_c}$  and  $\sum_m \alpha_{jm} \leq 1$ , and (d) follows by letting  $(i^*, m^*) = \arg \max_{1 \leq i \leq N_1, 1 \leq m \leq N_t} s_{im}$ . If  $N_2 \geq N_t$ , we can consider a distinct set of  $N_t$  rank-one power allocations each of which excites only one mode. Using this set in the above framework allows us to meet the upper bound. ■

Note that the condition  $N_2 \geq N_t$  implies that  $2^B = N_1 N_2 \geq N_t$ . But this inequality does not impose any constraint on  $N_1$ . Nevertheless, we can say that if  $B$  is sufficiently large ( $B \gg \log(N_t)$ ) so that at least  $\log(N_t)$  bits can be allocated to quantize the power allocation component of  $\mathbf{Q}^\ell$ , rank-one codebooks are always optimal irrespective of the constellation of input symbols, SNR, and channel correlation. This optimality is not completely surprising since the quantized feedback system closely approximates a perfect feedback system when  $B \gg \log(N_t)$ . That is, allocating more than  $\log(N_t)$  bits into quantizing  $\{\mathbf{\Lambda}_j\}$  is clearly sub-optimal. Other than this fact, the optimal allocation of  $B$  bits into  $N_1$  and  $N_2$  is still unclear.

#### B. Weak Optimality of Rank-One Codebooks when $N_2 < N_t$

From the notation of Theorem 4, when  $N_2 < N_t$  we can rewrite the optimization defining  $\mathbf{\Lambda}_{j^*}$  as

$$\max_{\{\alpha_{jm} : \sum_{m=1}^{N_t} \alpha_{jm} \leq 1 \text{ for all } j\}} \mathbf{E} \left[ \max_{i,j} \varphi \left( \frac{\rho N_c}{K} \sum_{m=1}^{N_t} \alpha_{jm} s_{im} \right) \right]. \quad (20)$$

Direct optimization of (20) requires the exact distribution function of  $s_{im}$ , which is a complicated function of the spatial correlation, thus rendering the above problem intractable. We now consider an alternate formulation of the above problem wherein the objective function is the minimization of  $\Delta I$  with

$$\Delta I = \mathbf{E} \left[ \varphi \left( \frac{\rho N_c}{K} \lambda_{\max}(\mathbf{H}^\dagger \mathbf{H}) \right) - \max_{i,j} \varphi \left( \frac{\rho N_c}{K} \sum_{m=1}^{N_t} \alpha_{jm} s_{im} \right) \right]. \quad (21)$$

That is, the objective is to minimize the difference in average mutual information between the perfect CSI benchmark and a quantized feedback scheme. We now propose an upper bound for  $\Delta I$  that makes the study of optimal signaling tractable in a weak sense.

*Lemma 1:* The quantity  $\Delta I$  can be upper bounded by  $\Delta \text{SNR}$ , the difference in average received SNR, defined as  $\Delta \text{SNR} \triangleq \frac{\rho N_c}{K} \cdot \mathbf{E}_H \left[ \lambda_{\max}(\mathbf{H}^H \mathbf{H}) - \sum_{m=1}^{N_t} \alpha_{jm} s_{im} \right]$ .

*Proof:* See Appendix D. ■

Thus, in a weak sense, the optimization in (20) is equivalent to maximizing the average received SNR of the quantized feedback scheme:

$$\max_{\{\alpha_{jm} : \sum_{m=1}^{N_t} \alpha_{jm} \leq 1 \text{ for all } j\}} \mathbf{E} \left[ \frac{\rho N_c}{K} \max_{i,j} \sum_{m=1}^{N_t} \alpha_{jm} s_{im} \right]. \quad (22)$$

With this new metric, we now establish the optimality of rank-one codebooks.

*Theorem 5:* Let the  $B$ -bit quantized feedback system and the corresponding  $N_1$  and  $N_2$  be described as before. The average SNR at the receiver is maximized by a rank-one codebook.

*Proof:* See Appendix E. ■

While Theorem 5 is true for all values of  $N_1$  and  $N_2$ , it is of relevance only when  $N_2 < N_t$ . It is important to note that the optimality of a rank-one codebook in terms of the new metric *does not* necessarily imply the optimality of rank-one codebooks in terms of the average mutual information. In fact, a signaling design that is based on maximizing the average SNR may destroy the degrees of freedom that are needed to achieve full diversity [31]. Nevertheless, for any choice of correlation and SNR, we expect a natural transition from weak optimality to strong optimality as  $B$  increases. Numerical studies suggest that for most reasonable correlation values, *rank-one codebooks are still optimal from a practical viewpoint*. This will be the focus of our future work. Furthermore, following the approach in Theorem 3, the system equation for the optimal coding scheme reduces to

$$\mathbf{Y} = \mathbf{H} \mathbf{U}_{i^*} [j^*] \mathbf{x}_{\text{trans}} + \mathbf{W} \quad (23)$$

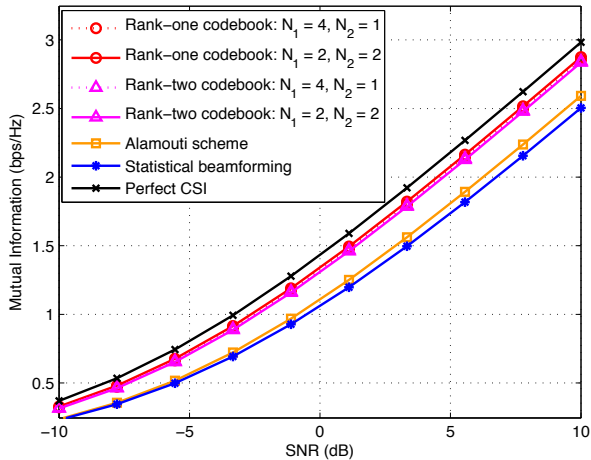


Fig. 1. Average mutual information of different codebooks in a  $2 \times 2$  i.i.d. channel.

