Precoding by Pairing Subchannels to Increase MIMO Capacity With Discrete Input Alphabets

Saif Khan Mohammed, *Member, IEEE*, Emanuele Viterbo, *Fellow, IEEE*, Yi Hong, *Senior Member, IEEE*, and Ananthanarayanan Chockalingam, *Senior Member, IEEE*

***Abstract*—We consider Gaussian multiple-input multiple-output (MIMO) channels with discrete input alphabets. We propose a nondiagonal precoder based on the X-Codes in [1] to increase the mutual information. The MIMO channel is transformed into a set of parallel subchannels using singular value decomposition (SVD) and X-Codes are then used to pair the subchannels. X-Codes are fully characterized by the pairings and a 2×2 real rotation matrix for each pair (parameterized with a single angle). This precoding structure enables us to express the total mutual information as a sum of the mutual information of all the pairs. The problem of finding the optimal precoder with the above structure, which maximizes the total mutual information, is solved by: i) optimizing the rotation angle and the power allocation within each pair and ii) finding the optimal pairing and power allocation among the pairs. It is shown that the mutual information achieved with the proposed pairing scheme is very close to that achieved with the optimal precoder by Cruz *et al.*, and is significantly better than Mercury/waterfilling strategy by Lozano *et al.* Our approach greatly simplifies both the precoder optimization and the detection complexity, making it suitable for practical applications.**

***Index Terms*—Condition number, multiple-input multiple-output (MIMO), mutual information, orthogonal frequency division multiplexing (OFDM), precoding, singular value decomposition (SVD).**

I. Introduction

MANY modern communication channels are modeled as a Gaussian multiple-input multiple-output (MIMO) channel. Examples include multitone digital subscriber line (DSL), orthogonal frequency division multiplexing (OFDM), and multiple transmit-receive antenna systems. It is known that the capacity of the Gaussian MIMO channel is achieved by beamforming a Gaussian input alphabet along the right singular vectors of the MIMO channel. The received vector is projected along the left singular vectors, resulting in a set of parallel Gaussian subchannels. Optimal power allocation between the subchannels is achieved by waterfilling [2]. In practice, the input alphabet is *not Gaussian* and is generally chosen from a finite signal set.

We distinguish between two kinds of MIMO channels: i) *diagonal* (or parallel) channels and ii) *nondiagonal* channels.

For a diagonal MIMO channel with discrete input alphabets, assuming only power allocation on each subchannel (i.e., a diagonal precoder), Mercury/waterfilling was shown to be optimal by Lozano *et al.* in [3]. With discrete input alphabets, Cruz *et al.* later proved in [4] that the optimal precoder is, however, nondiagonal, i.e., precoding needs to be performed across all the subchannels.

For a general nondiagonal Gaussian MIMO channel, it was also shown in [4] that the optimal precoder is nondiagonal. Such an optimal precoder is given by a fixed point equation, which requires a high complexity numeric evaluation. Since the precoder jointly codes all the n inputs, joint decoding is also required at the receiver. Thus, the decoding complexity can be very high, specially for large n, as in the case of DSL and OFDM applications. This motivates our quest for a practical low complexity precoding scheme achieving near optimal capacity.

In this paper, we consider a general MIMO channel and a nondiagonal precoder based on X-Codes [1]. The MIMO channel is transformed into a set of parallel subchannels using singular value decomposition (SVD) and X-Codes are then used to pair the subchannels. X-Codes are fully characterized by the pairings and the 2-dimensional real rotation matrices for each pair. These rotation matrices are parameterized with a single angle. This precoding structure enables us to express the total mutual information as a sum of the mutual information of all the pairs.

The problem of finding the optimal precoder with the above structure, which maximizes the total mutual information, can be split into two tractable problems: i) optimizing the rotation angle and the power allocation within each pair and ii) finding the optimal pairing and power allocation among the pairs. It is shown by simulation that the mutual information achieved with the proposed pairing scheme is very close to that achieved with the optimal precoder in [4], and is significantly better than the Mercury/waterfilling strategy in [3]. Our approach greatly simplifies both the precoder optimization and the detection complexity, making it suitable for practical applications.

Manuscript received February 18, 2010; revised October 28, 2010; accepted January 03, 2011. Date of current version June 22, 2011. S. K. Mohammed was supported in part by the Italian Ministry of University and Research (MIUR) with the collaborative research program: Bando per Borse A Favore Di Giovani Ricercatori Indiani (A.F. 2008) and in part by the Swedish Foundation for Strategic Research (SSF) and ELLIIT. A. Chockalingam and S. K. Mohammed were supported in part by the DRDO-IISc Program on Advanced Research in Mathematical Engineering. S. K. Mohammed, E. Viterbo, and Y. Hong performed this work while at DEIS, University of Calabria, Rende (CS), Italy. This work was presented in part at the IEEE International Symposium on Information Theory, Austin, TX, June 2010.

S. K. Mohammed is with the Department of Electrical Engineering (ISY), Linköping University, 581 83 Linköping, Sweden (e-mail: saif@isy.liu.se).

E. Viterbo and Y. Hong are with the Department of Electrical and Computer Systems Engineering, Monash University at Clayton, Melbourne, Victoria 3800, Australia (e-mail: emanuele.viterbo@monash.edu; yi.hong@monash.edu; yi.winnie.hong@gmail.com).

A. Chockalingam is with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012, India (e-mail: achockal@ece.iisc.ernet.in).

Communicated by L. Zheng, Associate Editor for Communications.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2011.2146050

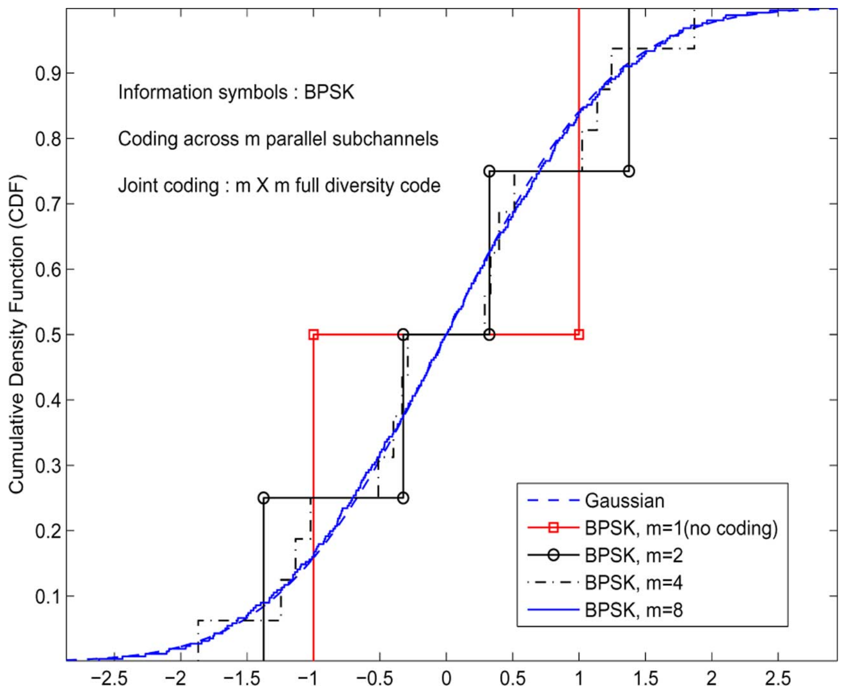

Fig. 1. Plot of the cumulative density function of the coded symbols when joint coding is performed across m subchannels. Input alphabet $\mathcal{S}$ is BPSK.

The rest of the paper is organized as follows. Section II introduces the system model and SVD precoding. In Section III, we provide a brief review of the optimal precoding with discrete inputs in [4] and the relevant MIMO capacity. In Section IV, we present the proposed precoding scheme using X-Codes with discrete inputs and the relevant capacity expressions. In Section V, we consider the first problem, which is to find the optimal rotation angle and power allocation within a given pair. This problem is equivalent to optimizing the mutual information for a Gaussian MIMO channel with two subchannels. In Section VI, using the results from Section V, we attempt to optimize the mutual information for a Gaussian MIMO channel with n subchannels, where $n > 2$. Finally, in Section VII we discuss the application of our precoding scheme to OFDM systems. Conclusions are drawn in Section VIII.

Notations: The field of complex numbers is denoted by $\mathbb{C}$ and let $\mathbb{R}^+$ be the set of positive real numbers. Superscripts ^{T} and † denote transposition and Hermitian transposition, respectively. The $n \times n$ identity matrix is denoted by $\mathbf{I}_n$, and the zero matrix is denoted by $\mathbf{0}$. The $\mathbb{E}[\cdot]$ is the expectation operator, $\|\cdot\|$ denotes the Euclidean norm of a vector, and $\|\cdot\|_F$ the Frobenius norm of a matrix. Finally, we let $\mathrm{tr}(\cdot)$ denote the trace of a matrix.

II. System Model and Precoding With Gaussian Inputs

We consider a $n_t \times n_r$ MIMO channel, where the channel state information (CSI) is known perfectly at both transmitter and receiver. Let $\mathbf{x} = (x_1, \ldots, x_{n_t})^T$ be the vector of input symbols to the channel, and let $\mathbf{H} = \{h_{ij}\}$, $i = 1, \ldots, n_r$, $j = 1, \ldots, n_t$, be a full rank $n_r \times n_t$ channel coefficient matrix, with h_{ij} representing the complex channel gain between the jth input and the ith output. The vector of n_r channel output symbols is given by

$$\mathbf{y} = \sqrt{P_T}\mathbf{H}\mathbf{x} + \mathbf{w} \tag{1}$$

where $\mathbf{w}$ is an uncorrelated Gaussian noise vector, such that $\mathbb{E}[\mathbf{w}\mathbf{w}^\dagger] = \mathbf{I}_{n_r}$, and P_T is the total transmitted power. The power constraint is given by

$$\mathbb{E}[\|\mathbf{x}\|^2] = 1. \tag{2}$$

The maximum multiplexing gain of this channel is $n = \min(n_r, n_t)$. Let $\mathbf{u} = (u_1, \ldots, u_n)^T \in \mathbb{C}^n$ be the vector of n information symbols to be sent through the MIMO channel, with $\mathbb{E}[|u_i|^2] = 1$, $i = 1, \ldots, n$. Then the vector $\mathbf{u}$ can be precoded using a $n_t \times n$ matrix $\mathbf{T}$, resulting in $\mathbf{x} = \mathbf{T}\mathbf{u}$.

