Optimal Linear Dispersion Codes for Correlated MIMO Channels

Che Lin and Venugopal V. Veeravalli, *Fellow, IEEE*

***Abstract*— The design of space-time codes for frequency flat, spatially correlated MIMO fading channels is considered. The focus of the paper is on the class of space-time block codes known as Linear Dispersion (LD) codes, introduced by Hassibi and Hochwald. The LD codes are optimized with respect to the mutual information between the inputs to the space-time encoder and the output of the channel. The use of the mutual information as both a design criterion and a performance measure is justified by allowing soft decisions at the output of the space-time decoder. A spatial Fourier (virtual) representation of the channel is exploited to allow for the analysis of MIMO channels with quite general fading statistics. Conditions, known as Generalized Orthogonal Conditions (GOC's), are derived for an LD code to achieve an upper bound on the mutual information, with the understanding that LD codes that achieve the upper bound, if they exist, are optimal. Explicit code constructions and properties of the optimal power allocation schemes are also derived. In particular, it is shown that optimal LD codes correspond to beamforming to a single virtual transmit angle at low SNR, and a necessary and sufficient condition for beamforming to be optimal is provided. Due to the nature of the code construction, it is further observed that the optimal LD codes can be designed to adapt to the statistics of different scattering environments. Finally, numerical results are provided to illustrate the optimal code design for three examples of sparse scattering environments. The performance of the optimal LD codes for these scattering environments is compared with that of LD codes designed assuming the i.i.d. Rayleigh fading (rich scattering) model, and it is shown that the optimal LD codes perform significantly better. The optimal LD codes are also compared to beamforming LD codes and it is shown that beamforming is nearly optimal over a range of SNR's of interest.**

***Index Terms*— Beamforming, fading channels, multiple-antennas, space-time codes, virtual representation.**

I. INTRODUCTION

THE use of antenna arrays at the transmitter and receiver to form a multi-input multi-output (MIMO) system has emerged as a powerful technique to improve the information rates and reliability of wireless links at low cost. The initial theoretical work of [1] and [2] has sparked considerable interest in designing practical schemes that can approach the capacity of MIMO channels. Practical approaches to coding for MIMO channels typically separate the encoding into an outer code concatenated with an inner block code, a space-time code, which is matched to the MIMO channel.

While it is optimal to jointly optimize the inner and outer codes, to simplify the optimization, the inner codes are generally designed separately using certain criteria. Among these criteria, the pairwise error probability (PEP) [3]-[6] and diversity gain [7] are most extensively used throughout the literature. Recent work on Low-Density Parity Check Codes (see, e.g, [8]) has shown that it is possible to construct outer codes that come close to achieving the mutual information between the input and output of the inner space-time code. This motivates the use of mutual information as a design criterion for space-time block code design. Hassibi and Hochwald [9] have applied the mutual information criterion to design optimal codes within the class of Linear Dispersion (LD) codes with i.i.d. Gaussian input symbol. Jiang [10] has considered design of optimal LD codes for i.i.d. Rayleigh MIMO fading channel with binary input symbols and conjectured that the optimizing LD code is the generalized orthogonal design introduced in [7]. Bresler and Hajek [11] later extended Jiang's work to real input symbols with arbitrary distribution and proved the conjecture proposed by Jiang. It is worth noting that in [9], mutual information is merely treated as a design criterion to devise good LD codes, while Bit Error Rate (BER) is used to measure the performance of the code. This is due to the implicit assumption that hard decisions are made at the LD decoder. If soft decisions are allowed (for instance, passing the likelihood function from the space-time decoder to the outer decoder), mutual information may be a more reasonable performance measure to consider, i.e., two candidate inner codes should be compared in terms of the mutual information they achieve. Therefore, in this paper, we treat mutual information as both a design criterion and a performance measure.

Much of the existing work on space-time block code design using the mutual information criterion has focused on the i.i.d. Rayleigh fading model for the channel. While this model is reasonable for rich scattering environments, correlation between the elements of the channel matrix needs to be considered in more general scattering scenarios. Significant performance degradation may occur if a space-time code intended for an uncorrelated scattering environment is used on a correlated MIMO channel. Therefore, adaptations of the code design to the scattering environments may be essential and beneficial. The main contributions of this work are two-fold: a) we extend the optimal LD code design of [10] and [11] to general MIMO channels with uniform linear arrays (ULA's) at the transmitter and receiver, and b) we study techniques for adaptive code design in changing scattering environments.

Manuscript received August 21, 2006; accepted December 7, 2006. The associate editor coordinating the review of this paper and approving it for publication was V. Bhargava. This research was supported in part by the NSF CAREER/PECASE award CCR 0049089 and the NSF award CCF 0431088, through the University of Illinois, and by a Vodafone Foundation Graduate Fellowship.

The authors are with the Coordinated Science Laboratory, Dept. of Electrical and Computer Engineering, Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801 (e-mail: {clin, vvv}@uiuc.edu).

Digital Object Identifier 10.1109/TWC.2008.060606.

II. Channel and System Model

A. Channel Model

Consider a frequency flat, multiple-antenna communication system with n_t transmit antennas and n_r receive antennas. In a discrete-time, complex baseband model, the transmitted signal matrix $\boldsymbol{X} \in \mathbb{C}^{n_t \times T}$ and the received signal matrix $\boldsymbol{Y} \in \mathbb{C}^{n_r \times T}$ are related by

$$\boldsymbol{Y} = \sqrt{\frac{\Gamma}{n_t}} \boldsymbol{H}\boldsymbol{X} + \boldsymbol{W}$$

where $\boldsymbol{H} \in \mathbb{C}^{n_r \times n_t}$ denotes the channel matrix, and $\boldsymbol{W} \in \mathbb{C}^{n_r \times T}$ denotes the complex additive white Gaussian noise with i.i.d. entries $\boldsymbol{W}_{ij} \sim \mathcal{CN}(0,1)$. We assume that the channel remains constant over the blocklength of T, i.e., T is smaller than the coherence time of the channel. We also assume that the channel changes in an ergodic fashion from block to block, and an average input power constraint of $\mathrm{E}[\mathrm{Tr}\{\boldsymbol{X}\boldsymbol{X}^\dagger\}] \leq n_t T$. Since each pair of transmit and receive antennas observes the same scattering environment, it is reasonable to assume that each entry of the channel matrix $\boldsymbol{H}$ is identically distributed. If we further assume that the channel matrix is normalized such that $\mathrm{E}[|\boldsymbol{H}_{i,j}|^2] = 1$, then Γ represents the effective signal-to-noise ratio (SNR) at each receive antenna. We assume that the receiver knows the realization of the channel matrix $\boldsymbol{H}$ while the transmitter only knows its distribution. In a rich scattering environment without line-of-sight (LOS) component, the distribution of $\boldsymbol{H}$ can be well modeled by i.i.d. zero-mean circularly-symmetric complex Gaussian random variables. However, in other scenarios, this may not be true and a general distribution of $\boldsymbol{H}$ needs to be considered. Our model and assumptions for the distribution of $\boldsymbol{H}$ are introduced in Section II-C.

B. System Model and Linear Dispersion Codes

As mentioned in the introduction, the coding problem for the MIMO channel $\boldsymbol{H}$ can be separated into the design of an inner space-time block code and an outer code, as shown in Figure 1. We assume that there are K streams of input symbols $\boldsymbol{x}_1[t], \ldots, \boldsymbol{x}_K[t]$ for the space-time encoder at a given symbol time t and that they satisfy the following assumption

Assumption 1: The streams of input symbols $\boldsymbol{x}_1[t], \ldots, \boldsymbol{x}_K[t]$ satisfy:

(i) For each stream, input symbols $\boldsymbol{x}_k[t]$ are i.i.d. across time and are drawn from some *real* constellations with marginal distribution $p(x_k)$.

(ii) Different streams are independent from each other.

The second assumption can be justified if $\boldsymbol{x}_1[t], \ldots, \boldsymbol{x}_K[t]$ 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. Since the input symbols $\boldsymbol{x}_k[t]$ are i.i.d. across time, we will drop the time index t for the rest of this paper.