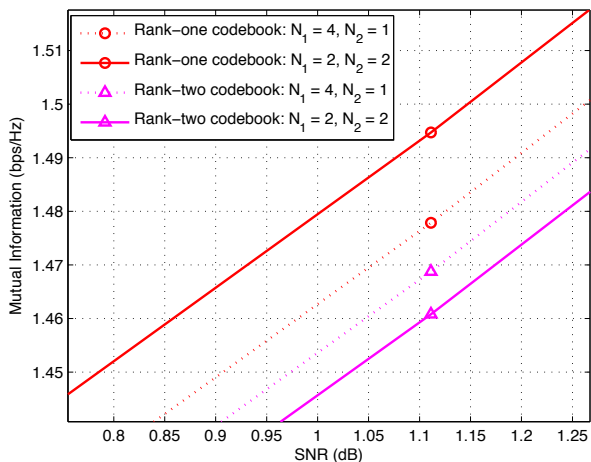


Fig. 2. A magnified view of the performance of rank-one and rank-two codebooks in the  $2 \times 2$  i.i.d. case.

where  $\mathbf{x}_{\text{trans}}$  is as in (16) and for any given realization  $\mathbf{H} = \mathbf{H}$ ,  $\mathbf{Q}^{i^*, j^*} = \mathbf{U}_{i^*} \mathbf{\Lambda}_{j^*} \mathbf{U}_{i^*}^\dagger$  with  $\mathbf{\Lambda}_{j^*} = \frac{N_t N_c}{K} \text{diag}(e_{j^*})$  and  $e_{j^*}$  is the  $j^*$ -th standard basis vector of  $\mathbb{R}^{N_t}$ .

It is critical to note the difference between a beamforming scheme in the classical sense and the system equation in (23). In the classical sense, the beamforming direction is *fixed and independent* of the channel state, but perhaps dependent on the channel statistics, which evolves over slower time scales. In (23), the beamforming direction is dependent on the channel state and is based on the feedback information. Despite the adaptation of this direction in response to the reverse link feedback, the low-complexity gain associated with beamforming (in the classical sense) can be accrued because we still need only a single radio link chain to implement this scheme. The need to adapt the beamforming direction at the transmitter at a fast rate<sup>6</sup> may impose additional constraints on the hardware, but these are expected to be sub-dominant in comparison with the performance improvement obtained by utilizing the feedback information.

To summarize, the main conclusion of this work is: Coding

<sup>6</sup>The rate has to be slightly faster than the rate at which the channel evolves.

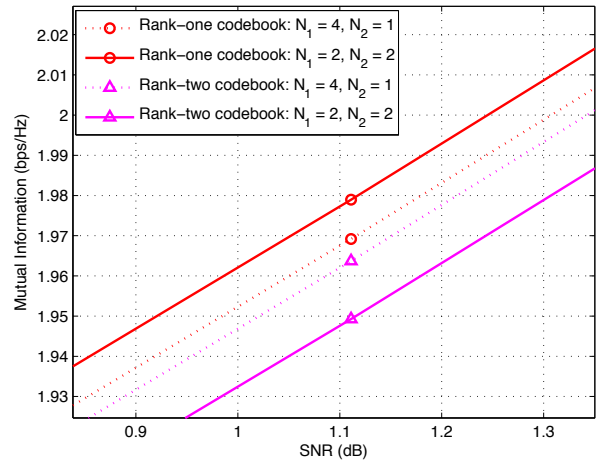


Fig. 3. A magnified view of the performance of rank-one and rank-two codebooks in the  $4 \times 4$  i.i.d. case.

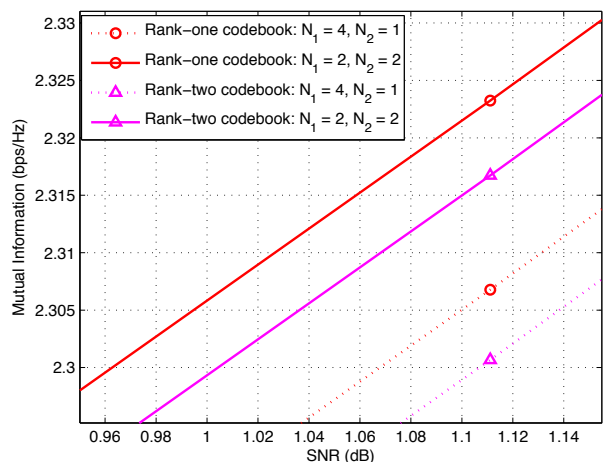


Fig. 4. A magnified view of the performance of rank-one and rank-two codebooks in the  $4 \times 4$  correlated case.

across time is not necessary to maximize the average mutual information if the quality of CSI feedback is sufficiently good. The same conclusion holds with a low quality of CSI feedback if the objective is relaxed to that of maximizing the average received SNR.