The capacity of the deterministic Gaussian MIMO channel is then achieved by solving

Problem 1:

$$\begin{aligned} C(\mathbf{H}, P_T) &= \max_{\mathbf{K_x} \mid \mathrm{tr}(\mathbf{K_x})=1} I(\mathbf{x}; \mathbf{y}|\mathbf{H}) \\ &\geq \max_{\mathbf{K_u}, \mathbf{T} \mid \mathrm{tr}(\mathbf{T}\mathbf{K_u}\mathbf{T}^\dagger)=1} I(\mathbf{u}; \mathbf{y}|\mathbf{H}) \end{aligned} \tag{3}$$

where $I(\mathbf{x}; \mathbf{y}|\mathbf{H})$ is the mutual information between $\mathbf{x}$ and $\mathbf{y}$, and $\mathbf{K_x} \triangleq \mathbb{E}[\mathbf{x}\mathbf{x}^\dagger]$, $\mathbf{K_u} \triangleq \mathbb{E}[\mathbf{u}\mathbf{u}^\dagger]$ are the covariance matrices of $\mathbf{x}$ and $\mathbf{u}$, respectively. The inequality in (3) follows from the data processing inequality [2].

Let us consider the SVD of the channel $\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}$, where $\mathbf{U} \in \mathbb{C}^{n_r \times n}$, $\mathbf{\Lambda} \in \mathbb{C}^{n \times n}$, $\mathbf{V} \in \mathbb{C}^{n \times n_t}$, $\mathbf{U}^\dagger\mathbf{U} = \mathbf{V}\mathbf{V}^\dagger = \mathbf{I}_n$, and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ with $\lambda_1 \geq \lambda_2, \ldots, \geq \lambda_n \geq 0$.

Telatar showed in [6] that the Gaussian MIMO capacity $C(\mathbf{H}, P_T)$, is achieved when $\mathbf{x}$ is Gaussian distributed and $\mathbf{T}\mathbf{K_x}\mathbf{T}^\dagger$ is diagonal. Diagonal $\mathbf{T}\mathbf{K_x}\mathbf{T}^\dagger$ can be achieved by using the optimal precoder matrix $\mathbf{T} = \mathbf{V}^\dagger\mathbf{P}$, where $\mathbf{P} = \mathrm{diag}(p_1, \ldots, p_n) \in (\mathbb{R}^+)^n$ is the diagonal power allocation matrix such that $\mathrm{tr}(\mathbf{P}\mathbf{P}^\dagger) = 1$. Furthermore, u_i, $i = 1, \ldots, n$, are i.i.d. Gaussian (i.e., no coding is required across the input symbols u_i). With this, the second line of (3) is actually an equality. Also, projecting the received vector $\mathbf{y}$

along the columns of $\mathbf{U}$ is information lossless and transforms the nondiagonal MIMO channel into an equivalent diagonal channel with n noninterfering subchannels. The equivalent diagonal system model is then given by

$$\mathbf{r} \triangleq \mathbf{U}^{\dagger}\mathbf{y} = \sqrt{P_T}\mathbf{\Lambda}\mathbf{P}\mathbf{u} + \tilde{\mathbf{w}} \tag{4}$$

where $\tilde{\mathbf{w}}$ is the equivalent noise vector, having the same statistics as $\mathbf{w}$. The total mutual information is now given by

$$I(\mathbf{x};\mathbf{y}|\mathbf{H}) = \sum_{i=1}^{n} \log_2\left(1 + {\lambda_i}^2 p_i^2 P_T\right). \tag{5}$$

Note that now the mutual information is a function of only the power allocation matrix $\mathbf{P}$, with the constraint $\mathrm{tr}(\mathbf{P}\mathbf{P}^{\dagger}) = 1$. Optimal power allocation is achieved through waterfilling between the n parallel channels of the equivalent system in (4)[2].

III. Optimal Precoding With Discrete Inputs

In real systems, discrete input alphabets are used. Subsequently, we assume that the ith information symbol is given by $u_i \in \mathcal{U}_i$, where $\mathcal{U}_i \subset \mathbb{C}$ is a finite signal set. Let $\mathcal{S} \triangleq \mathcal{U}_1 \times \mathcal{U}_2 \times \cdots \times \mathcal{U}_n$ be the overall input alphabet. The capacity of the Gaussian MIMO channel with discrete input alphabet $\mathcal{S}$ is defined by the following problem.

Problem 2:

$$C_{\mathcal{S}}(\mathbf{H}, P_T) = \max_{\mathbf{T}|\mathbf{u}\in\mathcal{S}, \|\mathbf{T}\|_F=1} I(\mathbf{u};\mathbf{y}|\mathbf{H}). \tag{6}$$

Note that there is no maximization over the pdf of $\mathbf{u}$, since we fix $\mathbf{K}_{\mathbf{u}} = \mathbf{I}_n$. The optimal precoder $\mathbf{T}^*$, which solves Problem 2, is given by the following fixed point equation given in [4]:

$$\mathbf{T}^* = \frac{\mathbf{H}^{\dagger}\mathbf{H}\mathbf{T}^*\mathbf{E}}{\|\mathbf{H}^{\dagger}\mathbf{H}\mathbf{T}^*\mathbf{E}\|_F} \tag{7}$$

where $\mathbf{E}$ is the minimum mean-square error (MMSE) matrix of $\mathbf{u}$ given by

$$\mathbf{E} = \mathbb{E}[(\mathbf{u} - \mathbb{E}[\mathbf{u}|\mathbf{y}])(\mathbf{u} - \mathbb{E}[\mathbf{u}|\mathbf{y}])^{\dagger}]. \tag{8}$$

The optimal precoder is derived using the relation between MMSE and mutual information [7]. We observe that, with discrete input alphabets, it is no longer optimal to beamform along the column vectors of $\mathbf{V}^{\dagger}$ and then use waterfilling on the parallel subchannels. Even when $\mathbf{H}$ is diagonal (parallel noninterfering subchannels), the optimal precoder $\mathbf{T}^*$ is *non diagonal*, and can be computed numerically (using a gradient based method) as discussed in [4]. However, the complexity of computing $\mathbf{T}^*$ is prohibitively high for practical applications, especially when n is large. This problem can be further aggravated if the channel changes frequently.

We propose a suboptimal precoding scheme based on X-Codes [1], which achieves close to the optimal capacity $C_{\mathcal{S}}(\mathbf{H}, P_T)$, at low encoding and decoding complexities. In the proposed precoding scheme, the MIMO channel is first transformed into a set of parallel channels by precoding along the right singular vectors of $\mathbf{H}$ (i.e., columns of $\mathbf{V}$) and projecting the received vector along the left singular vectors of $\mathbf{H}$ (i.e., columns of $\mathbf{U}$). The subchannels are then grouped into pairs of subchannels, with joint coding/decoding within each pair.

As we shall see later in Section VI, simply pairing subchannels can result in significant increase in the mutual information between $\mathbf{u}$ and $\mathbf{y}$. Here we provide some insights and reasoning as to why this is so. It is known that the optimal capacity achieving input distribution (**Problem** 1) is Gaussian [6]. By jointly coding over groups of m subchannels (pairing is a special case with $m = 2$), each coded output symbol can be made to have zero mean, finite variance, and a probability density function (pdf) similar to the Gaussian distribution for the same variance. However, it is not so simple to quantify the closeness of the *discrete* pdf of the coded output symbols to the *continuous* Gaussian pdf. For the purpose of illustration, in Fig. 1, we compare the cumulative density function (cdf) of a real Gaussian random variable with mean 0 and variance 1, with the cdf of a coded output symbol when joint coding is performed across m subchannels. Joint coding is performed using a $m \times m$ real orthogonal matrix as the linear code generator matrix. For the purpose of illustration we have used $m \times m$ algebraic rotation matrices which generate full diversity code [5]. We also note that $m = 1$ corresponds to the case of no coding. The input information symbols are assumed to be BPSK. It is observed from the figure that, with increasing m the cdf of the coded output symbols approaches the Gaussian cdf A simple way of quantifying the closeness is in terms of the maximum absolute difference between the Gaussian cdf and the cdf of the coded output symbol. With such a measure of closeness, we observe that the maximum absolute difference for $m = 1, 2, 4, 8$ is 0.34, 0.17, 0.11, and 0.02, respectively. Therefore it seems that most of the reduction in the maximum absolute difference is when m is increased from 1 to 2 (i.e., by simply coding across a pair of subchannels). Also, further increase in m beyond $m = 2$, results in smaller reduction in the maximum absolute difference. This observation makes us believe that most of the increase in mutual information can be obtained by coding across only a pair of subchannels. Later in Section VI, we shall see that, indeed, coding across a pair of subchannels results in significant increase in mutual information when compared to the scenario where no coding is performed across subchannels.

IV. Precoding With X-Codes

X-Codes are based on a pairing of n subchannels $\ell = \{(i_k, j_k) \in [1, n] \times [1, n], i_k < j_k, k = 1, \ldots n/2\}$. For a given n, there are $(n-1)(n-3)\cdots 3 \cdot 1$ possible pairings. Let $\mathcal{L}$ denote the set of all possible pairings. For example, with $n = 4$, we have

$$\mathcal{L} = \{\{(1,4),(2,3)\}\ \{(1,2),(3,4)\}\ \{(1,3),(2,4)\}\}.$$

X-Codes are generated by a $n \times n$ real orthogonal matrix, denoted by $\mathbf{G}$. When precoding with X-Codes, the precoder matrix is given by $\mathbf{T} = \mathbf{V}^{\dagger}\mathbf{P}\mathbf{G}$, where $\mathbf{P} = \mathrm{diag}(p_1, p_2, \ldots, p_n) \in (\mathbb{R}^+)^n$ is the diagonal power allocation matrix such that $\mathrm{tr}(\mathbf{P}\mathbf{P}^{\dagger}) = 1$. The kth pair consists

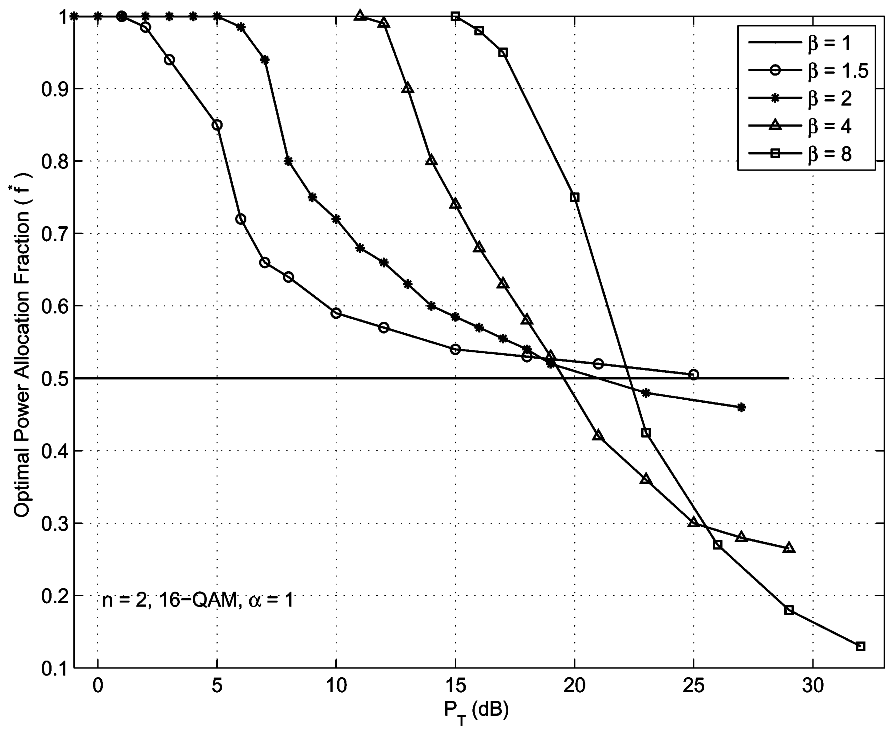

Fig. 2. Plot of f^* versus P_T for $n = 2$ parallel channels with $\beta = 1$, 1.5, 2, 4, 8 and $\alpha = 1$. Input alphabet is 16-QAM.