Linear dispersion codes were first introduced in [9] and subsume both the V-BLAST signaling scheme [14] and the block codes of [7]. In this paper, we focus on applying LD codes as our inner space-time codes under the system model discussed above. The definition of a LD code involves a set of dispersion matrices $\{A_k\} \in \mathbb{C}^{n_t \times T}$ such that our space-time code $\boldsymbol{X}$ is given by

$$\boldsymbol{X} = \sum_{k=1}^{K} \boldsymbol{x}_k A_k \tag{1}$$

where the symbols $\{\boldsymbol{x}_k\}_{k=1}^K$ satisfy Assumption 1. Namely, at a given symbol time, the outer encoder produces a set of independent symbols $\{\boldsymbol{x}_k\}$. Information contained in $\{\boldsymbol{x}_k\}$ is then spread across the spatial and temporal dimensions through $\{A_k\}_{k=1}^K$. After normalizing $\boldsymbol{x}_k$ such that $\mathrm{E}[\boldsymbol{x}_k^2] = 1$, the power constraint is applied to A_k so that $\sum_{k=1}^K \mathrm{Tr}\{A_k A_k^\dagger\} \leq n_t T$. It may at first seem restrictive to assume that $\{\boldsymbol{x}_k\}$ are real. However, any sequence of complex symbols with independent real and imaginary parts can be generated under the model using two separate input symbol streams. Therefore, under the assumption that the complex input symbol has independent real and imaginary parts, the LD codes defined here are equivalent to those defined in [9].

Since a LD code is required to be an invertible mapping to guarantee successful decoding, and since the likelihood function $p(x_1, \ldots, x_K|Y, H)$ is a sufficient statistic for the estimation of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K$ from $(\boldsymbol{Y}, \boldsymbol{H})$, we obtain

$$\begin{aligned} I(\boldsymbol{X}; \boldsymbol{Y}|\boldsymbol{H}) &= I(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K; \boldsymbol{Y}|\boldsymbol{H}) \\ &= I(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K; \hat{\boldsymbol{Y}}|\boldsymbol{H}) \end{aligned} \tag{2}$$

where $\boldsymbol{Y}$ is the output of the channel and $\hat{\boldsymbol{Y}}$ is the likelihood function produced by the ML decoder. From (2), we see that for a fixed distribution on $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K$, $I(\boldsymbol{X}; \boldsymbol{Y}|\boldsymbol{H})$ is an achievable rate for the effective MIMO channel (See Figure 1) under the LD restriction. Therefore, our goal is to find LD codes that maximize the achievable rate for the effective MIMO channel.

C. Virtual Representation

The virtual representation is a succinct method to capture the scattering environment in general scenarios ([12], [13]). In [12], the virtual representation is applied to compute the capacity of correlated MIMO channels. Assuming ULA's at the transmitter and receiver, the virtual representation matrix of the channel is the two-dimensional spatial Fourier transformation

$$\tilde{\boldsymbol{H}} = S_r^\dagger \boldsymbol{H} S_t \tag{3}$$

where S_r and S_t are unitary spatial Fourier transform matrices. If the scattering does not have a Line of Sight (LOS) component, under the standard uncorrelated scattering assumption, we have the following properties of $\tilde{\boldsymbol{H}}$ [12]:

Property 1: The virtual channel matrix satisfies

(i) The elements of $\tilde{\boldsymbol{H}}$ are independent.

(ii) $\tilde{\boldsymbol{H}}_{k,\ell}$ is a zero-mean proper-complex random variable, and $\tilde{\boldsymbol{H}}_{k,\ell}$ has same distribution as $-\tilde{\boldsymbol{H}}_{k,\ell}$.

The virtual representation can be easily extended when there is a LOS path, as in [12]. For simplicity of presentation, we restrict to non-LOS scenarios. By applying the virtual representation (3) of the channel matrix $\boldsymbol{H}$, we convert our

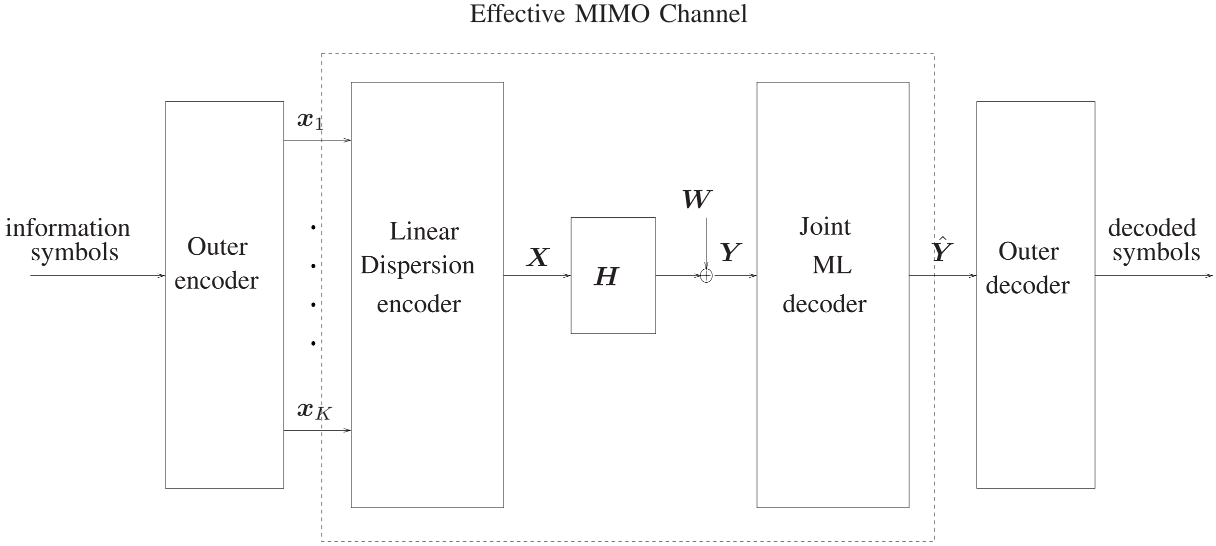

Fig. 1. System Model

signal space into the virtual domain. In the virtual domain, our channel model becomes

$$\tilde{\boldsymbol{Y}} = \sqrt{\frac{\Gamma}{n_t}} \sum_{k=1}^{K} \boldsymbol{x}_k \tilde{\mathbf{H}} \tilde{A}_k + \tilde{\boldsymbol{W}} \tag{4}$$

where $\tilde{A}_k = S_t^\dagger A_k$, $\tilde{\boldsymbol{Y}} = S_r^\dagger \boldsymbol{Y}$, and $\tilde{\boldsymbol{W}} = S_r^\dagger \boldsymbol{W}$. Since S_t is unitary, the power constraint in the virtual domain remains the same and is given by $\sum_{k=1}^{K} \text{Tr}\{\tilde{A}_k \tilde{A}_k^\dagger\} \leq n_t T$. We also define V as the variance matrix such that $V_{k,\ell} = \text{Var}(\tilde{\boldsymbol{H}}_{k,\ell})$. For the discussions that follow, we will focus on this channel model (4) in the virtual domain and make use of *Property* 1.

III. Mutual Information Criterion for Optimal Linear Dispersion Code

Under the mutual information criterion discussed earlier, our goal is to find the optimal $\{\tilde{A}_k\}_{k=1}^K$ such that the mutual information $I(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K; \tilde{\boldsymbol{Y}} | \tilde{\boldsymbol{H}})$ is maximized. We first state the following theorem.

Theorem 1: Let $\tilde{\boldsymbol{X}}$ be a LD code (1) with K symbols and corresponding dispersion matrices $\{\tilde{A}_k\}_{k=1}^K \in \mathbb{C}^{n_t \times T}$. Assuming that the real symbols $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K$ satisfy *Assumption* 1, the following upper bound on the mutual information holds

$$\begin{aligned} I(\tilde{\boldsymbol{X}}; \tilde{\boldsymbol{Y}} | \tilde{\boldsymbol{H}}) &= I(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K; \tilde{\boldsymbol{Y}} | \tilde{\boldsymbol{H}}) \\ &\leq \sum_{k=1}^{K} I\left(\boldsymbol{x}_k; \sqrt{\frac{\Gamma}{n_t}} \boldsymbol{x}_k \tilde{\boldsymbol{H}} \tilde{A}_k + \tilde{\boldsymbol{W}} \,\middle|\, \tilde{\boldsymbol{H}}\right) \end{aligned} \tag{5}$$

Equality holds if and only if $\tilde{A}_k \tilde{A}_j^\dagger + \tilde{A}_j \tilde{A}_k^\dagger = 0, \ \forall k, j$ such that $k \neq j$. Moreover, the following conditions are equivalent:

(i) $\tilde{A}_k \tilde{A}_j^\dagger + \tilde{A}_j \tilde{A}_k^\dagger = 0 \ \ \forall k, j \ \ k \neq j$

(ii) $p(x_1, \ldots, x_K | \tilde{Y}, \tilde{H}) = \prod_{k=1}^{K} p(x_k | \tilde{Y}, \tilde{H})$

(iii) $p(\tilde{Y} | x_1, \ldots, x_K, \tilde{H}) = c \prod_{k=1}^{K} p(\tilde{Y} | x_k, \tilde{H})$, where $c = p(\tilde{Y} | \tilde{H})^{-(K-1)}$

Proof : This result is proved in Lemma 3.1 in [10] for the special case where each element of $\tilde{\boldsymbol{H}}$ is i.i.d. $\mathcal{CN}(0,1)$ and $\{\boldsymbol{x}_k\}$ are binary and equiprobable. As in [10], *Assumption* 1 is crucial in deriving the upperbound. The key steps of the proof do not exploit the specific distribution of $\tilde{\boldsymbol{H}}$ and $\{\boldsymbol{x}_k\}$. The extension is thus straightforward. □

We referred to Condition (i) as the first Generalized Orthogonal Condition (GOC's). Conditions (ii) and (iii) in Theorem 1 state that the *a posteriori* probability and the output likelihood function factor. If any of these conditions holds, the complexity of the LD decoder is greatly reduced since the joint maximum-likelihood decoding is equivalent to applying maximum-likelihood decoding to each symbol.