## V. SIMULATION RESULTS

We now present numerical studies to demonstrate that the rank-one codebook is a reasonable choice for most practical scenarios of interest. We study three settings here: 1) a  $2 \times 2$  i.i.d. channel, 2) a  $4 \times 4$  i.i.d. channel, and 3) a  $4 \times 4$  correlated channel with variance of channel entries given by

$$\mathbf{V}_4 = \frac{16}{2.6} \begin{bmatrix} 0.1 & 0 & 0.4 & 0 \\ 0 & 0.1 & 0.4 & 0 \\ 0 & 0 & 0.4 & 0.4 \\ 0 & 0 & 0.4 & 0.4 \end{bmatrix}. \quad (24)$$

In all the cases, the channel power  $\mathbf{E}[\text{Tr}(\mathbf{H}\mathbf{H}^\dagger)]$  is normalized to  $N_t N_r$ . Furthermore, we compare the mutual information between the best rank-one and best rank-two codebooks. The cases studied are: a)  $B = 2$ ,  $N_1 = 4$ ,  $N_2 = 1$ , and b)  $B = 2$ ,

$N_1 = 2, N_2 = 2$ . We now elaborate on how to obtain the best rank-one and rank-two codebooks.

From (19), the structure for each codeword is  $\mathbf{Q}^{i,j} = \mathbf{U}_i \mathbf{\Lambda}_j \mathbf{U}_i^\dagger$ . Fixing  $\{\mathbf{U}_i\}$ , the different rank-one codebooks are characterized by different choices of rank-one  $\{\mathbf{\Lambda}_j\}$ . For example, with  $N_r = N_t = 4, N_1 = 4$  and  $N_2 = 1$ , there are four choices of rank-one codebooks:  $\mathbf{\Lambda}_i = \text{diag}(e_i)$  where  $e_i$  is the  $i$ -th standard basis vector of  $\mathbb{R}^4$ . Similarly, there are six possible rank-one choices for  $N_r = N_t = 4, N_1 = 2$  and  $N_2 = 2$ . The best rank-one codebook is the one that maximizes the average mutual information. The above procedure is difficult to extend for rank-two codebooks. This is because even though there are only finite choices for the positions of the modes that can be excited, the power allocations between these two excited modes can run through a continuum. For example, with  $N_r = N_t = 4, N_1 = 4$  and  $N_2 = 1$ ,  $\mathbf{\Lambda}_1 = \text{diag}(1/2, 1/2, 0, 0)$ ,  $\mathbf{\Lambda}_1 = \text{diag}(1/3, 2/3, 0, 0)$ , or  $\mathbf{\Lambda}_1 = \text{diag}(0, 2/3, 0, 1/3)$  are all feasible choices for rank-two codebooks. This difficulty forces us to study this case by randomly generating 50 different sets of  $\{\mathbf{\Lambda}_j\}$  and picking the ‘best’ rank-two codebook from this random set. Further, since there is no proper distance metric to pack unitary matrices, a random family of  $\{\mathbf{U}_i\}$  are generated via random vector quantization (RVQ). Numerical studies show that there is roughly very similar performance with different choices of  $\{\mathbf{U}_i\}$  and hence, only one such choice is highlighted.

In the simulations, the choice of  $K$  used is  $N_c$ . This is because while rank-one codebooks meeting the GOC exist for up to  $K = 2N_c$ , the study in Sec. III-B suggests that rank-two codebooks that meet the GOC may not exist. We illustrate our results with Gaussian inputs, but numerical studies show that input constellation plays a minimal role in the trends.

Fig. 1 plots the mutual information with the best rank-one and rank-two codebooks for  $N_1 = 4, N_2 = 1$ , and for  $N_1 = 2, N_2 = 2$ . Benchmark plots of the perfect CSI (upper bound), only statistical information (lower bound), and statistical beamforming (lower bound) are also presented. We observe a 3 dB reduction in transmit power with two bits of feedback at a rate of 1 bps/Hz. A magnified view of this plot in Fig. 2 shows that the best rank-one codebook outperforms all other rank-two codebooks. In all subsequent plots, we focus only on a magnified view of the comparison between rank-one and rank-two codebooks since all plots show very similar trends for the mutual information, and the main focus is on our conjecture that a rank-one scheme leads to good performance in practice. Fig. 3 and Fig. 4 both verify that a rank-one codebook outperforms rank-two codebooks in the  $4 \times 4$  i.i.d. and  $4 \times 4$  correlated channels, respectively, thus suggesting that in most practical scenarios of interest, beamforming is a good candidate for optimal signaling.

## VI. CONCLUSION

In this work, we have studied the cases of coding across space and across space-time in a unified fashion by considering a family of linear dispersion codes that satisfy an orthogonality constraint. Our results show that there is no need to code across time either when the channel information at the transmitter is perfect or when the channel information

is of a sufficiently good quality. On the other hand, even when the channel information is not of a good quality, the low-complexity beamforming scheme possesses some attractive optimality properties; namely, it maximizes the average received SNR. From a design viewpoint, beamforming is particularly attractive due to the low-complexity of its design augmented with the low-cost ensured by using a single radio link chain.

Note that the orthogonal LD codes are of a complexity comparable to the OSTBC which are commonly used in standardization efforts. However, even in the case of rank-one signaling, one may be able to send  $K > 2N_c$  data-streams with dispersion matrices that do not meet the orthogonality constraint. The obvious disadvantage of this strategy is that the data-streams may have to be separated at the receiver with more complex decoding architectures. More so, the objective of maximizing the average mutual information of the inner space-time code can be met by *precoding schemes* that multiplex more than one data-stream, albeit at the cost of some decoding complexity. Related work in [31] suggests that relaxing the orthogonality condition could result in a different set of results. Thus, the orthogonality/low-complexity assumption of this work is critical to our conclusion and the trade-off between mutual information and decoding complexity is not clear. This is an interesting research topic that we would like to explore in the future. Furthermore, our work provides a good justification as to why there has been significant recent attention on limited feedback precoding/beamforming schemes [3], [4], [32]–[38] rather than on limited feedback space-time coding schemes.