of subchannels i_k and j_k. For the kth pair, the information symbols u_{i_k} and u_{j_k} are jointly coded using a 2×2 real orthogonal matrix $\mathbf{A}_k$ given by

$$\mathbf{A}_k = \begin{bmatrix} \cos(\theta_k) & \sin(\theta_k) \\ -\sin(\theta_k) & \cos(\theta_k) \end{bmatrix} \quad k = 1, \dots n/2. \tag{9}$$

The angle θ_k can be chosen to maximize the mutual information for the kth pair. Each $\mathbf{A}_k$ is a submatrix of the code generator matrix $\mathbf{G} = (g_{i,j})$ as shown

$$\begin{aligned} g_{i_k,i_k} &= \cos(\theta_k) \qquad g_{i_k,j_k} = \sin(\theta_k) \\ g_{j_k,i_k} &= -\sin(\theta_k) \quad g_{j_k,j_k} = \cos(\theta_k). \end{aligned} \tag{10}$$

It was shown in [1] that, for achieving the best diversity gain, an optimal pairing is one in which the kth subchannel is paired with the $(n-k+1)$th subchannel. For example, with this pairing and $n = 6$, the X-Code generator matrix is given by

$$\mathbf{G} = \begin{bmatrix} \cos(\theta_1) & & & & & \sin(\theta_1) \\ & \cos(\theta_2) & & & \sin(\theta_2) & \\ & & \cos(\theta_3) & \sin(\theta_3) & & \\ & & -\sin(\theta_3) & \cos(\theta_3) & & \\ & -\sin(\theta_2) & & & \cos(\theta_2) & \\ -\sin(\theta_1) & & & & & \cos(\theta_1) \end{bmatrix}.$$

The special case with $\theta_k = 0$, $k = 1, 2, \dots, n/2$, results in no coding across subchannels (i.e., a diagonal precoder).

Given the generator matrix $\mathbf{G}$, the subchannel gains $\mathbf{\Lambda}$, and the power allocation matrix $\mathbf{P}$, the mutual information between $\mathbf{u}$ and $\mathbf{y}$ is given by

$$\begin{aligned} I_{\mathcal{S}}(\mathbf{u};\mathbf{y}|\mathbf{\Lambda},\mathbf{P},\mathbf{G}) &= h(\mathbf{y}|\mathbf{\Lambda},\mathbf{P},\mathbf{G}) - h(\mathbf{w}) \\ &= -\int_{\mathbf{y}\in\mathbb{C}^{n_r}} p(\mathbf{y}|\mathbf{\Lambda},\mathbf{P},\mathbf{G}) \log_2(p(\mathbf{y}|\mathbf{\Lambda},\mathbf{P},\mathbf{G}))d\mathbf{y} - n\log_2(\pi e) \end{aligned} \tag{11}$$

where the received vector pdf is given by

$$p(\mathbf{y}|\mathbf{\Lambda},\mathbf{P},\mathbf{G}) = \frac{1}{|\mathcal{S}|\pi^n} \sum_{\mathbf{u}\in\mathcal{S}} e^{-\|\mathbf{y}-\sqrt{P_T}\mathbf{U\Lambda PGu}\|^2} \tag{12}$$

and when $n = n_r$ (i.e., $n_r \leq n_t$), it is equivalently given by

$$p(\mathbf{y}|\mathbf{\Lambda},\mathbf{P},\mathbf{G}) = \frac{1}{|\mathcal{S}|\pi^n} \sum_{\mathbf{u}\in\mathcal{S}} e^{-\|\mathbf{r}-\sqrt{P_T}\mathbf{\Lambda PGu}\|^2} \tag{13}$$

where $\mathbf{r} = (r_1, r_2, \dots, r_n)^T \triangleq \mathbf{U}^\dagger\mathbf{y}$.

We next define the capacity of the MIMO Gaussian channel when precoding with $\mathbf{G}$. In the following, we assume that $n_r \leq n_t$, so that $I_{\mathcal{S}}(\mathbf{u};\mathbf{y}|\mathbf{\Lambda},\mathbf{P},\mathbf{G}) = I_{\mathcal{S}}(\mathbf{u};\mathbf{r}|\mathbf{\Lambda},\mathbf{P},\mathbf{G})$.[1] Note that, when $n_r > n_t$, the receiver processing $\mathbf{r} = \mathbf{U}^\dagger\mathbf{y}$ becomes information lossy, and $I_{\mathcal{S}}(\mathbf{u};\mathbf{y}|\mathbf{\Lambda},\mathbf{P},\mathbf{G}) > I_{\mathcal{S}}(\mathbf{u};\mathbf{r}|\mathbf{\Lambda},\mathbf{P},\mathbf{G})$.

We introduce the following definitions. For a given pairing ℓ, let $\mathbf{r}_k \triangleq (r_{i_k}, r_{j_k})^T$, $\mathbf{u}_k \triangleq (u_{i_k}, u_{j_k})^T$, $\mathbf{\Lambda}_k \triangleq \mathrm{diag}(\lambda_{i_k}, \lambda_{j_k})$, $\mathbf{P}_k \triangleq \mathrm{diag}(p_{i_k}, p_{j_k})$ and $\mathcal{S}_k \triangleq \mathcal{U}_{i_k} \times \mathcal{U}_{j_k}$. Due to the pairing structure of $\mathbf{G}$ the mutual information $I_{\mathcal{S}}(\mathbf{u};\mathbf{r}|\mathbf{\Lambda},\mathbf{P},\mathbf{G})$ can be expressed as the sum of mutual information of all the $n/2$ pairs as follows:

$$I_{\mathcal{S}}(\mathbf{u};\mathbf{r}|\mathbf{\Lambda},\mathbf{P},\mathbf{G}) = \sum_{k=1}^{n/2} I_{\mathcal{S}_k}(\mathbf{u}_k;\mathbf{r}_k|\mathbf{\Lambda}_k,\mathbf{P}_k,\theta_k). \tag{14}$$

Having fixed the precoder structure to $\mathbf{T} = \mathbf{V}^\dagger\mathbf{PG}$, we can formulate the following:

Problem 3:

$$C_X(\mathbf{H}, P_T) = \max_{\mathbf{G},\mathbf{P}|\mathbf{u}\in\mathcal{S},\mathrm{tr}(\mathbf{PP}^\dagger)=1} I_{\mathcal{S}}(\mathbf{u};\mathbf{r}|\mathbf{\Lambda},\mathbf{P},\mathbf{G}). \tag{15}$$

[1]It is to be noted that, this assumption is made only when precoding with X-Codes, and therefore Problems 1 and 2 *do not* assume $n_r \leq n_t$.

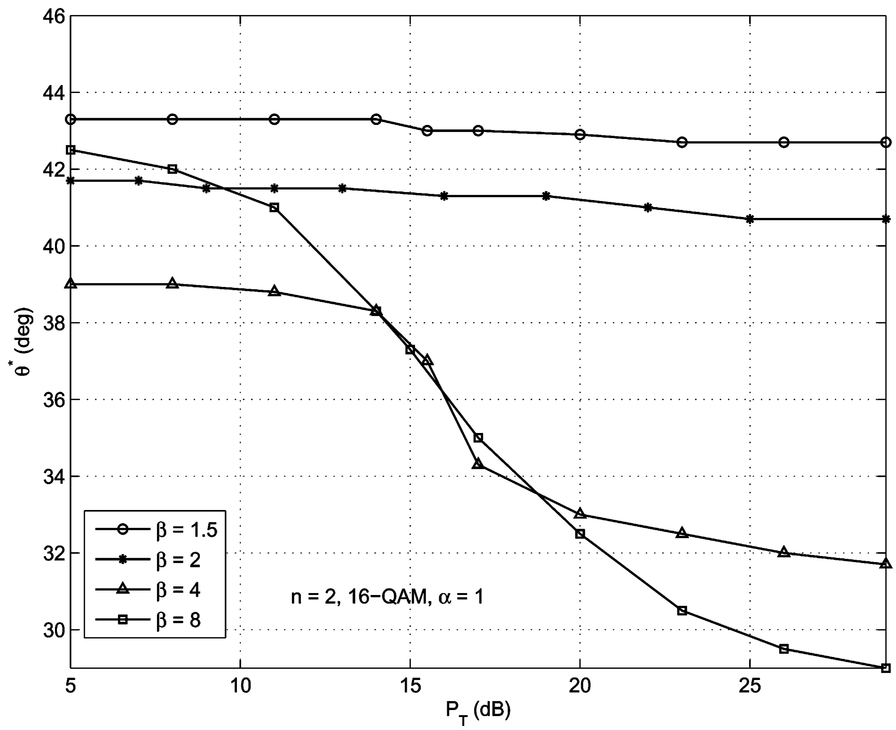

Fig. 3. Plot of θ^* versus P_T for $n = 2$ parallel channels with $\beta = 1.5$, 2, 4, 8 and $\alpha = 1$. Input alphabet is 16-QAM.

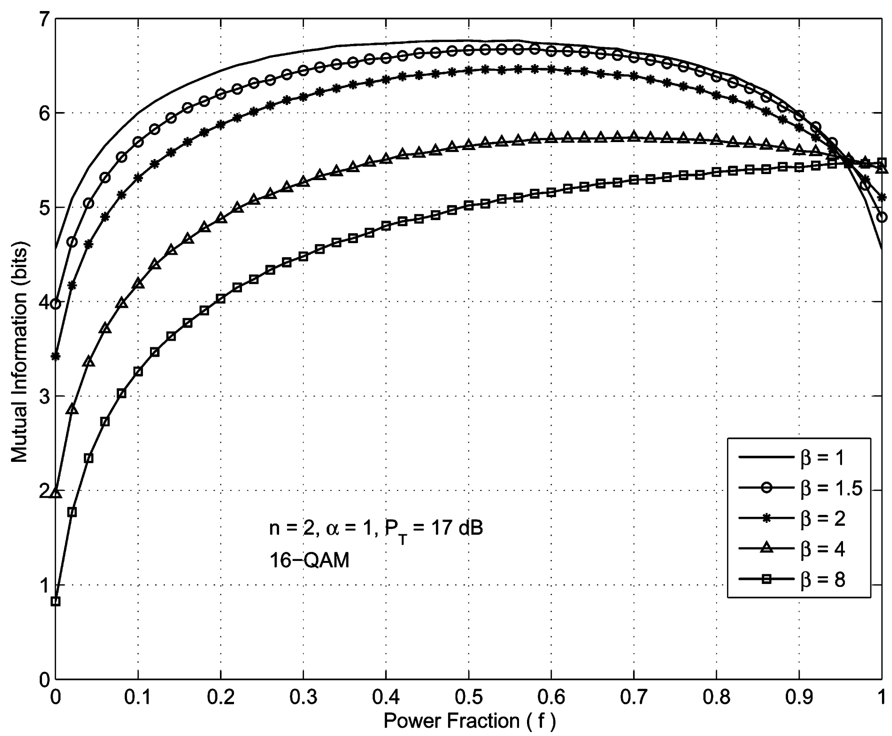

Fig. 4. Mutual Information of X-Codes versus power allocation fraction f for $n = 2$ parallel channels with $\beta = 1$, 1.5, 2, 4, 8, $\alpha = 1$ and $P_T = 17$ dB. Input alphabet is 16-QAM.