In the next theorem, we maximize the upper bound given in Theorem 1. The upperbound is a summation of terms of the form

$$I\left(\boldsymbol{x}_k; \sqrt{\frac{\Gamma}{n_t}} \boldsymbol{x}_k \tilde{\boldsymbol{H}} \tilde{A}_k + \tilde{\boldsymbol{W}} \,\middle|\, \tilde{\boldsymbol{H}}\right)$$

Therefore, maximizing the upper bound is equivalent to maximizing each term subject to the individual power constraint $\text{Tr}\{\tilde{A}_k \tilde{A}_k^\dagger\} \leq \alpha_k n_t T$. In [10], it was conjectured that the optimal LD codes for i.i.d. Rayleigh fading MIMO channels with binary input symbols satisfy the first GOC and $\tilde{A}_k^o \tilde{A}_k^{o\dagger} = \alpha_k T \mathrm{I}$. Bresler and Hajek [11] later generalized [10] to real input symbols with arbitrary distribution and proved the conjecture proposed in [10]. We further extend their results (Lemma 2) and use techniques in [12] to show that the optimal $\tilde{A}_k^o \tilde{A}_k^{o\dagger}$ is diagonal. For notational simplicity, we ignore the subscript k in the following discussion.

Theorem 2: Suppose that the channel matrix $\tilde{\boldsymbol{H}}$ satisfies *Property* 1. Under an input power constraint $\text{Tr}\{\tilde{A}\tilde{A}^\dagger\} \leq \alpha n_t T, \ 0 \leq \alpha \leq 1$, a dispersion matrix $\tilde{A}^o$ that maximizes the mutual information $I\left(\boldsymbol{x}; \sqrt{\frac{\Gamma}{n_t}} \boldsymbol{x} \tilde{\boldsymbol{H}} \tilde{A} + \tilde{\boldsymbol{W}} \,\middle|\, \tilde{\boldsymbol{H}}\right)$ must be such that $\tilde{A}^o \tilde{A}^{o\dagger} = \Lambda^o$, where Λ^o is diagonal.

Proof: As in the proof of Lemma 2 in [11], we define $\varphi(a) = h(\sqrt{a}\boldsymbol{x} + \tilde{\boldsymbol{n}} | \tilde{\boldsymbol{H}} = \tilde{H})$, where $\tilde{\boldsymbol{n}}$ is a real zero-mean Gaussian random variable with variance $1/2$. The proof of *Theorem* 2 then follows techniques similar to those in [12] except with the difference that having a concave function φ in our expression complicates the proof. Moreover, the fact that φ is concave instead of strictly concave weakens the result in terms of the uniqueness of the optimal diagonal matrix Λ^o. The details are given in Appendix I. □

If we denote the optimal diagonal matrix for the kth input symbol by $\Lambda_k^o = \text{diag}\{\lambda_{k,1}^o, \ldots, \lambda_{k,n_t}^o\}$, then the following power constraints hold

$$\sum_{j=1}^{n_t} \lambda_{k,j}^o \leq \alpha_k n_t T \quad \text{and} \quad \sum_{k=1}^{K} \alpha_k = 1. \tag{6}$$

i.e., we assume separate power constraints $\{\alpha_k n_t T\}_{k=1}^K$ for each symbol and they sum up to $n_t T$. Therefore, for any channel matrix $\boldsymbol{H}$ that satisfies *Property* 1, we have the main result of this paper which is stated below.

Theorem 3: Let $\tilde{\boldsymbol{X}}$ be a LD code (1) with K symbols and corresponding dispersion matrices $\{\tilde{A}_k\}_{k=1}^K \in \mathbb{C}^{n_t \times T}$, and assume that input symbols $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K$ satisfy *Assumption* 1. A universal upperbound on the mutual information (5) is achieved if there exist LD codes satisfying the set of Generalized Orthogonal Conditions (GOC's):

(*i*) $\tilde{A}_k \tilde{A}_j^\dagger + \tilde{A}_j \tilde{A}_k^\dagger = 0 \ \ \forall k, j \ \ k \neq j$.

(*ii*) $\tilde{A}_k \tilde{A}_k^\dagger = \Lambda_k^o, \ \ k = 1, \ldots, K$, where Λ_k^o is a diagonal matrix that maximizes $I\left(\boldsymbol{x}_k; \sqrt{\frac{\Gamma}{n_t}} \boldsymbol{x}_k \tilde{\boldsymbol{H}} \tilde{A}_k + \tilde{\boldsymbol{W}} \middle| \tilde{\boldsymbol{H}}\right)$ and satisfies the power constraints (6).

Such LD codes maximize the mutual information $I(\tilde{\boldsymbol{X}}; \tilde{\boldsymbol{Y}} | \tilde{\boldsymbol{H}}) = I(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K; \tilde{\boldsymbol{Y}} | \tilde{\boldsymbol{H}})$ and are thus optimal under the mutual information criterion.

Proof: The second GOC maximizes each term on the RHS of (5) and therefore gives us a universal upperbound on the mutual information $I(\tilde{\boldsymbol{X}}; \tilde{\boldsymbol{Y}} | \tilde{\boldsymbol{H}})$ independent of the dispersion matrices $\{\tilde{A}_k\}$. The first GOC further implies that the universal upperbound is reached. Therefore, $I(\tilde{\boldsymbol{X}}; \tilde{\boldsymbol{Y}} | \tilde{\boldsymbol{H}})$ is maximized when both the first and second GOC's are satisfied. □

IV. Design of Optimal Dispersion Matrices

The first GOC implies that equality is reached in (5) and we obtain

$$\begin{aligned} &I(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K; \tilde{\boldsymbol{Y}} | \tilde{\boldsymbol{H}}) \\ &= \sum_{k=1}^{K} I\left(\boldsymbol{x}_k; \sqrt{\frac{\Gamma}{n_t}} \boldsymbol{x}_k \tilde{\boldsymbol{H}} \tilde{A}_k + \tilde{\boldsymbol{W}} \middle| \tilde{\boldsymbol{H}}\right) \end{aligned} \tag{7}$$

As mentioned in Section III, this equality can be interpreted as the decomposition of the joint ML decoder into individual ML decoders for each symbol. Therefore, each symbol $\boldsymbol{x}_k$ sees its own channel and we have an equivalent set of K parallel channels. Finding the optimal Λ_k^o and the corresponding dispersion matrix $\tilde{A}_k^o$ is then equivalent to obtaining the optimal power allocation for the kth subchannel. In the following section, we focus on finding the optimal power allocation for the kth subchannel and ignore the subscript k. Optimal power allocations for other subchannels can be found through similar procedures.

A. Necessary and Sufficient Condition for Beamforming to be Optimal

Although simple convex optimization algorithms can be applied to obtain optimal power allocations numerically, it is insightful to understand the asymptotic behavior of the optimal Λ^o in the low SNR regime. Based on the capacity result in [12], one should expect that beamforming is asymptotically optimal in the low SNR regime. This can actually be shown using techniques similar to those in [12]. Furthermore, one can establish an explicit necessary and sufficient condition such that beamforming is optimal. The following theorem describes the condition.