Much work needs to be done to understand how these results translate to more practical scenarios of interest where the channel information at the receiver or the statistical information at the transmitter may not be perfect, and the channel is not block fading or wideband. It should also be noted that the results in this work critically depend on the error-free nature of the reverse link. Extension to more realistic scenarios where the reverse link suffers from delay and error are of interest. Construction of dispersion matrices that satisfy desired low-complexity properties is another area of interest. It is also important to note that we have only scratched the surface on understanding the trade-off between reliability and throughput with constraints on the complexity of the encoder-decoder pair. While reliability is an important design metric in certain situations, throughput is probably a more important aspect in the design of high data-rate wireless systems. In such settings, it is of interest to understand how and when low-complexity, adaptive signaling techniques can be leveraged to achieve near-optimal performance.

## APPENDIX

### A. Proof of Theorem 1

Denote  $\tilde{\mathbf{A}}_k \tilde{\mathbf{A}}_k^\dagger$  by  $\tilde{\mathbf{Q}}_k$  and  $\frac{1}{K} \sum_{k=1}^K \tilde{\mathbf{Q}}_k$  by  $\hat{\mathbf{Q}}$ , and observe that  $\hat{\mathbf{Q}}$  is a positive semi-definite matrix with trace constrained by  $\frac{N_t N_c}{K}$ . With  $\gamma \triangleq \mathbf{E}[I(\mathbf{X}; \mathbf{Y}|\mathbf{H} = \mathbf{H})] + K \mathbf{E}[h(\mathbf{n})]$ , we



have

$$\begin{aligned} \gamma &\stackrel{(a)}{\leq} \mathbf{E} \left[ \sum_{k=1}^K I_k \right] + K \mathbf{E} [h(\mathbf{n})] \\ &= \mathbf{E} \left[ \sum_k \varphi \left( \frac{\rho}{N_t} \text{Tr} (\tilde{\mathbf{Q}}_k \mathbf{H}^\dagger \mathbf{H}) \right) \right] \\ &\stackrel{(b)}{\leq} K \cdot \mathbf{E} \left[ \varphi \left( \frac{\rho}{N_t K} \sum_k \text{Tr} (\tilde{\mathbf{Q}}_k \mathbf{H}^\dagger \mathbf{H}) \right) \right] \\ &= K \cdot \mathbf{E} \left[ \varphi \left( \frac{\rho}{N_t} \text{Tr} (\hat{\mathbf{Q}} \mathbf{H}^\dagger \mathbf{H}) \right) \right] \end{aligned}$$

where equality holds in (a) if the GOC condition is satisfied and (b) follows from the concavity of  $\varphi(\cdot)$ . Optimizing over the choice of  $\hat{\mathbf{Q}}$  in the above equation results in an upper bound for  $\mathbf{E} [I(\mathbf{X}; \mathbf{Y} | \mathbf{H} = \mathbf{H})]$ . Denote by  $\mathbf{Q}_{\text{opt}}$  the solution to the following optimization problem:

$$\mathbf{Q}_{\text{opt}} = \arg \max_{\text{Tr}(\mathbf{Q}) = \frac{N_t N_c}{K}} \mathbf{E} \left[ \varphi \left( \frac{\rho}{N_t} \text{Tr} (\mathbf{H} \mathbf{Q} \mathbf{H}^\dagger) \right) \right] \quad (25)$$

where  $\mathbf{Q} = \tilde{\mathbf{A}} \tilde{\mathbf{A}}^\dagger$ . A choice of dispersion matrices  $\{\tilde{\mathbf{A}}_k\}$  that satisfies  $\tilde{\mathbf{A}}_k \tilde{\mathbf{A}}_k^\dagger = \mathbf{Q}_k = \mathbf{Q}_{\text{opt}}$  for all  $k$  and that meets the GOC condition would result in an equality in the upper bound and hence achieves the ergodic capacity. The fact that  $\mathbf{Q}_{\text{opt}}$  in (25) is diagonal follows from [15]. ■

### B. Proof of Theorem 2

The connection between minimum mean squared error (MMSE) estimation and mutual information established in [39] implies that  $\frac{d\varphi(a)}{da} = \frac{1}{2} \text{mse}(a)$  where  $\text{mse}(a)$  is the mean squared error for the channel under consideration at an SNR of  $a$ . The positivity and the monotonous decrease of the  $\text{mse}(\cdot)$  function implies that  $\varphi(\cdot)$  (and hence  $I_k(\cdot)$ ) is concave and non-decreasing. We first upper bound  $I_k$  and this leads to an upper bound on  $I(\mathbf{X}; \mathbf{Y} | \mathbf{H} = \mathbf{H})$ . For this, note that

$$\begin{aligned} I_k + h(\mathbf{n}) &= \varphi \left( \frac{\rho}{N_t} \text{Tr} (\mathbf{Q}_k \mathbf{H}^\dagger \mathbf{H}) \right) = \varphi \left( \frac{\rho}{N_t} \sum_{i=1}^{N_t} \lambda_i (\mathbf{Q}_k \mathbf{H}^\dagger \mathbf{H}) \right) \\ &\stackrel{(a)}{\leq} \varphi \left( \frac{\rho}{N_t} \lambda_1 (\mathbf{H}^\dagger \mathbf{H}) \text{Tr} (\mathbf{Q}_k) \right) \end{aligned}$$

where (a) follows from the fact that  $\lambda_i (\mathbf{Q}_k \mathbf{H}^\dagger \mathbf{H}) \leq \lambda_i (\mathbf{Q}_k) \lambda_1 (\mathbf{H}^\dagger \mathbf{H})$  and the monotonicity of  $\varphi(\cdot)$ . The concavity of  $\varphi$  implies that  $\eta = \sum_{k=1}^K I_k + K h(\mathbf{n})$  satisfies

$$\begin{aligned} \eta &\leq \sum_{k=1}^K \varphi \left( \frac{\rho}{N_t} \lambda_1 (\mathbf{H}^\dagger \mathbf{H}) \text{Tr} (\mathbf{Q}_k) \right) \\ &\leq K \varphi \left( \frac{\rho}{N_t} \frac{N_t N_c}{K} \lambda_1 (\mathbf{H}^\dagger \mathbf{H}) \right) = \varphi \left( \frac{\rho N_c}{K} \lambda_1 (\mathbf{H}^\dagger \mathbf{H}) \right). \end{aligned}$$