It is clear that the solution of the above problem is still a formidable task, although it is simpler than Problem 2. In fact, instead of the $n \times n$ variables of $\mathbf{T}$, we now deal with n variables for power allocation in $\mathbf{P}$, $n/2$ variables for the angles defining $\mathbf{A}_k$, and the pairing $\ell \in \mathcal{L}$. In the following, we will show how to efficiently solve Problem 3 by splitting it into two simpler problems.

Power allocation can be divided into power allocation *among* the $n/2$ pairs, followed by power allocation *between* the two subchannels of each pair.[2] Let $\bar{\mathbf{P}} = \mathrm{diag}(\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_{n/2})$ be a

[2]We draw the attention of the reader to the distinction between the usage of the words "among" and "between." In this paper, we use "among" when referring to more than 2 entities. The word "between" is used when there are exactly 2 entities involved.

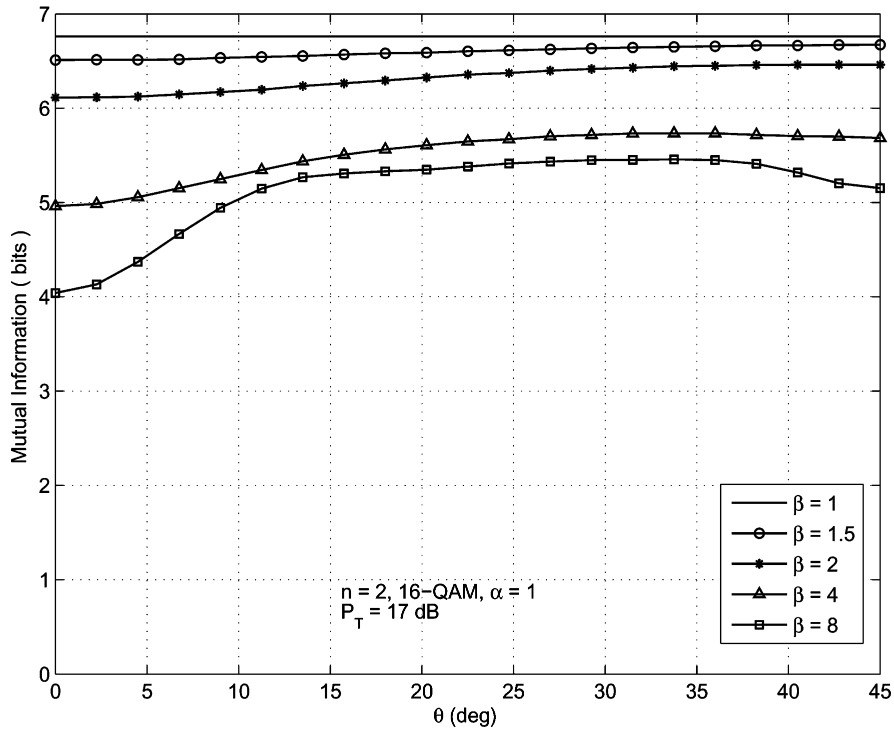

Fig. 5. Mutual information of X-Codes versus rotation angle θ for $n = 2$ parallel channels with $\beta = 1, 1.5, 2, 4, 8$, $\alpha = 1$ and $P_T = 17$ dB. Input alphabet is 16-QAM.

diagonal matrix, where $\bar{p}_k \triangleq \sqrt{p_{i_k}^2 + p_{j_k}^2}$ with $\bar{p}_k^2$ being the power allocated to the kth pair. The power allocation within each pair can be simply expressed in terms of the fraction $f_k \triangleq p_{i_k}^2/\bar{p}_k^2$ of the power assigned to the first subchannel of the pair. The mutual information achieved by the kth pair is then given by

$$I_{\mathcal{S}_k}(\mathbf{u}_k; \mathbf{r}_k | \mathbf{\Lambda}_k, \mathbf{P}_k, \theta_k) = I_{\mathcal{S}_k}(\mathbf{u}_k; \mathbf{r}_k | \mathbf{\Lambda}_k, \bar{p}_k, f_k, \theta_k)$$
$$= -\int_{\mathbf{r}_k \in \mathbb{C}^2} p(\mathbf{r}_k) \log_2 p(\mathbf{r}_k)\, d\mathbf{r}_k - 2\log_2(\pi e) \quad (16)$$

where $p(\mathbf{r}_k)$ is given by

$$p(\mathbf{r}_k) = \frac{1}{|\mathcal{S}_k|\pi^2} \sum_{\mathbf{u}_k \in \mathcal{S}_k} e^{-\|\mathbf{r}_k - \sqrt{P_T}\bar{p}_k \mathbf{\Lambda}_k \mathbf{F}_k \mathbf{A}_k \mathbf{u}_k\|^2} \quad (17)$$

where $\mathbf{F}_k \triangleq \mathrm{diag}(\sqrt{f_k}, \sqrt{1-f_k})$ and $\mathbf{A}_k$ is given by (9).

The capacity of the discrete input MIMO Gaussian channel when precoding with X-Codes can be expressed as

Problem 4:

$$C_X(\mathbf{H}, P_T) = \max_{\ell \in \mathcal{L}, \bar{\mathbf{P}} | \mathrm{tr}(\bar{\mathbf{P}}\bar{\mathbf{P}}^\dagger)=1} \sum_{k=1}^{n/2} C_{\mathcal{S}_k}(k, \ell, \bar{p}_k) \quad (18)$$

where $C_{\mathcal{S}_k}(k, \ell, \bar{p}_k)$, the capacity of the kth pair in the pairing ℓ, is achieved by solving

Problem 5:

$$C_{\mathcal{S}_k}(k, \ell, \bar{p}_k) = \max_{\theta_k, f_k} I_{\mathcal{S}_k}(\mathbf{u}_k; \mathbf{r}_k | \mathbf{\Lambda}_k, \bar{p}_k, f_k, \theta_k). \quad (19)$$

In other words, we have split Problem 3 into two different simpler problems. First, given a pairing ℓ and power allocation *among* the $n/2$ pairs $\bar{\mathbf{P}}$, we can solve Problem 5 for each $k = 1, 2, \ldots, n/2$. Problem 4 uses the solution to Problem 5 to find the optimal pairing ℓ^* and the optimal power allocation $\bar{\mathbf{P}}^*$ *among* the $n/2$ pairs. For small n, the optimal pairing and power allocation *among* the pairs can always be computed numerically and by brute force enumeration of all possible pairings. This is, however, prohibitively complex for large n, and we shall discuss heuristic approaches in Section VI.

We will show in the following that, although suboptimal, precoding with X-Codes will provide a close to optimal capacity with the additional benefit that the detection complexity at the receiver is highly reduced, since there is coupling only *between* pairs of subchannels, as compared to the case of full-coupling for the optimal precoder in [4].

In Section V, we solve Problem 5, which is equivalent to finding the optimal rotation angle and power allocation for a Gaussian MIMO channel with only $n = 2$ subchannels.

V. Gaussian MIMO Channels With $n = 2$

With $n = 2$, there is only one pair and only one possible pairing. Therefore, we drop the subscript k in Problem 5 and we find $C_X(\mathbf{H}, P_T)$ in Problem 3. The processed received vector $\mathbf{r} \in \mathbb{C}^2$ is given by

$$\mathbf{r} = \sqrt{P_T}\mathbf{\Lambda}\mathbf{F}\mathbf{A}\mathbf{u} + \mathbf{z} \quad (20)$$

where $\mathbf{z} = \mathbf{U}^\dagger \mathbf{w}$ is the equivalent noise vector with the same statistics as $\mathbf{w}$. Let $\alpha \triangleq \lambda_1^2 + \lambda_2^2$ be the overall channel power gain and $\beta \triangleq \lambda_1/\lambda_2$ be the *condition number* of the channel. Then (20) can be rewritten as

$$\mathbf{r} = \sqrt{\tilde{P}_T}\tilde{\mathbf{\Lambda}}\mathbf{F}\mathbf{A}\mathbf{u} + \mathbf{z} \quad (21)$$

where $\tilde{P}_T \triangleq P_T\alpha$ and $\tilde{\lambda} = \mathrm{diag}(\tilde{\lambda}_1, \tilde{\lambda}_2) \triangleq \mathbf{\Lambda}/\sqrt{\alpha} = \mathrm{diag}(\beta/\sqrt{1+\beta^2}, 1/\sqrt{1+\beta^2})$. The equivalent channel $\tilde{\mathbf{\Lambda}}$

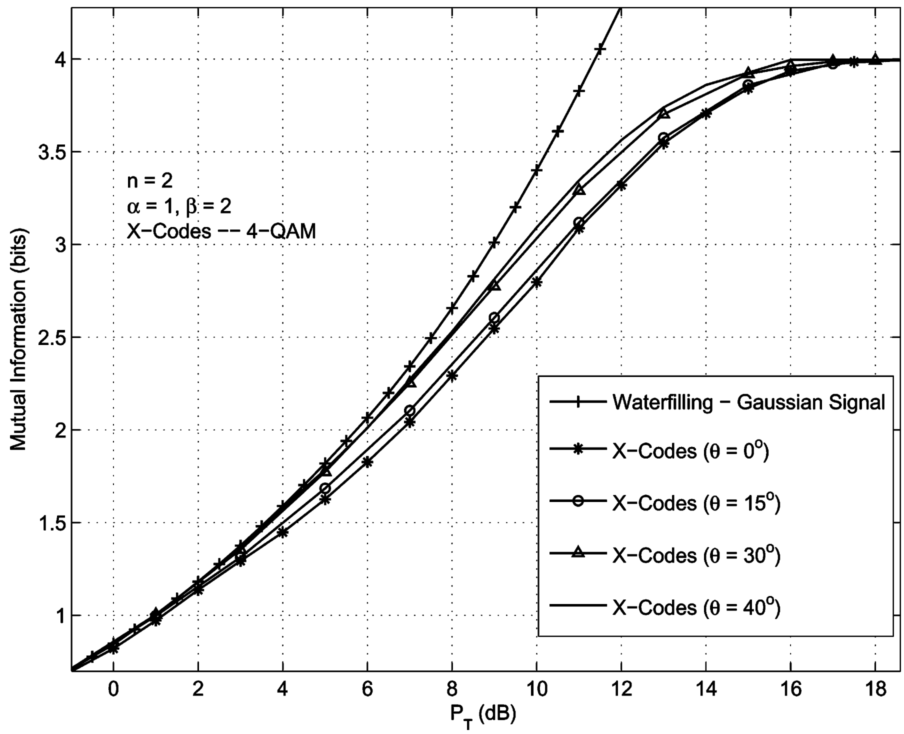

Fig. 6. Mutual information versus P_T for X-Codes for different θs, $n = 2$ parallel channels, $\alpha = 1$, $\beta = 2$, and 4-QAM input alphabet.

now has a normalized gain of $\sqrt{\tilde{\lambda}_1^2 + \tilde{\lambda}_2^2} = 1$, and its subchannel gains $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ are dependent only upon β. Our goal is, therefore, to find the optimal rotation angle θ^* and the fractional power allocation f^*, which maximize the mutual information of the equivalent channel with condition number β and gain $\alpha = 1$. The total available transmit power is now $\tilde{P}_T$.