Theorem 4: Beamforming in the ith virtual transmit angle is optimal if and only if

$$\text{E}[\varphi'(\Gamma\alpha T \chi_i)] \sum_{k=1}^{n_r} V_{k,l^o} - \text{E}[\varphi'(\Gamma\alpha T \chi_i)\chi_i] \leq 0 \tag{8}$$

where

$$l^o = \arg \max_{1 \leq j \leq n_t, j \neq i} \sum_{k=1}^{n_r} V_{k,j} \quad \text{and} \quad \chi_j = \sum_{k=1}^{n_r} |\tilde{\boldsymbol{H}}_{k,j}|^2$$

Proof: Following the arguments similar to those given in the proof of Theorem 3 in [12], we first establish a necessary and sufficient condition for beamforming to be optimal for the first virtual angle. Generalizing the condition to the ith transmit virtual angle is then straightforward. For details of the proof, refer to Appendix II . □

B. Explicit Code Construction

In section III, we derived conditions (GOC's) for a LD code to achieve an upper bound on the mutual information. The LD codes that achieve the upper bound, if they exist, are optimal under the mutual information criterion. The second GOC involves finding the optimal diagonal power allocation matrices. Assume that the same power constraints hold for all subchannels and denote the optimal power allocation matrix by $\Lambda^o = \text{diag}\{\lambda_1^o, \ldots, \lambda_{n_t}^o\}$. Let N be the total number of non-zero diagonal elements of Λ^o and i_m be the index for the m^{th} non-zero diagonal elements. We can then observe that every row of $\tilde{A}_k$ is the all-zero vector except the N rows with indices $\{i_m\}_{m=1}^N$. And we can express $\tilde{A}_k$ as

$$\tilde{A}_k = [\underline{0}, \ldots, \underline{0}, \underline{a}_k^{i_1 \top}, \ldots, \underline{a}_k^{i_N \top}, \underline{0}, \ldots, \underline{0}]^\top$$

where $\underline{a}_k^{i_m}$ is the i_mth row vector for the dispersion matrix $\tilde{A}_k$ and $\underline{0}$ is the all-zero column vector. Therefore, finding dispersion matrices $\{\tilde{A}_k\}_{k=1}^K$ satisfying the set of GOC's is equivalent to finding $\{B_k\}_{k=1}^K$ such that

$$\begin{aligned} B_k B_k^\dagger &= \Lambda^{sub} \\ B_k B_j^\dagger + B_j B_k^\dagger &= 0 \ \ \forall j \neq k \end{aligned}$$

where $\Lambda^{sub} = \text{diag}\{\lambda_{i_1}^o, \ldots, \lambda_{i_N}^o\}$ and $B_k = [\underline{a}_k^{i_1 \top}, \ldots, \underline{a}_k^{i_N \top}]^\top$, which are obtained by stacking all non-zero diagonal elements of Λ_k^o and non-zero rows of $\tilde{A}_k$ together, respectively. If we further define

$U_k = [\Lambda^{sub}]^{-1/2} B_k$, our problem becomes finding $\{U_k\}_{k=1}^{K}$ with the properties:

$$U_k U_k^\dagger = I \quad (9)$$
$$U_k U_j^\dagger + U_j U_k^\dagger = 0 \quad \forall j \neq k \quad (10)$$

These $\{U_k\}$ correspond to the generalized orthogonal design introduced in [7] and can be constructed explicitly if

$$N \leq \rho(T) \text{ and } K \leq T \quad (11)$$

where $\rho(T)$ is defined in [7]. As a consequence, finding the optimal LD codes reduces to finding the generalized orthogonal design that satisfies the constraints (9) and (10).

C. Adaptation to Scattering Environments

The nature of the code construction mentioned above allows us to design codes that can adapt to different scattering environments. In particular, let $\{U_k^I\}$ and $\{U_k^{II}\}$ be the transformed dispersion matrices that correspond to scattering environment I and II, respectively. Moreover, let Λ_I^o be the optimal power allocation matrix for environment I and Λ_{II}^o be that for environment II. If the rank of Λ_{II}^o is less than that of Λ_I^o, each U_k^{II} is simply a sub-matrix of U_k^I. Therefore, our codes can adapt from scattering environment I to environment II in a straightforward fashion. Starting with a set of transformed dispersion matrices of a scattering environment with a full-rank optimal power allocation matrix, the transformed dispersion matrices of that with a smaller rank optimal power allocation matrix can therefore be obtained directly. The adaptation is crucial since in practice even the statistics of the wireless environment are prone to change, especially when users of the wireless service are mobile. The simple adaptation scheme allows real-time adjustment of the optimal LD codes when the statistics of the scattering environment change drastically. For example, when a mobile in the outdoor environment enters an indoor environment. The adaptation scheme mentioned above should be carefully distinguished from the MIMO Precoder System discussed in [17] to [19]. Our code design adapts to the statistics of the fading channel while in the MIMO Precoder System, the precoder is selected according the realization of the fading channel matrix.

D. Effect of Varying Parameters K and T

For all the discussions above, we fixed the number of input symbols to be K and the blocklength of our space-time code to be T. The question then arises as to what the optimal values of K and T are if we are given the freedom to vary these two parameters. For any fixed T, it is argued in [9] that the maximum mutual information increases with increasing values of K if $K \leq 2n_r T$. However, we notice that in order to make our LD code an invertible mapping to guarantee a successful decoding, we further require that $K \leq 2n_t T$. The maximum value for K is therefore

$$K_{max} = 2\min\{n_r, n_t\}T$$

and has a nice interpretation as the maximum number of real degrees of freedom of the MIMO channel. From the independence bound on the mutual information, choosing $T = 1$ is optimal, which gives rise to the optimality of the V-BLAST signaling scheme. It is therefore tempting to pick $T = 1$ and $K = 2\min\{n_r, n_t\}$. However, for $T = 1$, the first GOC can never be satisfied. Moreover, with a larger value of K, we need more dispersion matrices $\tilde{A}_k$ that simultaneously meet the first GOC. The first GOC is hence more difficult to be satisfied. It then becomes an engineering problem to find the optimal K and T such that both the GOC's can be satisfied and the overall mutual information is maximized. We further illustrate this design issue by discussing the optimality of the Alamouti scheme under the mutual information criterion.

The Alamouti scheme [15] ($n_t = T = 2$) can be expressed as a LD code defined in (1) by

$$\begin{bmatrix} \boldsymbol{s}_1 & -\boldsymbol{s}_2^* \\ \boldsymbol{s}_2 & \boldsymbol{s}_1^* \end{bmatrix} = \boldsymbol{x}_1 \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix} + \boldsymbol{x}_2 \begin{bmatrix} \frac{i}{\sqrt{2}} & 0 \\ 0 & \frac{-i}{\sqrt{2}} \end{bmatrix} + \boldsymbol{x}_3 \begin{bmatrix} 0 & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 \end{bmatrix} + \boldsymbol{x}_4 \begin{bmatrix} 0 & \frac{i}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & 0 \end{bmatrix}$$

where $\boldsymbol{s}_1 = \frac{1}{\sqrt{2}}\boldsymbol{x}_1 + i\frac{1}{\sqrt{2}}\boldsymbol{x}_2$ and $\boldsymbol{s}_2 = \frac{1}{\sqrt{2}}\boldsymbol{x}_3 + i\frac{1}{\sqrt{2}}\boldsymbol{x}_4$. It is straightforward to show that the Alamouti scheme satisfies both GOC's and hence achieves the maximum achievable rate of the effective MIMO channel for $K = 4$ and $T = 2$. If we further assume that $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_4$ are the capacity achieving, zero-mean, real Gaussian random variables, we then expect that the Alamouti scheme achieves the capacity of the MIMO channel. However, the optimality of the Alamouti scheme actually depends on the number of receive antennas n_r. For $n_r = 1$, $K = 4$ is the optimal value since the maximum value of K is $4\min\{n_r, n_t\}$. In this case, the Alamouti scheme achieves the capacity of a 2x1 MIMO channel. However, for $n_r \geq 2$, $K = 4$ is smaller than k_{max} and hence the Alamouti scheme is suboptimal. The above observations are consistent with the results given in [9].

V. Simulation Results

In this section, we present simulation results that illustrate the advantage of designing the LD codes according to the correlation of the channel. We assume that equal power constraints are applied to all input symbols such that the optimal power allocations are the same for all subchannels. If we further let $K = T$, where T is chosen such that (11) is satisfied and omit the subscript k for notational simplicity, our optimization problem becomes that of finding the optimal power allocation matrix Λ^o for each subchannel such that

$$I(\Lambda) = \mathrm{E}\left[\varphi\left(\frac{\Gamma}{n_t}\mathrm{Tr}\{\tilde{\boldsymbol{H}}\Lambda\tilde{\boldsymbol{H}}^\dagger\}\right)\right] - h(\tilde{\boldsymbol{n}})$$

is maximized. The maximum achievable rate for the effective MIMO channel is then $\frac{K}{T}I(\Lambda^o) = I(\Lambda^o)$. The optimal power allocation matrix Λ^o can be found numerically by using the Stochastic Quasi-Gradient Algorithm [20]. We apply the Kiefer-Wolfowitz procedure, which approximates the gradient function by the divided differences of $I(\Lambda)$. A convergence analysis of the Kiefer-Wolfowitz procedure can be found in [21]. Since the fluctuations of the estimate for $I(\Lambda)$ are large, we apply the Control Variate Method [22] to reduce the

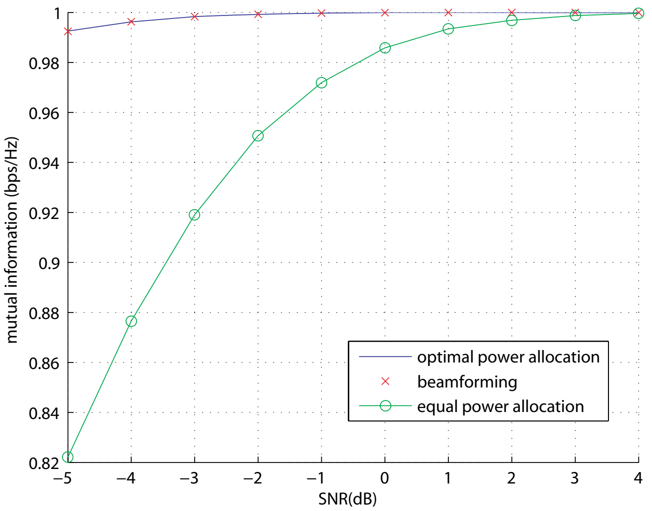

Fig. 2. Comparison of the mutual information using optimal power allocation, beamforming, and equal power allocation for scattering environment I using BPSK.