We now show that the upper bound is in fact achievable. Consider the maximization of  $\sum_k I_k$  over the set  $\mathcal{Q} = \{\mathbf{Q}_k = \mathbf{Q} \text{ for all } k, \mathbf{Q} \succ \mathbf{0} \text{ and } \text{Tr}(\mathbf{Q}) = \frac{N_t N_c}{K}\}$ . We then

have  $\omega = \varphi \left( \frac{\rho}{N_t} \text{Tr} (\mathbf{H} \mathbf{Q} \mathbf{H}^\dagger) \right)$  satisfying

$$\begin{aligned} \omega &= \varphi \left( \frac{\rho}{N_t} \text{Tr} (\mathbf{Q} \mathbf{H}^\dagger \mathbf{H}) \right) = \varphi \left( \frac{\rho}{N_t} \sum_{i=1}^{N_t} \lambda_i (\mathbf{Q} \mathbf{H}^\dagger \mathbf{H}) \right) \\ &\stackrel{(a)}{\leq} \varphi \left( \frac{\rho}{N_t} \sum_{i=1}^{N_t} \lambda_i (\mathbf{Q}) \lambda_i (\mathbf{H}^\dagger \mathbf{H}) \right) \\ &\stackrel{(b)}{\leq} \varphi \left( \frac{\rho}{N_t} \cdot \frac{N_t N_c}{K} \lambda_1 (\mathbf{H}^\dagger \mathbf{H}) \right) = \varphi \left( \frac{\rho N_c}{K} \lambda_1 (\mathbf{H}^\dagger \mathbf{H}) \right) \end{aligned}$$

where (a) follows from the monotonicity of  $\varphi(\cdot)$  and the fact that if  $\mathbf{A}$  and  $\mathbf{B}$  are  $n \times n$  positive semi-definite matrices, then  $\sum_{i=1}^n \lambda_i (\mathbf{A} \mathbf{B}) \leq \sum_{i=1}^n \lambda_i (\mathbf{A}) \lambda_i (\mathbf{B})$  and (b) follows from trivially upper bounding  $\lambda_i (\mathbf{H}^\dagger \mathbf{H})$  with  $\lambda_1 (\mathbf{H}^\dagger \mathbf{H})$ . Also note that the upper bound is achieved by beamforming along the dominant eigen-direction of  $\mathbf{H}^\dagger \mathbf{H}$ . Thus, we have

$$I(\mathbf{X}; \mathbf{Y} | \mathbf{H} = \mathbf{H}) \leq \sum_{k=1}^K I_k = K \varphi \left( \frac{\rho N_c}{K} \lambda_1 (\mathbf{H}^\dagger \mathbf{H}) \right) - K h(\mathbf{n})$$

with equality if and only if  $\mathbf{Q}$  is as above and the GOC conditions are satisfied. ■

### C. Proof of Theorem 2

First, note that any (generic) codebook can be written as

$$\begin{aligned} \mathcal{C} &= \{c^\ell, \ell = 1, \dots, 2^B\} \text{ where } c^\ell \triangleq (\mathbf{Q}_1^\ell, \dots, \mathbf{Q}_K^\ell) \\ &\text{s.t. } \sum_{k=1}^K \text{Tr}(\mathbf{Q}_k^\ell) \leq N_t N_c. \end{aligned}$$

Further, define a codebook  $\mathcal{D}$  as

$$\begin{aligned} \mathcal{D} &= \{d^\ell, \ell = 1, \dots, 2^B\} \text{ where } d^\ell \triangleq (\mathbf{Q}^\ell, \dots, \mathbf{Q}^\ell) \\ &\text{s.t. } \text{Tr}(\mathbf{Q}^\ell) \leq \frac{N_t N_c}{K}. \end{aligned}$$

Denoting the families of codebooks of the type  $\mathcal{C}$  and  $\mathcal{D}$  by  $\mathcal{C}$  and  $\mathcal{D}$ , respectively, we have  $\mathcal{D} \subset \mathcal{C}$ . With a codebook  $\mathcal{C}$  from  $\mathcal{C}$ , the average mutual information is

$$\begin{aligned} \mathbf{E}_{\mathbf{H}} \left[ \max_i \sum_{k=1}^K \varphi \left( \frac{\rho}{N_t} \text{Tr} (\mathbf{H} \mathbf{Q}_k^i \mathbf{H}^\dagger) \right) \right] \\ \stackrel{(a)}{\leq} \mathbf{E}_{\mathbf{H}} \left[ \max_i K \varphi \left( \frac{\rho}{N_t} \text{Tr} (\mathbf{H} \hat{\mathbf{Q}}^i \mathbf{H}^\dagger) \right) \right] \end{aligned}$$

where  $\hat{\mathbf{Q}}^i = \frac{1}{K} \sum_{k=1}^K \mathbf{Q}_k^i$  satisfies  $\text{Tr}(\hat{\mathbf{Q}}^i) \leq \frac{N_t N_c}{K}$  and (a) follows from the concavity of  $\varphi(\cdot)$ . Thus, the average mutual information with a codebook  $\mathcal{C}$  can be upper bounded by an appropriately generated codebook from  $\mathcal{D}$ . Since  $\mathcal{D} \subset \mathcal{C}$ , the upper bound is tight. ■