It is difficult to get analytic expressions for the optimal θ^* and f^*, and therefore we can use numerical techniques to evaluate them and store them in lookup tables to be used at run time. For a given application scenario, given the distribution of β, we decide upon a few discrete values of β which are representative of the actual values observed in real channels. For each such quantized value of β, we numerically compute a table of the optimal values f^* and θ^* as a function of $\tilde{P}_T$. These tables are constructed offline. During the process of communication, the transmitter knows the value of α and β from channel measurements. It then finds the lookup table with the closest value of β to the measured one. The optimal values f^* and θ^* are then found by indexing the appropriate entry in the table with $\tilde{P}_T$ equal to $P_T\alpha$.

In Fig. 2, we graphically plot the optimal power fraction f^* to be allocated to the stronger subchannel in the pair, as a function of P_T. The input alphabet is 16-QAM and $\beta = 1, 1.5, 2, 4, 8$. For $\beta = 1$, both subchannels have equal gains, and therefore, as expected, the optimal power allocation is to divide power equally *between* the two subchannels. However with increasing β, the power allocation becomes more asymmetrical. For a fixed P_T, a higher fraction of the total power is allocated to the stronger subchannel with increasing β.

For a fixed β, it is observed that at low P_T it is optimal to allocate all power to the stronger subchannel. In contrast, at high P_T, it is the weaker subchannel which gets most of the power. In the high P_T regime, these results are in contrast with the waterfilling scheme, where almost all subchannels are allocated equal power. However, a similar observation has also been made for the Mercury/waterfilling scheme [3]. We next present an intuitive explanation for the fact that, at high P_T, it is the weaker subchannel which is allocated a higher fraction of the total power.

The mutual information with a finite input set of cardinality M is limited to $\log_2(M)$ bits and the mutual information curve when plotted w.r.t. P_T flattens out as $P_T \rightarrow \infty$. Therefore, at high P_T there is little incentive to allocate further power to a strong subchannel since its mutual information is already very close to $\log_2(M)$ bits, and being in the "flat" region of the mutual information curve results in very little increase in mutual information for a given increase in P_T. A weak subchannel on the other hand, has a mutual information far from $\log_2(M)$ bits and an appreciable increase in mutual information for a given increase in P_T is possible. Therefore, in terms of the increase in mutual information at high P_T, for a similar increase in P_T, a weak subchannel would benefit more when compared to a strong subchannel.

In Fig. 3, the optimal rotation angle θ^* is plotted as a function of P_T. The input alphabet is 16-QAM and $\beta = 1.5$, 2, 4, 8. For $\beta = 1$ the mutual information is independent of θ for all values of P_T. For $\beta = 1.5$, 2, the optimal rotation angle is almost invariant to P_T. For larger β, the optimal rotation angle varies with P_T and approximately ranges *between* 30–40 ° for all P_T values of interest.

Fig. 4 shows the variation of the mutual information with the power fraction f for $\alpha = 1$. The power P_T is fixed at 17 dB and the input alphabet is 16-QAM. We observe that for all values of β, the mutual information is a concave function of f. We also observe that the sensitivity of the mutual information to variation in f increases with increasing β. However, for all β, the mu-

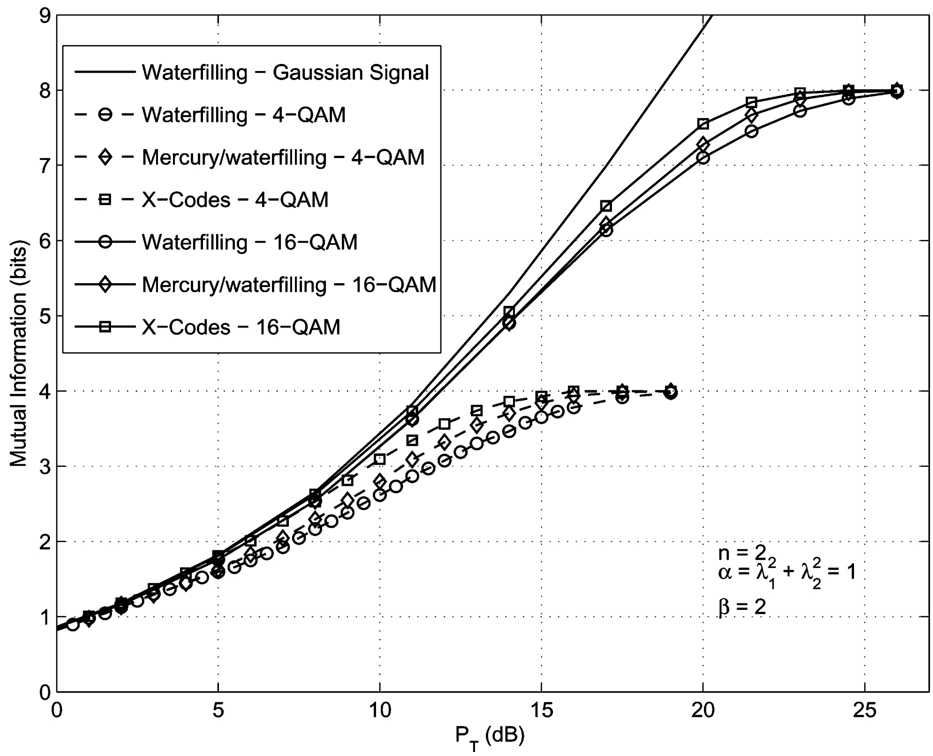

Fig. 7. Mutual information versus P_T for $n = 2$ parallel channels with $\beta = 2$ and $\alpha = 1$, for 4-QAM and 16-QAM.

tual information is fairly stable (has a "plateau") around the optimal power fraction. This is good for practical implementation, since this implies that an error in choosing the correct power allocation would result in a very small loss in the achieved mutual information.

In Fig. 5, we plot the variation of the mutual information w.r.t. the rotation angle θ. The power P_T is fixed at 17 dB and the input alphabet is 16-QAM. For $\beta = 1$, the mutual information is obviously constant with θ. With increasing β, mutual information is observed to be increasingly sensitive to θ. However, when compared with Fig. 4, it can also be seen that the mutual information appears to be more sensitive to the power allocation fraction f, than to θ.

In Fig. 6, we plot the mutual information of X-Codes for different rotation angles with $\alpha = 1$ and $\beta = 2$. For each rotation angle, the power allocation is optimized numerically. We observe that, the mutual information is quite sensitive to the rotation angle except in the range 30–40°.

We next present some simulation results to show that indeed our simple precoding scheme can significantly increase the mutual information, compared to the case of no precoding across subchannels (i.e., Mercury/waterfilling). For the sake of comparison, we also present the mutual information achieved by the waterfilling scheme with discrete input alphabets.

We restrict the discrete input alphabets $\mathcal{U}_i$, $i = 1, 2$, to be square M-QAM alphabets consisting of two $\sqrt{M}$-PAM alphabets in quadrature. Mutual information is evaluated by solving Problem 5 (i.e., numerically maximizing w.r.t. the rotation angle and power allocation).

In Fig. 7, we plot the maximal mutual information versus P_T, for a system with two subchannels, $\beta = 2$ and $\alpha = 1$. Mutual information is plotted for 4- and 16-QAM signal sets. It is observed that for a given achievable mutual information, coding across subchannels is more power efficient. For example, with 4-QAM and an achievable mutual information of 3 bits, X-Codes require only 0.8 dB more transmit power when compared to the ideal Gaussian signalling with waterfilling. This gap increases to 1.9 dB for Mercury/waterfilling and 2.8 dB for the waterfilling scheme with 4-QAM as the input alphabet. A similar trend is observed with 16-QAM as the input alphabet. The proposed precoder clearly performs better, since the mutual information is optimized w.r.t. the rotation angle θ and power allocation, while Mercury/waterfilling, as a special case of X-Code, only optimizes power allocation and fixes $\theta = 0$.

In Fig. 8, we compare the mutual information achieved by X-Codes and the Mercury/waterfilling strategy for $\alpha = 1$ and $\beta = 1, 2, 4$. The input alphabet is 4-QAM. It is observed that both the schemes have the same mutual information when $\beta = 1$. However with increasing β, the mutual information of Mercury/waterfilling strategy is observed to degrade significantly at high P_T, whereas the performance of X-Codes does not vary as much. The degradation of mutual information for the Mercury/waterfilling strategy is explained as follows. For the Mercury/waterfilling strategy, with increasing β, all the available power is allocated to the stronger channel till a certain transmit power threshold. However, since finite signal sets are used, mutual information is bounded from above until the transmit power exceeds this threshold. This also explains the reason for the intermediate change of slope in the mutual information curve with $\beta = 4$ (see the rightmost curve in Fig. 8). On the other hand, due to coding across subchannels, this problem does not arise when precoding with X-Codes. Therefore, in terms of achievable mutual information, rotation coding is observed to be more robust to ill-conditioned channels.