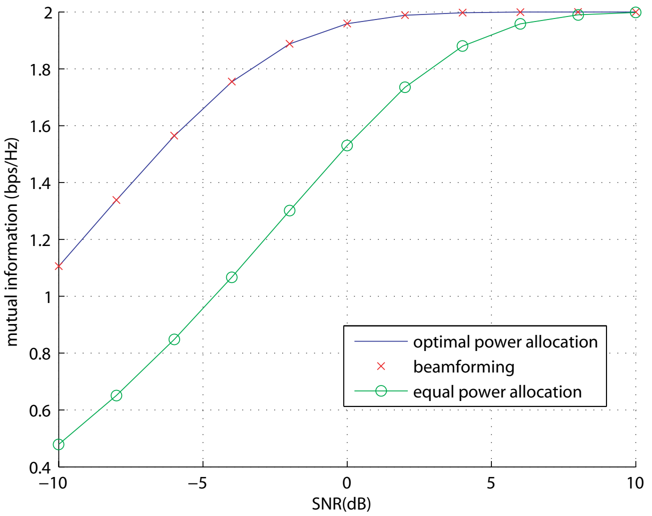

Fig. 3. Comparison of the mutual information using optimal power allocation, beamforming, and equal power allocation for scattering environment I using 4-PAM.

variance of $I(\Lambda)$. For the results that follow, we apply several different input distributions that correspond to different signal constellations for the input symbols, including BPSK, 4-PAM, and Gaussian inputs.

A. Scattering Environment I

In the first scattering environment, we have a system of 5 transmit and 5 receive antennas [12], where each element of $\tilde{\boldsymbol{H}}$ is a zero-mean proper complex Gaussian random variable with the following variance matrix,

$$V = \frac{25}{5.7}\begin{pmatrix} 0.1 & 0 & 1 & 0 & 0 \\ 0 & 0.1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0.25 & 0 \\ 0 & 0 & 1 & 0 & 0.25 \end{pmatrix}$$

The entries of V are normalized so that $\sum_{k,\ell} V_{k,\ell} = n_t n_r = 25$. Such a variance matrix corresponds to a physical environment with two very small scatterers, two bigger scatterers, and one large scattering cluster.

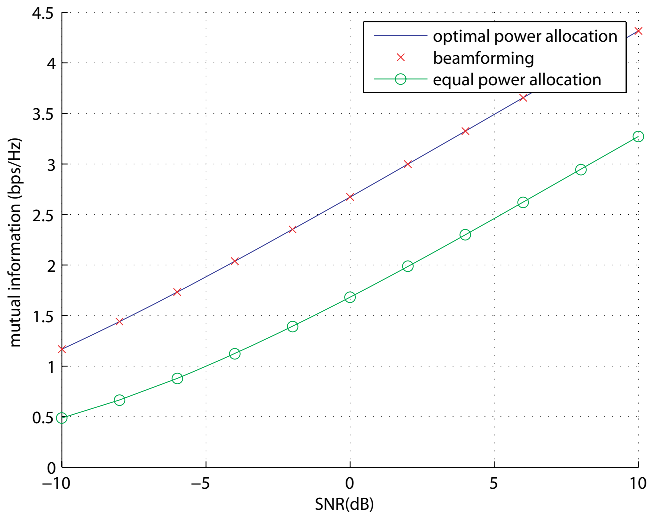

Fig. 4. Comparison of the mutual information using optimal power allocation, beamforming, and equal power allocation for scattering environment I using Gaussian inputs.

We obtain the optimal power allocation and compare the mutual information achieved by the optimal power allocation, beamforming, and equal power allocation. For Figure 2, BPSK is used for the input constellation. We observe that beamforming is almost always optimal even until the saturation of mutual information at 4dB and the mutual information achieved by the optimal power allocation and beamforming is larger than that achieved by equal power allocation. Similar observations can be made for the 4-PAM input constellations, with equiprobable input symbols (see Figure 3), and for Gaussian inputs (see Figure 4). Moreover, the advantage of using the optimal power allocation is even more significant with larger constellation sizes. Finally, we see that beamforming to the third transmit angle is close to optimal. This is as expected since the channel gains of the third virtual transmit angle are much higher than that of other transmit angles (the sum of the third column of V is much larger than sums of other columns).

B. Scattering Environment II

For the second scattering environment, we consider the following variance matrix

$$V = \frac{25}{1}\begin{pmatrix} 0.1 & 0 & 0.06 & 0 & 0 \\ 0 & 0.1 & 0.06 & 0 & 0 \\ 0 & 0 & 0.06 & 0 & 0 \\ 0 & 0 & 0.06 & 0.25 & 0 \\ 0 & 0 & 0.06 & 0 & 0.25 \end{pmatrix}$$

This environment is similar to Environment *I* except that we assume more absorption for the larger scattering cluster. In this case, beamforming is expected to be suboptimal starting at a lower SNR level than that in Environment *I*. Indeed, the optimal power allocation outperforms both beamforming and equal power allocation in terms of the mutual information as seen in Figure 5 for the BPSK constellation. For 4-PAM and Gaussian inputs, we again observe that the mutual information

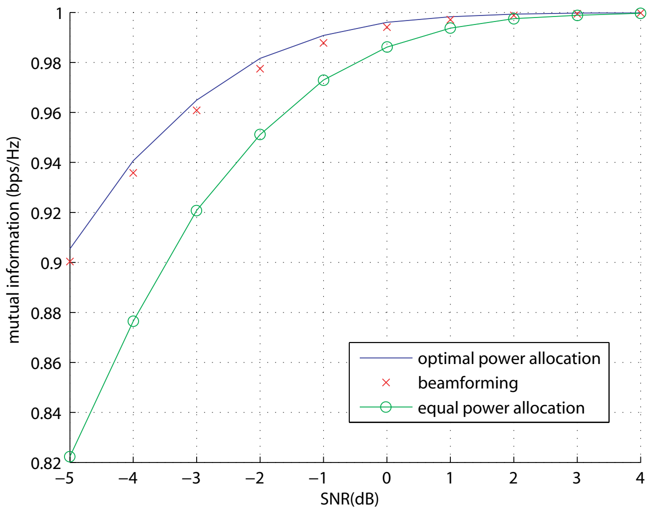

Fig. 5. Comparison of the mutual information using optimal power allocation, beamforming, and equal power allocation for scattering environment II using BPSK.

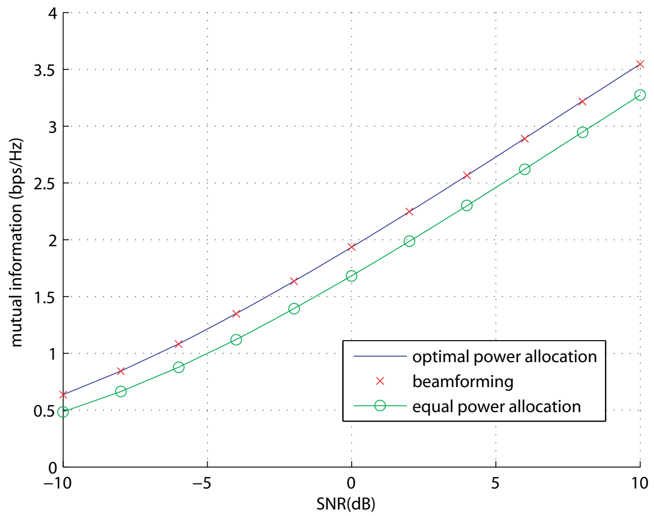

Fig. 7. Comparison of the mutual information using optimal power allocation, beamforming, and equal power allocation for scattering environment II using Gaussian inputs.