### D. Proof of Lemma 1

From the fundamental theorem of calculus and the MMSE connection in App. B, we have

$$\Delta I = \mathbf{E}_{\mathbf{H}} \left[ \int_A^B \text{mse}(x) dx \right] \quad (30)$$

where  $A = \frac{\rho N_c}{K} \sum_{m=1}^{N_t} \alpha_{j_m} s_{im}$ ,  $B = \frac{\rho N_c}{K} \lambda_{\max} (\mathbf{H}^\dagger \mathbf{H})$ , and  $\text{mse}(\cdot)$  is the MSE function. Note that  $A \leq B$ . Since a

Gaussian input maximizes the  $\text{mse}(\cdot)$  function [39], an upper bound to  $\Delta I$  can be achieved by replacing the entropy function  $\varphi(x)$  with that corresponding to a Gaussian input, (or equivalently,  $\log(1+x)$ ). Thus, we have

$$\begin{aligned} \Delta I &\leq \mathbf{E}_H [\log(1+B) - \log(1+A)] \\ &= \mathbf{E}_H \left[ \log \left( 1 + \frac{B-A}{1+A} \right) \right] \\ &\stackrel{(a)}{\leq} \frac{\rho N_c}{K} \cdot \mathbf{E}_H \left[ \frac{\lambda_{\max}(\mathbf{H}^H \mathbf{H}) - \sum_{m=1}^{N_t} \alpha_{jm} s_{im}}{1+A} \right] \\ &\stackrel{(b)}{\leq} \frac{\rho N_c}{K} \cdot \mathbf{E}_H \left[ \lambda_{\max}(\mathbf{H}^H \mathbf{H}) - \sum_{m=1}^{N_t} \alpha_{jm} s_{im} \right] \end{aligned}$$

where (a) follows from log-inequality and (b) trivially. ■

### E. Proof of Theorem 5

We need the following proposition towards proving the theorem.

*Proposition 3:* For any  $\{a_{j,k}, j = 1, \dots, M\}$  such that  $\sum_{k=1}^N a_{j,k} = 1$ , we have

$$\begin{aligned} &\max_{j=1, \dots, M} \sum_{k=1}^N a_{j,k} y_{j,k} \\ &\leq \sum_{k_1=1}^N a_{1,k_1} \cdots \sum_{k_M=1}^N a_{M,k_M} \max\{y_{1,k_1}, \dots, y_{M,k_M}\}. \end{aligned}$$

*Proof:* Note that  $\max\{y_i, z\} \geq y_i$  and  $\max\{y_i, z\} \geq z$  for any set of real numbers  $\{y_i\}_{i=1, \dots, N}$  and  $z$ . Thus, we have the inequality

$$\max \left\{ \sum_{i=1}^N \beta_i y_i, z \right\} \leq \sum_{i=1}^N \beta_i \max\{y_i, z\} \quad (31)$$

where  $\sum_{i=1}^N \beta_i = 1$ .

We now prove the proposition by induction. With  $M = 1$ , equality holds and the statement is trivially valid. If the proposition holds for some  $M$ , we then have the set of equations as in the bottom of the page, where (a) follows by applying (31) on the first term in the max and the hypothesis in Prop. 3 on the second term, and (b) by applying (31) on the second term in the max. ■

**Proof of Theorem 5:** To prove the theorem, we upper bound the average SNR at the receiver

$$\begin{aligned} \mathbf{E} \left[ \frac{\rho N_c}{K} \max_{i,j} \sum_{m=1}^{N_t} \alpha_{jm} s_{im} \right] &\stackrel{(a)}{\leq} \mathbf{E} \left[ \frac{\rho N_c}{K} \max_j \sum_{m=1}^{N_t} \alpha_{jm} \max_i s_{im} \right] \\ &\stackrel{(b)}{\leq} \mathbf{E} \left[ \frac{\rho N_c}{K} \sum_{m_1=1}^{N_t} \alpha_{1,m_1} \cdots \sum_{m_{N_2}=1}^{N_t} \alpha_{N_2,m_{N_2}} \right. \\ &\quad \left. \max \left\{ \max_i s_{im_1}, \dots, \max_i s_{im_{N_2}} \right\} \right] \\ &\stackrel{(c)}{\leq} \mathbf{E} \left[ \frac{\rho N_c}{K} \max \left\{ \max_i s_{im_1^*}, \dots, \max_i s_{im_{N_2}^*} \right\} \right] \end{aligned}$$

where (a) and (b) follow from Prop. 3, and (c) follows by letting

$$\begin{aligned} (m_1^*, \dots, m_{N_2}^*) &= \arg \max_{\{(m_1, \dots, m_{N_2}) : 1 \leq m_j \leq N_t \text{ for all } j=1, \dots, N_2\}} \\ &\quad \mathbf{E} \left[ \max_j \left\{ \max_i s_{im_j} \right\} \right]. \end{aligned}$$

The upper bound can be achieved by letting  $\alpha_{jm}^* = \delta_{jm}^*$  which again does not depend on the channel realization. That is, the optimal codebook is of rank-one. ■