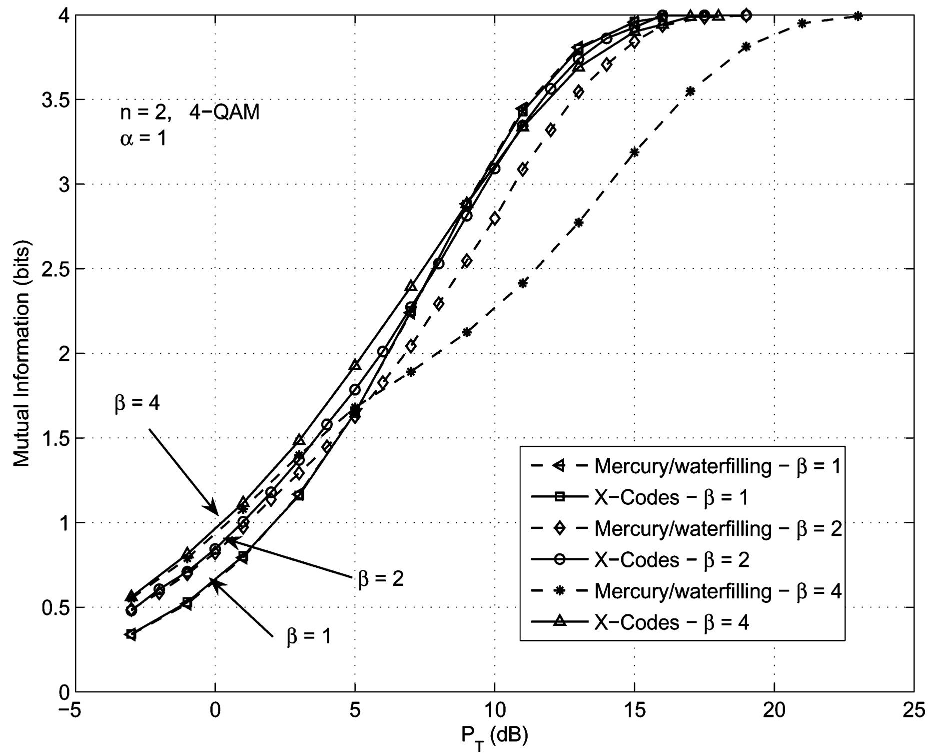

Fig. 8. Mutual information versus P_T for $n = 2$ parallel channels with varying $\beta = 1, 2, 4$, $\alpha = 1$ and 4-QAM input alphabet.

For low values of P_T, mutual information of both the schemes are similar, and improves with increasing β. This is due to the fact that, at low P_T, mutual information increases linearly with P_T, and, therefore, all power is assigned to the stronger channel. With increasing β, the stronger channel has an increasing fraction of the total channel gain, which results in increased mutual information.

In Fig. 9, the mutual information with X-Codes is plotted for $\beta = 1, 2, 4, 8$ and with 16-QAM as the input alphabet. It is observed that at low values of P_T, a higher value of β is favorable. However at high P_T, with 16-QAM input alphabets, the performance degrades with increasing β. This degradation is more significant compared to the degradation observed with 4-QAM input alphabets. Therefore it can be concluded that the mutual information is more sensitive to β with 16-QAM input alphabets as compared to 4-QAM.

VI. Gaussian MIMO Channels With $n > 2$

We now consider the problem of finding the optimal pairing and power allocation *among* the $n/2$ pairs for different Gaussian MIMO channels with even n and $n > 2$. We first observe that mutual information is indeed sensitive to the chosen pairing, and this therefore justifies the criticality of computing the optimal pairing. This is illustrated through Fig. 10, for $n = 4$ with a diagonal channel $\mathbf{\Lambda} = \mathrm{diag}(0.8, 0.4, 0.4, 0.2)$ and 16-QAM. Optimal power allocation *between* the two pairs is computed numerically. It is observed that the pairing $\{(1,4),(2,3)\}$ performs significantly better than the pairing $\{(1,3),(2,4)\}$.

In Fig. 11, we compare the mutual information achieved with optimal precoding [4], to that achieved by the proposed precoder with 4-QAM input alphabet. The 4×4 full channel matrix (non-diagonal channel) is given by [4, eq. (42)] (Gigabit DSL). For X-Codes, the optimal pairing is $\{(1,4),(2,3)\}$ and the optimal power allocation *between* the pairs is computed numerically. It is observed that X-Codes perform very close to the optimal precoding scheme. Specifically, for an achievable mutual information of 6 bits, compared to the optimal precoder [4], X-Codes need only 0.4 dB extra power whereas 2.3 dB extra power is required with Mercury/waterfilling.

Another interesting application is in wireless MIMO channels with perfect channel state information at both the transmitter and receiver. The channel coefficients are modeled as i.i.d complex normal random variables with unit variance.

In Fig. 12, we plot the ergodic capacity (i.e., the mutual information averaged over channel realizations) for a 4×4 wireless MIMO channel. For X-Codes, the best pairing and power allocation *between* pairs are chosen numerically using the optimal θ and power fraction tables created offline. It is observed that at high P_T, simple rotation based coding using X-Codes improves the mutual information significantly, when compared to Mercury/waterfilling. For example, for a target mutual information of 12 bits, X-Codes perform 1.2 dB away from the ideal Gaussian signalling scheme. This gap from the Gaussian signalling scheme increases to 3.1 dB for the Mercury/waterfilling scheme and to 4.4 dB for the waterfilling scheme with 16-QAM alphabets.

In this application scenario the low complexity of our precoding scheme becomes an essential feature, since the precoder can be computed on the fly using the look-up tables for each channel realization.

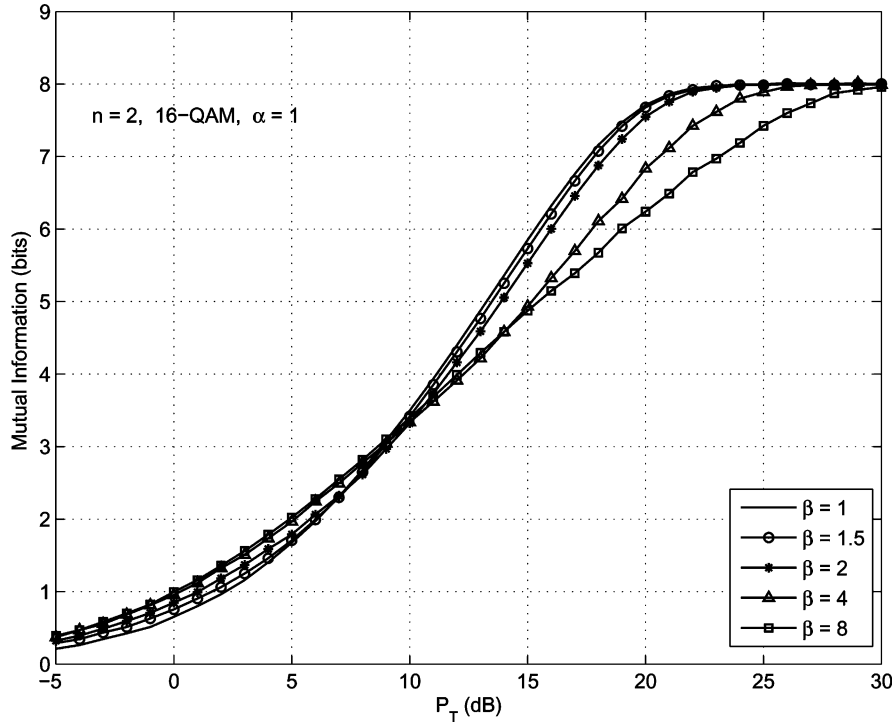

Fig. 9. Mutual information with X-Codes versus P_T for $n = 2$ parallel channels with varying $\beta = 1, 2, 4, 8$, $\alpha = 1$ and 16-QAM input alphabet.

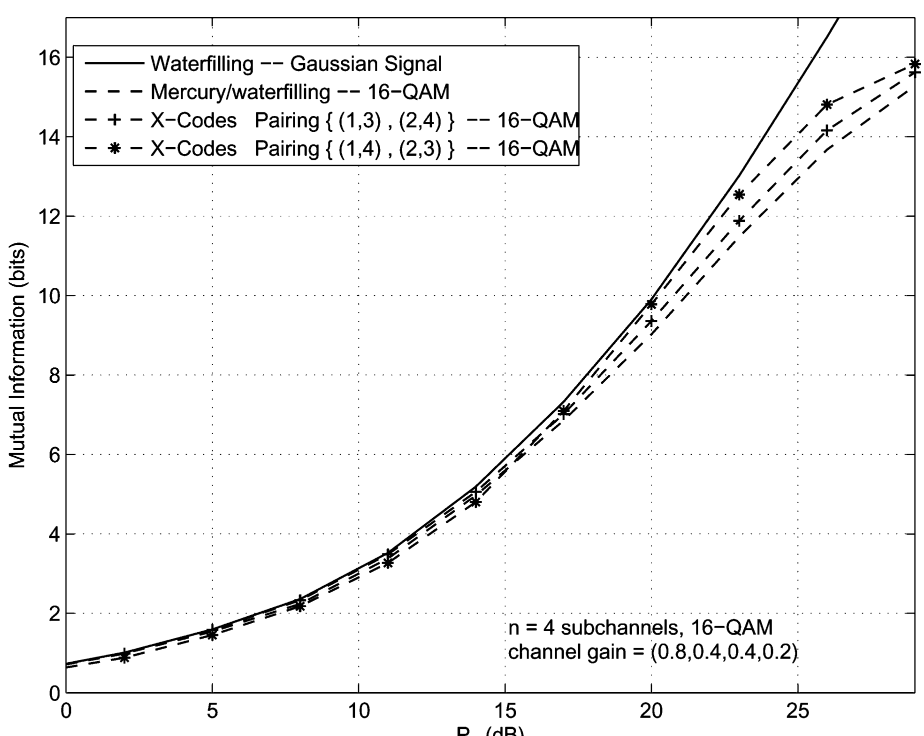

Fig. 10. Mutual information versus P_T with two different pairings for a $n = 4$ diagonal channel and 16-QAM input alphabet.

VII. Application to OFDM

In OFDM applications, n is large and Problem 4 becomes too complex to solve, since we can no more find the optimal pairing by enumeration.

It was observed in Section V, that for $n = 2$, a larger value of the condition number β leads to a higher mutual information at low values of P_T (low SNR). Therefore, we conjecture that pairing the kth subchannel with the $(n/2 + k)$th subchannel could have mutual information very close to optimal, since this pairing scheme attempts to maximize the minimum β *among* all pairs. We shall call this scheme the "conjectured" pairing scheme, and the X-Code scheme, which pairs the kth with the $(n-k+1)$th subchannel, the "X-pairing" scheme. Note

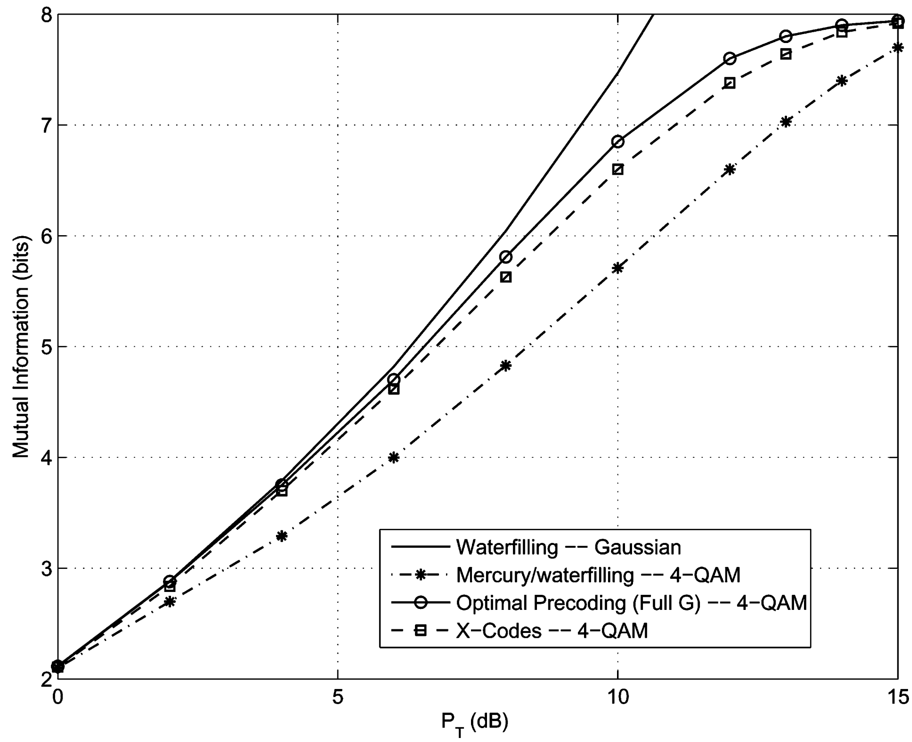

Fig. 11. Mutual information versus P_T for the Gigabit DSL channel given by (42) in [4].