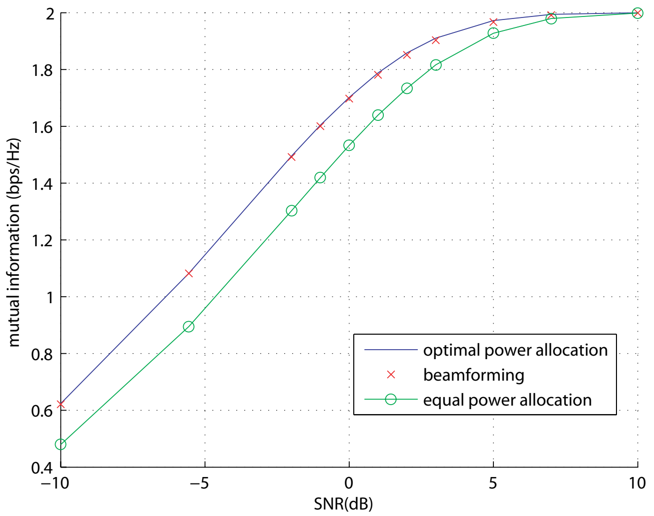

Fig. 6. Comparison of the mutual information using optimal power allocation, beamforming, and equal power allocation for scattering environment II using 4-PAM.

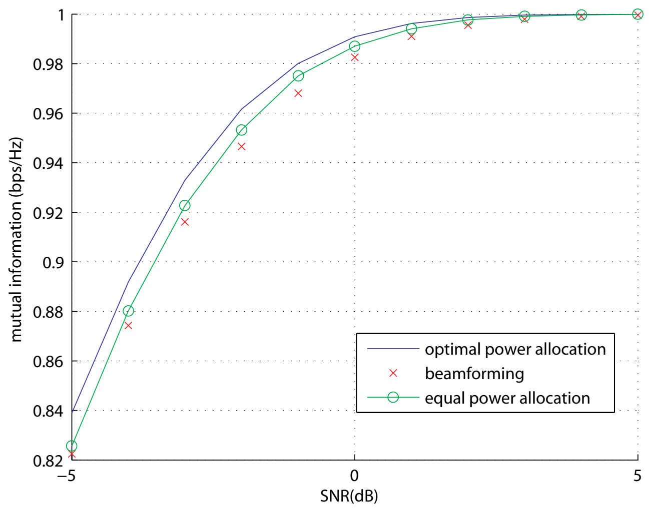

Fig. 8. Comparison of the mutual information using optimal power allocation, beamforming, and equal power allocation for scattering environment III using BPSK.

achieved by the optimal power allocation and beamforming is much larger than that achieved by equal power allocation (see Figure 6 and 7). The performance differences between optimal power allocation and beamforming, however, become smaller as the constellation size increases.

C. Scattering Environment III

In this scattering environment, the sums of each column of the variance matrix

$$V = \frac{25}{5.05}\begin{pmatrix} 1 & 0 & 0.21 & 0 & 0 \\ 0 & 1 & 0.21 & 0 & 0 \\ 0 & 0 & 0.21 & 0 & 0 \\ 0 & 0 & 0.21 & 0.5 & 0 \\ 0 & 0 & 0.21 & 0.5 & 1 \end{pmatrix}$$

are all fairly close to one. Therefore, equal power allocation is expected to be close to optimal. However, as observed in Figure 8-10, the performance gain of using equal power allocation over beamforming is not significant. With increasing constellation sizes, the differences become even more negligible. Although optimal power allocations outperform equal power allocations and beamformings in all cases, it is also tempting to simply use beamforming since the performance of beamforming is close to that of the optimal power allocation and equal power allocation.

Based on the observations for Scattering Environments I to III, we therefore conclude the following: In spatially correlated MIMO channels, beamforming is an attractive candidate for the transmission of symbols with large constellation sizes.

VI. Conclusions

We have considered the design of LD codes that maximize the mutual information in spatially correlated MIMO channels. By exploiting the virtual representation, we derived an upperbound on the mutual information for scattering environments with general fading statistics that satisfy the

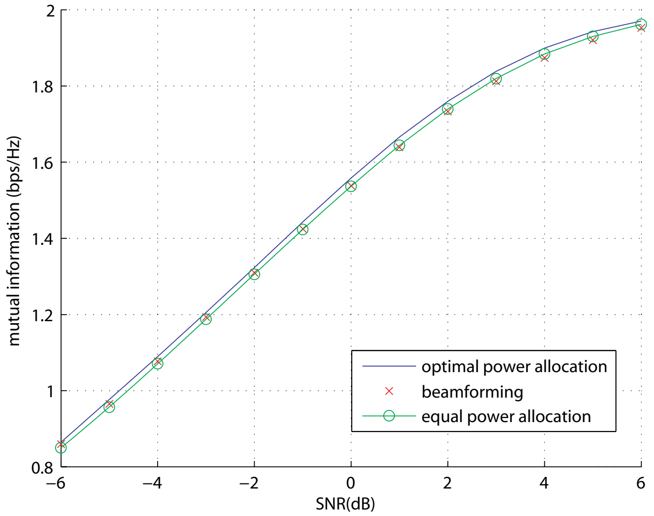

Fig. 9. Comparison of the mutual information using optimal power allocation, beamforming, and equal power allocation for scattering environment III using 4-PAM.

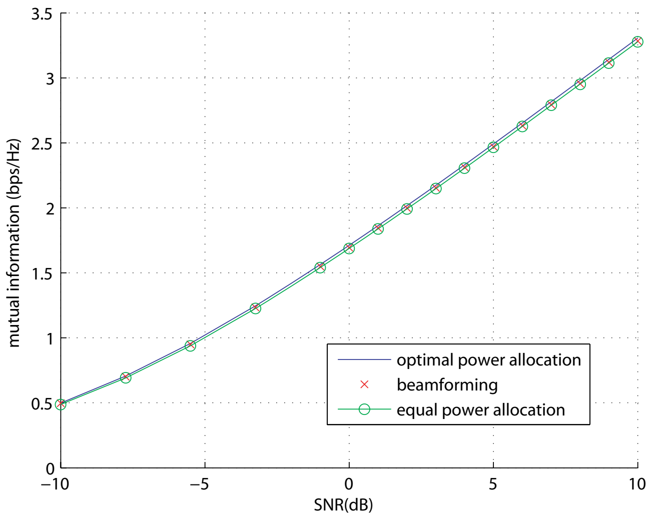

Fig. 10. Comparison of the mutual information using optimal power allocation, beamforming, and equal power allocation for scattering environment III using Gaussian inputs.

standard uncorrelated scattering assumption. Furthermore, we showed that if one can find dispersion matrices satisfying both GOC's, then the corresponding LD code achieves the upperbound and is hence optimal under the mutual information criterion. Equivalently, the channel decouples into K parallel subchannels, where the optimal power allocation is used for each of the subchannels. We provided an explicit construction for the optimal LD codes and discussed the effect of the design parameters K and T. In particular, we showed that it is necessary to minimize T and maximize K while keeping the dispersion matrices satisfying both GOC's . Finally, we presented numerical simulations for three scattering environments. Our results indicate that the mutual information achieved by the optimal power allocation for each subchannel significantly outperforms equal power allocation (which corresponds to ignoring the correlation of the fading channel). Therefore, designing space-time codes base on channel correlation is essential and advantageous.

Appendix I
Proof of Theorem 2

From the proof of Lemma 2 in [11], if we define $\varphi(a) = h(\sqrt{a}\boldsymbol{x} + \tilde{\boldsymbol{n}} | \tilde{\boldsymbol{H}} = \tilde{H})$, where $\tilde{\boldsymbol{n}}$ is a real zero-mean Gaussian random variable of variance $1/2$, then

$$\begin{aligned} &\arg\max_{A} I\left(\boldsymbol{x}; \sqrt{\frac{\Gamma}{n_t}}\boldsymbol{x}\tilde{\boldsymbol{H}}\tilde{A} + \tilde{\boldsymbol{W}} \,\middle|\, \tilde{\boldsymbol{H}}\right) \\ &\quad = \arg\max_{A} \mathrm{E}\left[\varphi\left(\frac{\Gamma}{n_t}\|\boldsymbol{v}\|^2\right)\right] \end{aligned}$$

where $\boldsymbol{v} = \mathrm{vec}(\tilde{\boldsymbol{H}}\tilde{A})$ and $\mathrm{vec}(M)$ is defined as the column vector obtained by stacking the columns of matrix M. Note that conditioned on $\tilde{\boldsymbol{H}} = \tilde{H}$, a is merely a constant. Therefore, $\varphi(a)$ is simply a deterministic, scalar function. Moreover, $\varphi(a)$ is a concave function of a (See Appendix III for detail.)