### REFERENCES

- [1] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge Univ. Press, 2003.
- [2] E. Jorswieck and H. Boche, "Channel Capacity and Capacity-Range of Beamforming in MIMO Wireless Systems under Correlated Fading with Covariance Feedback," *IEEE Trans. Wireless Commun.*, vol. 52, no. 10, pp. 1654–1657, Oct. 2004.
- [3] D. J. Love, R. W. Heath, Jr., and T. Strohmer, "Grassmannian Beamforming for Multiple-Input Multiple-Output Wireless Systems," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.
- [4] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On Beamforming with Finite Rate Feedback in Multiple Antenna Systems," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2562–2579, Oct. 2003.
- [5] V. Raghavan, R. W. Heath, Jr., and A. M. Sayeed, "Systematic Codebook Designs for Quantized Beamforming in Correlated MIMO Channels," *IEEE Journ. Selected Areas in Commun.*, vol. 25, no. 7, pp. 1298–1310, Sept. 2007.
- [6] A. Wittneben, "Base Station Modulation Diversity for Digital Simulcast," *Proc. IEEE Veh. Tech. Conf.*, pp. 848–853, May 1991.
- [7] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-Time Codes for High Data Rate Wireless Communications: Performance Analysis and Code Construction," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.
- [8] S. Alamouti, "A Simple Transmit Diversity Technique for Wireless Communications," *IEEE Journ. Selected Areas in Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

$$\max_{j=1, \dots, M+1} \sum_{k=1}^N a_{j,k} y_{j,k} = \max \left\{ \sum_{k_{M+1}=1}^N a_{M+1,k_{M+1}} y_{M+1,k_{M+1}}, \max_{j=1, \dots, M} \sum_{k=1}^N a_{j,k} y_{j,k} \right\} \quad (32)$$

$$\stackrel{(a)}{\leq} \sum_{k_{M+1}=1}^N a_{M+1,k_{M+1}} \max \left\{ y_{M+1,k_{M+1}}, \sum_{k_1=1}^N a_{1,k_1} \cdots \sum_{k_M=1}^N a_{M,k_M} \max\{y_{1,k_1}, \dots, y_{M,k_M}\} \right\} \quad (33)$$

$$\stackrel{(b)}{\leq} \sum_{k_1=1}^N a_{1,k_1} \cdots \sum_{k_M=1}^N a_{M,k_M} \sum_{k_{M+1}=1}^N a_{M+1,k_{M+1}} \max \{ y_{M+1,k_{M+1}}, \max\{y_{1,k_1}, \dots, y_{M,k_M}\} \} \quad (34)$$

$$= \sum_{k_1=1}^N a_{1,k_1} \cdots \sum_{k_M=1}^N a_{M,k_M} \sum_{k_{M+1}=1}^N a_{M+1,k_{M+1}} \max \{ y_{1,k_1}, \dots, y_{M+1,k_{M+1}} \} \quad (35)$$