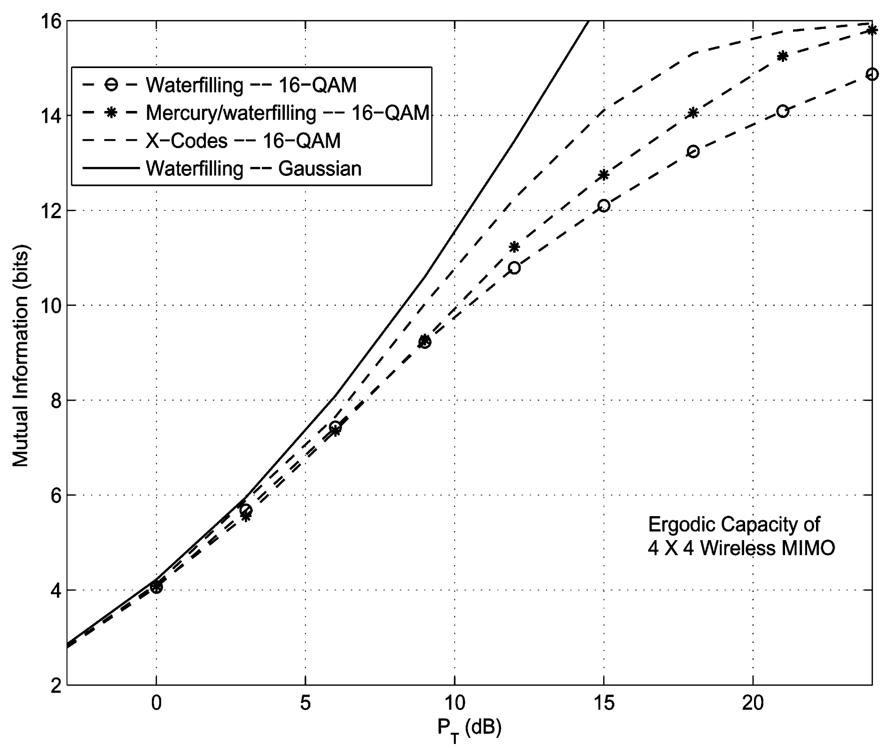

Fig. 12. 4×4 Wireless MIMO: Ergodic capacity vs. finite input precoding schemes.

that the "X-pairing" scheme was proposed in [1] as a scheme which achieved the optimal diversity gain when precoding with X-Codes.

Given a pairing of n subchannels, it is also difficult to compute the optimal power allocation *among* the $n/2$ pairs, $\bar{\mathbf{P}}$. However, it was observed that for channels with large n, taking $\bar{\mathbf{P}}$ to be the waterfilling power allocation *among* the $n/2$ pairs (with $\alpha_k \triangleq \sqrt{\lambda_{i_k}^2 + \lambda_{j_k}^2}$ as the equivalent channel gain of the kth pair) results in good performance.

Apart from the "conjectured" and the "X-pairing" schemes, we propose a pairing scheme which is based on the job assignment problem. The problem consists in matching m different workers to m different jobs that have to be completed. Consider the $m \times m$ cost matrix $\mathbf{C}$, whose (i,j)th entry $C_{i,j}$, is the cost

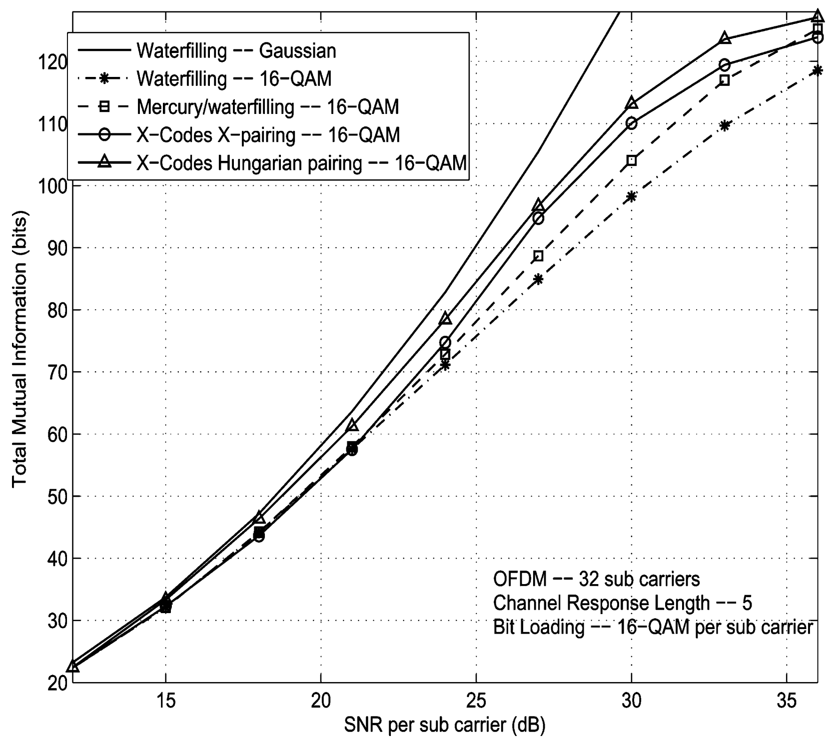

Fig. 13. Mutual information versus per subcarrier SNR for an OFDM system with 32 carriers. X-Codes versus Mercury/waterfilling.

involved when the ith worker is assigned to the jth job, $i, j = 1, 2, \ldots m$. The solution to the job assignment problem gives the optimal assignment of worker to jobs (with each worker getting assigned to exactly one job), such that the total cost of getting all the jobs completed is minimized. We call this as the minimization job assignment problem. Another form of the job assignment problem is where the total cost of getting all the jobs completed must be maximized, and we shall refer to this as the maximization job assignment problem. It is easy to see, that a maximization job assignment problem could be posed in terms of an equivalent minimization job assignment problem and vice versa.

The job assignment problem is efficiently solved using the Hungarian algorithm [8]. In this paper, we pose our problem of finding a good approximation to the optimal pairing as a job assignment problem and solve it using the Hungarian algorithm. We shall therefore refer to this pairing as the "Hungarian" pairing scheme. To find a good approximation to the optimal pairing, we split the n subchannels into two groups: i) Group-I: subchannels 1 to $n/2$, with the jth subchannel in the role of the jth job $(j = 1, 2, \ldots n/2)$; ii) Group-II: subchannels $n/2+1$ to n, with the $(n/2+i)$th subchannel in the role of the ith worker $(i = 1, 2, \ldots n/2)$. Therefore, there are $n/2$ workers and jobs.

For a given SNR (P_T), we initially assume uniform power allocation *among* the $n/2$ pairs and therefore assign a power of $2P_T/n$ to each pair. The value of $C_{i,j}$ is evaluated by finding the optimal mutual information achieved by an equivalent $n = 2$ channel with the $n/2+i$th and the jth subchannels as its two subchannels. This can be obtained by first choosing a table (see Section V) with the closest value of β to the given $\lambda_j/\lambda_{n/2+i}$, and then indexing the appropriate entry into the table with $\mathrm{SNR} = 2P_T(\lambda_j^2 + \lambda_{n/2+i}^2)/n$. The Hungarian algorithm then finds the pairing with the highest mutual information. Furthermore, the computational complexity of the Hungarian algorithm is $O(n^3)$, which is practically tractable. Power allocation *among* the $n/2$ pairs is then achieved through the waterfilling scheme.

To study the sensitivity of the mutual information to the pairing of subchannels, we also consider a "Random" pairing scheme. In the "Random" pairing scheme, we first choose a large number (≈ 50) of random pairings. For each chosen random pairing we evaluate the mutual information (through Monte Carlo simulations) with waterfilling power allocation *among* the $n/2$ pairs. Finally the average mutual information is computed. This gives us insight into the mean value of the mutual information w.r.t. pairing. It would also help us in assessing if the heuristic pairing schemes discussed above are worth pursuing.

We next illustrate the mutual information achieved by these heuristic schemes for an OFDM system with $n = 32$ subchannels and 16-QAM. The channel impulse response is $[-0.454+\mathrm{j}0.145, -0.258+\mathrm{j}0.198, 0.0783+\mathrm{j}0.069, -0.408-\mathrm{j}0.396, -0.532 - \mathrm{j}0.224\,]$. For the "conjectured" and the "X-pairing" schemes also, power allocation is achieved through waterfilling *among* the 16 pairs.