In the following proof, we use techniques similar to those introduced in the proof of Theorem 1 in [12]. However, the function φ in our expression complicates the proof. Moreover, the fact that φ is concave rather than strictly concave weakens our result. The uniqueness of the optimal diagonal matrix Λ^o is hence lost. Let $Q = \tilde{A}\tilde{A}^\dagger$ and let

$$\begin{aligned} I(Q) &\triangleq \mathrm{E}\left[\varphi\left(\frac{\Gamma}{n_t}\|\boldsymbol{v}\|^2\right)\right] \\ &= \mathrm{E}\left[\varphi\left(\frac{\Gamma}{n_t}\mathrm{Tr}\{\tilde{\boldsymbol{H}}\tilde{A}\tilde{A}^\dagger\tilde{\boldsymbol{H}}^\dagger\}\right)\right] \\ &= \mathrm{E}\left[\varphi\left(\frac{\Gamma}{n_t}\mathrm{Tr}\{\tilde{\boldsymbol{H}}Q\tilde{\boldsymbol{H}}^\dagger\}\right)\right] \end{aligned}$$

Denote

$$\begin{aligned} \Omega : &= \{Q : Q \geq 0, \mathrm{Tr}\{Q\} \leq \alpha n_t T\} \\ \Omega' : &= \{\Lambda \in \Omega : \Lambda \text{ is diagonal}\} \end{aligned}$$

where $Q \geq 0$ indicates that Q is positive semidefinite. Since Ω' is convex and compact and $I(Q)$ is differentiable and concave over Ω', if there exists a Λ^o such that $I(Q)$ is maximized over Ω', Λ^o satisfies the necessary and sufficient condition

$$\delta I(\Lambda^o; \Lambda - \Lambda^o) \leq 0, \forall \Lambda \in \Omega'$$

where the directional derivative $\delta I(\Lambda^o; \Lambda - \Lambda^o)$ can be expressed as

$$\begin{aligned} &\delta I(\Lambda^o; \Lambda - \Lambda^o) \\ &= \frac{d}{dx} I(\Lambda^o + x(\Lambda - \Lambda^o))|_{x=0} \\ &= \frac{d}{dx}\mathrm{E}\left[\varphi\left(\frac{\Gamma}{n_t}\mathrm{Tr}\{\tilde{\boldsymbol{H}}(\Lambda^o + x(\Lambda - \Lambda^o))\tilde{\boldsymbol{H}}^\dagger\}\right)\right]\Bigg|_{x=0} \\ &= \mathrm{E}\Bigg[\Phi\left(\frac{\Gamma}{n_t}\mathrm{Tr}\{\tilde{\boldsymbol{H}}\Lambda^o\tilde{\boldsymbol{H}}^\dagger\}\right) \\ &\qquad\qquad \frac{\Gamma}{n_t}\mathrm{Tr}\{\tilde{\boldsymbol{H}}(\Lambda - \Lambda^o)\tilde{\boldsymbol{H}}^\dagger\}\Bigg] \leq 0 \end{aligned} \tag{12}$$

where

$$\Phi\left(\frac{\Gamma}{n_t}\mathrm{Tr}\{\tilde{\boldsymbol{H}}\Lambda^o\tilde{\boldsymbol{H}}^\dagger\}\right) = \frac{d}{dy}\varphi(y)\Big|_{y=\frac{\Gamma}{n_t}\mathrm{Tr}\{\tilde{\boldsymbol{H}}\Lambda^o\tilde{\boldsymbol{H}}^\dagger\}}$$

Since $I(Q)$ is concave on the convex set Ω, it is sufficient to show that $\delta I(\Lambda^o; Q - \Lambda^o) \leq 0, \forall Q \in \Omega$ to establish that Λ^o

remains optimal on Ω. If we split Q into the diagonal Λ_Q and off-diagonal term F as

$$Q = \Lambda_Q + F$$

Then

$$\begin{aligned}\delta I(\Lambda^o; Q-\Lambda^o) = \\ \mathrm{E}\left[\Phi\left(\frac{\Gamma}{n_t}\mathrm{Tr}\{\tilde{\boldsymbol{H}}\Lambda^o\tilde{\boldsymbol{H}}^\dagger\}\right)\frac{\Gamma}{n_t}\mathrm{Tr}\{\tilde{\boldsymbol{H}}(\Lambda_Q-\Lambda^o)\tilde{\boldsymbol{H}}^\dagger\}\right] \\ + \mathrm{E}\left[\Phi\left(\frac{\Gamma}{n_t}\mathrm{Tr}\{\tilde{\boldsymbol{H}}\Lambda^o\tilde{\boldsymbol{H}}^\dagger\}\right)\frac{\Gamma}{n_t}\mathrm{Tr}\{\tilde{\boldsymbol{H}}F\tilde{\boldsymbol{H}}^\dagger\}\right] \quad (13)\end{aligned}$$

The first term on RHS is smaller than zero by (12). The second term can be expressed as

$$\begin{aligned}&\mathrm{E}\left[\Phi\left(\frac{\Gamma}{n_t}\mathrm{Tr}\{\tilde{\boldsymbol{H}}\Lambda^o\tilde{\boldsymbol{H}}^\dagger\}\right)\frac{\Gamma}{n_t}\mathrm{Tr}\{\tilde{\boldsymbol{H}}F\tilde{\boldsymbol{H}}^\dagger\}\right] \\ &= \mathrm{E}\left[\Phi\left(\frac{\Gamma}{n_t}\sum_{i=1}^{n_t}\lambda_i^o\mathrm{Tr}\{\underline{\tilde{\boldsymbol{h}}}_i\underline{\tilde{\boldsymbol{h}}}_i^\dagger\}\right)\frac{\Gamma}{n_t}\mathrm{Tr}\{\sum_{k,l=1,k\neq l}^{n_t}F_{k,\ell}\underline{\tilde{\boldsymbol{h}}}_k\underline{\tilde{\boldsymbol{h}}}_\ell^\dagger\}\right] \\ &= \sum_{k,l=1,k\neq l}^{n_t}\mathrm{E}\left[\Phi\left(\frac{\Gamma}{n_t}\sum_{i=1}^{n_t}\lambda_i^o\mathrm{Tr}\{\underline{\tilde{\boldsymbol{h}}}_i\underline{\tilde{\boldsymbol{h}}}_i^\dagger\}\right)\frac{\Gamma}{n_t}\mathrm{Tr}\{F_{k,\ell}\underline{\tilde{\boldsymbol{h}}}_k\underline{\tilde{\boldsymbol{h}}}_\ell^\dagger\}\right]\end{aligned}$$

A particular term (say $k=1, \ell=2$) in the above sum can be written as

$$\begin{aligned}&\mathrm{E}\left[\Phi\left(\frac{\Gamma}{n_t}\sum_{i=1}^{n_t}\lambda_i^o\mathrm{Tr}\{\underline{\tilde{\boldsymbol{h}}}_i\underline{\tilde{\boldsymbol{h}}}_i^\dagger\}\right)\frac{\Gamma}{n_t}\mathrm{Tr}\{F_{1,2}\underline{\tilde{\boldsymbol{h}}}_1\underline{\tilde{\boldsymbol{h}}}_2^\dagger\}\right] \\ &= \mathrm{E}\left[\mathrm{E}\left[\Phi\left(\frac{\Gamma}{n_t}\sum_{i=1}^{n_t}\lambda_i^o\mathrm{Tr}\{\underline{\tilde{\boldsymbol{h}}}_i\underline{\tilde{\boldsymbol{h}}}_i^\dagger\}\right)\frac{\Gamma}{n_t}\mathrm{Tr}\{F_{1,2}\underline{\tilde{\boldsymbol{h}}}_1\underline{\tilde{\boldsymbol{h}}}_2^\dagger\}\middle|\underline{\tilde{\boldsymbol{h}}}_2,\ldots,\underline{\tilde{\boldsymbol{h}}}_{n_t}\right]\right]\end{aligned}$$

RHS is an odd function of $\underline{\tilde{\boldsymbol{h}}}_1$. By *Property* 1, $\underline{\tilde{\boldsymbol{h}}}_1$ has the same distribution as $-\underline{\tilde{\boldsymbol{h}}}_1$ and $\underline{\tilde{\boldsymbol{h}}}_1$ is independent of $\underline{\tilde{\boldsymbol{h}}}_2,\ldots,\underline{\tilde{\boldsymbol{h}}}_{n_t}$. Therefore,

$$\mathrm{E}\left[\mathrm{E}\left[\Phi\left(\frac{\Gamma}{n_t}\sum_{i=1}^{n_t}\lambda_i^o\mathrm{Tr}\{\underline{\tilde{\boldsymbol{h}}}_i\underline{\tilde{\boldsymbol{h}}}_i^\dagger\}\right)\frac{\Gamma}{n_t}\mathrm{Tr}\{F_{1,2}\underline{\tilde{\boldsymbol{h}}}_1\underline{\tilde{\boldsymbol{h}}}_2^\dagger\}\middle|\underline{\tilde{\boldsymbol{h}}}_2,\ldots,\underline{\tilde{\boldsymbol{h}}}_{n_t}\right]\right] = 0$$