- [9] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-Time Block Codes from Orthogonal Designs," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.
- [10] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of Low-Density Parity-Check Codes for Modulation and Detection," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 670–678, Apr. 2004.
- [11] Y. Jiang, "An Information Theoretic Study on Linear Dispersion Codes and Low-Density Parity-Check Codes," *Ph.D. Thesis, University of Illinois at Urbana-Champaign*, 2005.
- [12] G. Bresler and B. Hajek, "Note On Mutual Information and Orthogonal Space-Time Codes," *Proc. IEEE Intern. Symp. Inform. Theory*, pp. 1315–1318, July 2006.
- [13] D. J. Love, R. W. Heath, Jr., W. Santipach, and M. L. Honig, "What is the Value of Limited Feedback for MIMO Channels?," *IEEE Commun. Magaz.*, vol. 42, no. 10, pp. 54–59, Oct. 2004.
- [14] G. Jongren, M. Skoglund, and B. Ottersten, "Design of Channel Estimate Dependent Space-Time Block Codes," *IEEE Trans. Commun.*, vol. 52, no. 7, pp. 1191–1203, July 2004.
- [15] C. Lin and V. V. Veeravalli, "Optimal Linear Dispersion Codes for Correlated MIMO Channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 657–666, Feb. 2008.
- [16] E. G. Larsson, G. Ganesan, P. Stoica, and W-H. Wong, "On the Performance of Orthogonal Space-Time Block Coding with Quantized Feedback," *IEEE Commun. Letters*, vol. 6, no. 11, pp. 487–489, Nov. 2002.
- [17] D. J. Love and R. W. Heath, Jr., "Limited Feedback Unitary Precoding for Orthogonal Space-Time Block Codes," *IEEE Trans. Sig. Proc.*, vol. 53, no. 1, pp. 64–73, Jan. 2005.
- [18] S. Ekbatani and H. Jafarkhani, "Combining Beamforming and Space-Time Coding Using Quantized Feedback," *To appear, IEEE Trans. Wireless Commun.*, 2007.
- [19] J. Akhtar and D. Gesbert, "Extending Orthogonal Block Codes with Partial Feedback," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 1959–1962, Nov. 2004.
- [20] M. Skoglund and G. Jongren, "On the Capacity of a Multiple-Antenna Communication Link with Channel Side Information," *IEEE Journ. Selected Areas in Commun.*, vol. 21, no. 3, pp. 395–405, April 2003.
- [21] B. Hassibi and B. Hochwald, "High-Rate Codes that are Linear in Space and Time," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 1804–1824, July 2002.
- [22] R. W. Heath, Jr. and A. J. Paulraj, "Linear Dispersion Codes for MIMO Systems Based on Frame Theory," *IEEE Trans. Sig. Proc.*, vol. 50, no. 10, pp. 2429–2441, Oct. 2002.
- [23] X. Wang, V. Krishnamurthy, and J. Wang, "Stochastic Gradient Algorithms for Design of Minimum Error-Rate Linear Dispersion Codes in MIMO Wireless Systems," *IEEE Trans. Sig. Proc.*, vol. 54, no. 4, pp. 1242–1255, Apr. 2006.
- [24] A. M. Sayeed, J. H. Kotecha, and Z. Hong, "Capacity Optimal Structured Linear Dispersion Codes for Correlated MIMO Channels," *Proc. IEEE Fall Veh. Tech. Conf.*, pp. 1623–1627, 2004.
- [25] V. Raghavan, J. H. Kotecha, and A. M. Sayeed, "Canonical Statistical Models for Correlated MIMO Fading Channels and Capacity Analysis," *Submitted, IEEE Trans. Inform. Theory*, 2007.
- [26] W. Weichselberger, M. Herdin, H. Ozelik, and E. Bonek, "A Stochastic MIMO Channel Model with Joint Correlation of Both Link Ends," *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 90–100, Jan 2006.
- [27] A. M. Tulino, A. Lozano, and S. Verdú, "Impact of Antenna Correlation on the Capacity of Multiantenna Channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2491–2509, July 2005.
- [28] A. M. Sayeed, "Deconstructing Multi-Antenna Fading Channels," *IEEE Trans. Sig. Proc.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.
- [29] A. M. Sayeed and V. V. Veeravalli, "Essential Degrees of Freedom in Space-Time Fading Channels," *Proc. 13th IEEE Intern. Symp. Personal Indoor and Mobile Radio Commun.*, vol. 4, pp. 1512–1516, Sept. 2002.
- [30] V. V. Veeravalli, Y. Liang, and A. M. Sayeed, "Correlated MIMO Rayleigh Fading Channels: Capacity, Optimal Signaling and Asymptotics," *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 2058–2072, June 2005.
- [31] T. Kim, M. Bengtsson, E. G. Larsson, and M. Skoglund, "Combining Long-Term and Low-Rate Short-Term Channel State Information over Correlated MIMO Channels," *To appear, IEEE Trans. Wireless Commun.*, 2008.
- [32] A. Narula, M.J. Lopez, M.D. Trott, and G.W. Wornell, "Efficient Use of Side Information in Multiple Antenna Data Transmission Over Fading Channels," *IEEE Journ. Selected Area in Commun.*, vol. 16, no. 8, pp. 1423–1436, Oct 1998.
- [33] E. Jorswieck and H. Boche, "Channel Capacity and Capacity Range of Beamforming in MIMO Wireless Systems under Correlated Fading with Covariance Feedback," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1543–1553, Sept. 2004.
- [34] V. K. N. Lau, Y. Liu, and T. A. Chen, "On the Design of MIMO Block Fading Channels with Feedback Link Capacity Constraint," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 62–70, Jan. 2004.
- [35] J. C. Roh and B. D. Rao, "Transmit Beamforming in Multiple-Antenna Systems with Finite Rate Feedback: A VQ-Based Approach," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1101–1112, Mar. 2006.
- [36] D. J. Love and R. W. Heath, Jr., "Limited Feedback Unitary Precoding for Spatial Multiplexing," *IEEE Trans. Inform. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.
- [37] J. C. Roh and B. D. Rao, "Multiple Antenna Channels with Partial Channel State Information at the Transmitter," *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 677–688, Mar. 2004.
- [38] D. J. Love and R. W. Heath, Jr., "Multimode Precoding for MIMO Wireless Systems," *IEEE Trans. Sig. Proc.*, vol. 53, no. 10, pp. 3674–3687, Oct. 2005.
- [39] D. Guo, S. Shamai, and S. Verdú, "Mutual Information and Minimum Mean-Square Error in Gaussian Channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.



**Che Lin (S'02)** received the B.Eng. degree in Electrical Engineering from National Taiwan University, Taipei, Taiwan in 1999. He received the M.S. degree in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign, IL in 2003, where he is currently pursuing the Ph.D. degree. His research interests include feedback systems, distributed algorithms, practical MIMO code design, optimization theory, and information theory.



**Vasanthan Raghavan (S'01-M'06)** received the B.Tech degree in Electrical Engineering from the Indian Institute of Technology at Madras, India in 2001, the M.S. and the Ph.D. degrees in Electrical and Computer Engineering in 2004 and 2006 respectively, and the M.A. degree in Mathematics in 2005, all from the University of Wisconsin-Madison, Madison, WI. He is currently a post-doctoral research associate with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL. His research interests span multi-antenna communication techniques, information theory, multihop networking, stochastic control, and random matrix theory.



**Venugopal V. Veeravalli (S'86-M'92-SM'98-F'06)** received the Ph.D. degree in 1992 from the University of Illinois at Urbana-Champaign, the M.S. degree in 1987 from Carnegie-Mellon University, Pittsburgh, PA, and the B.Tech degree in 1985 from the Indian Institute of Technology, Bombay, (Silver Medal Honors), all in Electrical Engineering. He joined the University of Illinois at Urbana-Champaign in 2000, where he is currently a Professor in the department of Electrical and Computer Engineering, and a Research Professor

in the Coordinated Science Laboratory. He served as a program director for communications research at the U.S. National Science Foundation in Arlington, VA from 2003 to 2005. He has previously held academic positions at Harvard University, Rice University, and Cornell University.

His research interests include distributed sensor systems and networks, wireless communications, detection and estimation theory, and information theory. He is a Fellow of the IEEE and was on the Board of Governors of the IEEE Information Theory Society from 2004 to 2007. He was an Associate Editor for Detection and Estimation for the IEEE Transactions on Information Theory from 2000 to 2003, and an associate editor for the IEEE Transactions on Wireless Communications from 1999 to 2000. Among the awards he has received for research and teaching are the IEEE Browder J. Thompson Best Paper Award, the National Science Foundation CAREER Award, and the Presidential Early Career Award for Scientists and Engineers (PECASE).