In Fig. 13 the total mutual information is plotted as a function of the SNR per sub carrier. It is observed that the proposed precoding scheme performs much better than the Mercury/waterfilling scheme. The proposed precoder with the "Hungarian" pairing scheme performs within 1.1 dB of the Gaussian signalling scheme for an achievable total mutual information of 96 bits (i.e., a rate of $96/128 = 3/4$). The proposed precoder with the "Hungarian" pairing scheme performs about 1.6 dB better than the Mercury/waterfilling scheme. The "X-pairing" scheme performs better than the Mercury/waterfilling and worse than the "Hungarian" pairing scheme. Even at a lower rate of 1/2

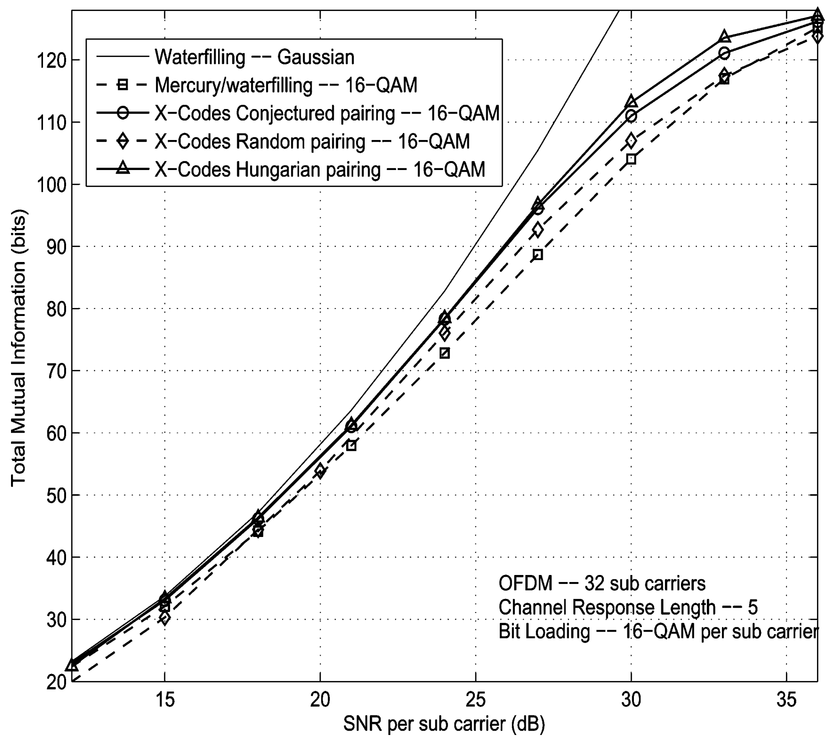

Fig. 14. Mutual information versus per subcarrier SNR for an OFDM system with 32 carriers. Comparison of heuristic pairing schemes.

(i.e., a total mutual information of 64 bits), the proposed precoder with the "Hungarian" pairing scheme performs about 0.7 dB better than the Mercury/waterfilling scheme.

In Fig. 14, we compare the mutual information achieved by the various heuristic pairing schemes. It is observed that the "conjectured" pairing scheme performs very close to the "Hungarian" pairing scheme except at very high SNR. For example, even for a high mutual information of 96 bits, the "Hungarian" pairing scheme performs better than the "conjectured" pairing scheme by only about 0.2 dB. However at very high rates (like 7/8 and above), the "Hungarian" pairing scheme is observed to perform better than the "conjectured" pairing scheme by about 0.7 dB. Therefore for low to medium rates, it would be better to use the "conjectured" pairing since it has the same performance at a lower computational complexity. The mutual information achieved by the "Random" pairing scheme is observed to be strictly inferior than the "conjectured" pairing scheme at all values of SNR, and at low SNR it is even worse than the Mercury/waterfilling strategy. This, therefore implies that the total mutual information is indeed sensitive to the chosen pairing. Further, till a rate of 1/2 (i.e., a mutual information of 64 bits) it appears that any extra optimization effort would not result in significant performance improvement for the "conjectured" pairing scheme, since it is already very close to the ideal Gaussian signalling schemes. However at higher rate and SNR it may still be possible to improve the mutual information by further optimizing the selection of pairing scheme and power allocation *among* the pairs. This is however a difficult problem that requires further investigation.

VIII. Conclusions

In this paper, we proposed a *low complexity* precoding scheme based on the pairing of subchannels, which achieves near optimal capacity for Gaussian MIMO channels with discrete inputs. The low complexity feature relates to both the evaluation of the optimal precoder matrix and the detection at the receiver. This makes the proposed scheme suitable for practical applications, even when the channels are time varying and the precoder needs to be computed for each channel realization.

The simple precoder structure, inspired by the X-Codes, enabled us to split the precoder optimization problem into two simpler problems. First, for a given pairing and power allocation *among* the pairs, we need to find the optimal power fraction allocation and rotation angle for each pair. Given the solution to the first problem, the second problem is then to find the optimal pairing and power allocation *among* the pairs.

For large n, typical of OFDM systems, we also discussed different heuristic approaches for optimizing the pairing of subchannels.

The proposed precoder was shown to perform better than the Mercury/waterfilling strategy for both diagonal and nondiagonal MIMO channels. Future work will focus on finding close to optimal pairings, and close to optimal power allocation strategies *among* the pairs.

Acknowledgment

S. K. Mohammed, E. Viterbo, and Y. Hong initialized this work while at DEIS, University of Calabria, Italy.

References

[1] S. K. Mohammed, E. Viterbo, Y. Hong, and A. Chockalingam, "MIMO precoding with X- and Y-codes," *IEEE Trans. Inf. Theory*, Nov. 2010.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

[3] A. Lozano, A. M. Tulino, and S. Verdu, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3033–3051, Jul. 2006.

[4] F. P. Cruz, M. R. D. Rodrigues, and S. Verdu, "MIMO Gaussian channels with arbitrary inputs: Optimal precoding and power allocation," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1070–1084, Mar. 2010.

[5] E. Bayer-Fluckiger, F. Oggier, and E. Viterbo, "Algebraic lattice constellations: bounds on performance," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 319–327, Jan. 2006.

[6] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[7] D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[8] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Res. Logist. Quart.*, vol. 2, pp. 83–97, 1955.

Saif Khan Mohammed (M'11) received the B.Tech. degree in computer science and engineering from the Indian Institute of Technology, New Delhi, in 1998, and the Ph.D. degree in electrical and communication engineering from the Indian Institute of Science, Bangalore, in December 2010.

From 1998 to 2000, he was with Philips, Inc., Bangalore, as an ASIC Design Engineer. From 2000 to 2003, he was with Ishoni Networks, Inc., Santa Clara, CA, as a Senior Chip Architecture Engineer. From 2003 to 2007, he was with Texas Instruments, Bangalore, as a Systems and Algorithms Designer in the Wireless Systems Group. His research interests also include low-complexity detection, estimation, and coding for wireless communications systems. Since October 2010, he is a Postdoctoral Researcher with the Communication Systems Division of the Electrical Engineering Department (ISY), University of Linköping, Sweden. His current research activities are primarily in the area of large multiple-input multiple-output (MIMO) systems.

Dr. Mohammed, during his Ph.D. research, was awarded the Young Indian Researcher Fellowship by the Italian Ministry of University and Research (MIUR).

Emanuele Viterbo (F'11) was born in Torino, Italy, in 1966. He received the Laurea degree in electrical engineering in 1989 and the Ph.D. degree in 1995 in electrical engineering, both from the Politecnico di Torino, Torino, Italy.

From 1990 to 1992, he was with the European Patent Office, The Hague, The Netherlands, as a patent examiner in the field of dynamic recording and error-control coding. Between 1995 and 1997, he held a Postdoctoral position with the Dipartimento di Elettronica, Politecnico di Torino. During 1997–1998, he was a Postdoctoral Research Fellow with the Information Sciences Research Center of AT&T Research, Florham Park, NJ. He became first Assistant Professor (1998) then Associate Professor (2005) with the Dipartimento di Elettronica at Politecnico di Torino. In 2006, he became a Full Professor with the DEIS at the University of Calabria, Italy. Since September 2010 he has been a Full Professor with the ECSE Department, Monash University, Melbourne, Australia. In 1993, he was a Visiting Researcher with the Communications Department of DLR, Oberpfaffenhofen, Germany. In 1994 and 1995, he was visiting the Ecole Nationale Supérieure des Télécommunications (E.N.S.T.), Paris. In 2003, he was a Visiting Researcher with the Maths Department, EPFL, Lausanne, Switzerland. In 2004, he was a Visiting Researcher with the Telecommunications Department, UNICAMP, Campinas, Brazil. In 2005, 2006, and 2009, he was Visiting Researcher with the ITR of UniSA, Adelaide, Australia. In 2007, he was Visiting Fellow with the Nokia Research Center, Helsinki, Finland. His main research interests are in lattice codes for the Gaussian and fading channels, algebraic coding theory, algebraic space-time coding, digital terrestrial television broadcasting, digital magnetic recording, and irregular sampling.

Prof. Emanuele Viterbo is an ISI Highly Cited Researcher and a member of Board of Governor (Information Theory Society). He was an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY, European *Transactions on Telecommunications*, and the *Journal of Communications and Networks*, and Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING: SPECIAL ISSUE MANAGING COMPLEXITY IN MULTIUSER MIMO SYSTEMS. He was awarded a NATO Advanced Fellowship in 1997 from the Italian National Research Council.

Yi Hong (SM'10) received the Ph.D. degree in electrical engineering and telecommunications from the University of New South Wales (UNSW), Sydney, Australia, in October 2004.

She is currently a lecturer with the Department of Electrical and Computer Systems Engineering, Monash University, Melbourne, Australia. She then joined the Institute of Telecom. Research, University of South Australia, Australia; also the Institute of Advanced Telecom, Swansea University, U.K.; and the University of Calabria, Italy. Her research interests include information and communication theory with applications to telecommunication engineering.

Dr. Hong, during her Ph.D. degree research, received an International Postgraduate Research Scholarship (IPRS) from the Commonwealth of Australia; a supplementary Engineering Award from the School of Electrical Engineering and Telecommunications, UNSW; and a Wireless Data Communication System Scholarship from UNSW. She received the NICTA-ACoRN Earlier Career Researcher award for a paper presented at the Australian Communication Theory Workshop (AUSCTW), Adelaide, Australia, in 2007. She is a Technical Program Committee Chair of AUSCTW'11, Melbourne. She was the Publicity Chair at the IEEE Information Theory Workshop 2009, Sicily, Italy. She is a Technical Program Committee member for many IEEE conferences such as IEEE ICC 2011, VTC 2011, PIMRC, and WCNC 2008. She is a member of ACoRN.

Ananthanarayanan Chockalingam (SM'98) received the B.E. (Honours) degree in electronics and communication engineering from the P. S. G. College of Technology, Coimbatore, India, in 1984, the M.Tech. degree with specialization in satellite communications from the Indian Institute of Technology, Kharagpur, in 1985, and the Ph.D. degree in electrical communication engineering (ECE) from the Indian Institute of Science (IISc), Bangalore, in 1993.

During 1986 to 1993, he was with the Transmission R & D Division, Indian Telephone Industries Limited, Bangalore. From December 1993 to May 1996, he was a Postdoctoral Fellow and an Assistant Project Scientist with the Department of Electrical and Computer Engineering, University of California, San Diego. From May 1996 to December 1998, he was with Qualcomm, Inc., San Diego, CA, as a Staff Engineer/Manager in the systems engineering group. In December 1998, he joined the faculty of the Department of ECE, IISc, Bangalore, where he is a Professor, working in the area of wireless communications and networking.

Dr. Chockalingam is a recipient of the Swarnajayanti Fellowship from the Department of Science and Technology, Government of India. He served as an Associate Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY from May 2003 to April 2007. He currently serves as an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He served as a Guest Editor for the IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS: SPECIAL ISSUE ON MULTIUSER DETECTION FOR ADVANCED COMMUNICATION SYSTEMS AND NETWORKS. Currently, he serves as a Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING: SPECIAL ISSUE ON SOFT DETECTION ON WIRELESS TRANSMISSION. He is a Fellow of the Institution of Electronics and Telecommunication Engineers and a Fellow of the Indian National Academy of Engineering.