This implies that the second term of (13) is zero. We can thus conclude that

$$\delta I(\Lambda^o; Q-\Lambda^o) \le 0, \forall Q \in \Lambda$$

The optimal $\tilde{A}_o\tilde{A}_o^\dagger$ is indeed diagonal. □

Appendix II
Proof of Theorem 4

We begin the proof by parameterizing Λ as

$$\Lambda = \mathrm{diag}\{\alpha n_t T - p, p\beta_2, \ldots, p\beta_{n_t}\}$$

where $0 \le p \le \alpha n_t T$ and

$$\beta_i \ge 0, \text{ for } 2 \le i \le n_t; \text{ and } \sum_{i=2}^{n_t}\beta_i \le 1$$

Writing the mutual information as a function of p gives

$$\begin{aligned}I(p) &= \mathrm{E}\left[\varphi\left(\frac{\Gamma}{n_t}\mathrm{Tr}\{\tilde{\boldsymbol{H}}\Lambda\tilde{\boldsymbol{H}}^\dagger\}\right)\right] \\ &= \mathrm{E}\left[\varphi\left(\frac{\Gamma}{n_t}(\alpha n_t T - p)\sum_{k=1}^{n_r}|\tilde{\boldsymbol{H}}_{k,1}|^2 + \frac{\Gamma}{n_t}\sum_{j=2}^{n_t}p\beta_j\sum_{k=1}^{n_r}|\tilde{\boldsymbol{H}}_{k,j}|^2\right)\right] \\ &= \mathrm{E}\left[\varphi\left(\frac{\Gamma}{n_t}(\alpha n_t T - p)\chi_1 + \frac{\Gamma}{n_t}\sum_{j=2}^{n_t}p\beta_j\chi_j\right)\right]\end{aligned}$$

Since $\varphi(a)$ is a concave function of a, $I(p)$ is also a concave function of p. As a result, $p=0$ maximizes $I(p)$ if and only if

$$\left.\frac{\partial I(p)}{\partial p}\right|_{p=0} \le 0$$

Evaluating the derivative of the mutual information gives

$$\begin{aligned}&\left.\frac{\partial I(p)}{\partial p}\right|_{p=0} \\ &= \mathrm{E}\left[\varphi'\left(\frac{\Gamma}{n_t}(\alpha n_t T)\chi_1\right)\frac{\Gamma}{n_t}\left(-\chi_1 + \sum_{j=2}^{n_t}\beta_j\chi_j\right)\right] \\ &= \frac{\Gamma}{n_t}\mathrm{E}\left[\varphi'(\Gamma\alpha T\chi_1)(-\chi_1)\right] + \frac{\Gamma}{n_t}\mathrm{E}\left[\varphi'(\Gamma\alpha T\chi_1)\sum_{j=2}^{n_t}\beta_j\chi_j\right] \le 0\end{aligned}$$

which is equivalent to

$$\mathrm{E}\left[\varphi'(\Gamma\alpha T\chi_1)\right]\sum_{j=2}^{n_t}\beta_j\sum_{k=1}^{n_r}V_{k,j} - \mathrm{E}\left[\varphi'(\Gamma\alpha T\chi_1)\chi_1\right] \le 0$$

and is a result of the independence of $\{\chi_j\}$. The first term is maximized when $\beta_{\ell^o} = 1$, where $\ell^o = \arg\max_{1\le j\le n_t, j\neq i}\sum_{k=1}^{n_r}V_{k,j}$. Therefore, the necessary and sufficient condition for beamforming in the first transmit virtual angle to be optimal is given by

$$\mathrm{E}\left[\varphi'(\Gamma\alpha T\chi_1)\right]\sum_{k=1}^{n_r}V_{k,\ell^o} - \mathrm{E}\left[\varphi'(\Gamma\alpha T\chi_1)\chi_1\right] \le 0$$

□

Appendix III
Concavity of φ Function

For a scalar Gaussian channel of the form

$$\tilde{\boldsymbol{y}} = \sqrt{a}\sqrt{\frac{\Gamma}{n_t}}\boldsymbol{x} + \tilde{\boldsymbol{n}}$$

It can be shown from [16] that for fixed $\sqrt{\frac{\Gamma}{n_t}}$,

$$\frac{dI(a)}{da} = \frac{\Gamma}{2n_t}\text{mmse}(a)$$

Moreover,

$$\begin{aligned} I(a) &= I\left(\boldsymbol{x}; \sqrt{a}\sqrt{\frac{\Gamma}{n_t}}\boldsymbol{x} + \tilde{\boldsymbol{n}} \,\Big|\, \tilde{\boldsymbol{H}} = \tilde{H}\right) \\ &= h\left(\sqrt{a}\sqrt{\frac{\Gamma}{n_t}}\boldsymbol{x} + \tilde{\boldsymbol{n}} \,\Big|\, \tilde{\boldsymbol{H}} = \tilde{H}\right) - h\left(\tilde{\boldsymbol{n}}\right) \end{aligned}$$

Therefore,

$$\frac{dI(a)}{da} = \frac{d\varphi(a)}{da} = \frac{\Gamma}{2n_t}\text{mmse}(a)$$

where $\varphi(a) = h(\sqrt{a}\sqrt{\frac{\Gamma}{n_t}}\boldsymbol{x} + \tilde{\boldsymbol{n}}|\tilde{\boldsymbol{H}} = \tilde{\boldsymbol{H}})$. Since $\text{mmse}(a)$ is a decreasing function of a, we can conclude that $\varphi(a)$ is a concave function of a.

References

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

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

[3] R. W. Heath and A. J. Paulraj, "Linear dispersion codes for MIMO systems based on frame theory," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2429–2441, Oct. 2002.

[4] G. V. Rangaraj, D. Jalihal, and K. Giridhar, "Analysis of space-time coding in correlated fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2882–2891, Nov. 2005.

[5] 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. Signal Processing*, vol. 54, no. 4, pp. 1242–1255, Apr. 2006.

[6] R. H. Gohary and T. N. Davidson, "Design of linear dispersion codes: asymptotic guidelines and their implementation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2892–2906, Nov. 2005.

[7] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.

[8] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Trans. Inf. Theory*, vol. 52, pp. 670–678, 2004.

[9] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 1804–1824, July 2002.

[10] Y. Jiang, "An information theoretic study on linear dispersion codes and low-density parity-check codes," Ph.d. thesis, UIUC, 2005.

[11] G. Bresler and B. Hajek, "Note on mutual information and orthogonal space-time codes," in *Proc. IEEE International Symposium on Information Theory*, pp. 1315–1318, July 2006.

[12] V. V. Veeravalli, Y. Liang, and A. M. Sayeed, "Correlated MIMO wireless channels: capacity, optimal signaling, and asymptotics," *IEEE Trans. Inf. Theory*, vol. 20, pp. 2–16, June 2005.

[13] A. M. Sayeed, "Deconstructing multi-antenna fading channels," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[14] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *Electron. Lett.*, vol. 35, pp. 14–16, Jan. 1999.

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

[16] D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 1261–1282, Apr. 2005.

[17] D. J. Love and R. W. Heath, "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.

[18] S. Zhou and B. Li, "BER criterion and codebook construction for finite-rate precoded spatial multiplexing with linear receivers," *IEEE Trans. Signal Processing*, vol. 54, no. 5, May 2006.

[19] E. N. Onggosanusi and A. G. Dabak, "A feedback-based adaptive multi-input multi-output signaling scheme," in *Proc. Asilomar Conf.*, Nov. 2002.

[20] Y. Ermoliev and R. J-B Wets (eds.), "Numerical techniques for stochastic optimization," Springer-Verlag, 1988.

[21] G. Ch. Pflug, *Optimization of Stochastic Models*. Kluwer Academic Publishers, 1996.

[22] L'Ecuyer, "Efficiency improvement via variance reduction," in *Proc. Winter Simulation Conference*, pp. 122–132, 1994.

Che Lin Received the B.Eng. Degree in Electrical Engineering from National Taiwan Univ., Taipei, Taiwan, in 1999. He received the M.S. degree in Electrical and Computer Engineering from Univ. of Illinois at Urbana-Champaign, IL, in 2001, and is currently pursing the Ph.D. degree at the same university. His research interests include limited feedback systems, space-time code design, MIMO channel modeling, and information 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 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 currently on the Board of Governors of the IEEE Information Theory Society. 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).