MIMO Precoding With X- and Y-Codes

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

Abstract—We consider a slow fading multiple-input multiple-output (MIMO) system with channel state information at both the transmitter and receiver. A well-known precoding scheme is based upon the singular value decomposition (SVD) of the channel matrix, which transforms the MIMO channel into parallel subchannels. Despite having low maximum likelihood decoding (MLD) complexity, this SVD precoding scheme provides a diversity gain which is limited by the diversity gain of the weakest subchannel. We therefore propose X- and Y-Codes, which improve the diversity gain of the SVD precoding scheme but maintain the low MLD complexity, by jointly coding information across a pair of subchannels. In particular, subchannels with high diversity gain are paired with those having low diversity gain. A pair of subchannels is jointly encoded using a 2×2 real matrix, which is fixed *a priori* and does not change with each channel realization. For X-Codes, these rotation matrices are parameterized by a single angle, while for Y-Codes, these matrices are left triangular matrices. Moreover, we propose X-, Y-Precoders with the same structure as X-, Y-Codes, but with encoding matrices adapted to each channel realization. We observed that X-Codes/Precoders are good for well-conditioned channels, while Y-Codes/Precoders are good for ill-conditioned channels.

Index Terms—Condition number, diversity, error probability, MIMO, precoding, singular value decomposition.

I. INTRODUCTION

W E consider slow fading $n_t \times n_r$ multiple-input multiple-output (MIMO) systems, where channel state information (CSI) is fully available both at transmitter and receiver. Channels in such systems are subject to block fading,

and therefore, reliability is a major concern. It is known that precoding techniques can provide large performance improvements in such scenarios by enhancing the communication reliability, which is typically quantified in terms of the diversity gain achieved by the precoding scheme.

Some state of the art precoding techniques are discussed next. The most straightforward precoding approach is based on direct channel inversion and also known as zero-forcing (ZF) precoding [4]. However, it suffers from a loss of power efficiency. Nonlinear precoding such as Tomlinson-Harashima (TH) precoding [5], [6] was exploited in [7]. Linear precoders, which involve simple linear pre- and post-processing, have been proposed in [8], [9] and references therein. Despite having low encoding and decoding complexity, the linear precoding schemes and the TH precoder have low diversity gain. Precoders based on lattice reduction techniques [10] and vector perturbation [11] can achieve high diversity gain, but at the cost of high complexity. We therefore see a tradeoff between diversity gain and encoding/decoding complexity. This motivates us to design precoding schemes which for a given rate of transmission (in bits communicated per channel use), achieve high diversity at low encoding/decoding complexities.

In this paper, we consider SVD precoding for MIMO systems, which is based on the SVD decomposition of the channel gain matrix, and which transforms the MIMO channel into parallel subchannels/streams [1], [2]. At the receiver, maximum likelihood decoding (MLD) of the transmitted information symbol vector reduces to separate ML decoding for the information symbol transmitted on each subchannel, thereby resulting in low ML detection complexity. The diversity gain achieved by the SVD precoding scheme is, however, limited by the subchannel with the lowest diversity gain. In some cases, like in Rayleigh fading MIMO channels with $n_r = n_t$, no diversity gain is achieved with this simple precoding scheme.

The diversity gain of a SVD precoded system can be improved by performing joint coding and joint ML detection across a group of subchannels, as with signal space diversity techniques in SISO Rayleigh fading channels, where multidimensional lattice coding is applied to a group of independently fading channel uses [19], [20]. Unfortunately, the complexity of joint ML detection increases exponentially as the number of subchannels which are jointly coded increases. Nevertheless, we show in this paper that we can get large improvements in achievable diversity gain by jointly coding *only over pairs of subchannels* as long as they are appropriately chosen. This approach results in a very low joint ML detection complexity, which only increases linearly with the number of pairs.

In this paper, we therefore propose codes named as X- and Y-Codes, due to the structure of the encoder matrix, which enable flexible *pairing*[1] of subchannels with different diversity orders. Specifically, the subchannels with low diversity orders can be paired together with those having high diversity orders, so that the overall diversity order is improved. The main contributions in this paper are as follows.

1) *X-Codes*: X-Codes are inspired by the signal space diversity techniques proposed in [20], based on rotated constellations. As shown in Figs. 1(a) and 2(a), in case no coding is performed across the two channel components (represented by the horizontal and the vertical axes), a deep fade along any one subchannel can result in an arbitrarily small minimum distance between the received codewords and hence the word error probability would increase. This problem is effectively resolved by rotating the 2-D codewords [see Figs. 1(b) and 2(b)]. Here, the minimum distance between the received codewords of a rotated constellation is larger and not vanishing even when there is a deep fade along one of the component subchannel. We therefore design 2-D real orthogonal rotation matrices, which are used to jointly code over pairs of subchannels, without increasing the transmit power. Since these matrices are effectively parameterized with a single angle, the design of X-Codes primarily involves choosing the optimal angle for each pair of subchannels. The angles are chosen *a priori* and do not change with each channel realization. This is why we use the term "Code" instead of "Precoder". The optimization of angles is based upon minimizing the average word error probability (i.e., averaged over the channel fading statistics) of the transmitted information symbol vector. At the receiver, we show that the MLD can be easily accomplished using n_r low complexity 2-D real ML decoders. Consider a pair of subchannels with subchannel gains $\lambda_1 \geq \lambda_2 \geq 0$. It is shown that when a pair of subchannels is *well conditioned* (i.e., λ_1/λ_2 close to 1), X-Codes have better error probability performance than that of other precoders. However, the error probability performance of X-Codes worsens when the pair of subchannels is *ill conditioned* (i.e., $\lambda_1/\lambda_2 \gg 1$). This can be explained as follows. When the subchannel pair is ill-conditioned, the error probability performance for the pair is determined primarily by the minimum Euclidean distance between the received codewords along the stronger subchannel component (λ_1). However, with the rotated constellation, the minimum received codeword distance along the stronger subchannel component may not be large enough, resulting in degradation of error performance in ill-conditioned channels. This along with the aim of further reducing the ML detection complexity, motivates the idea of Y-Codes.

2) *Y-Codes*: In a SVD precoded MIMO channel, the subchannel gains are the ordered singular values of the SVD decomposition of the MIMO channel gain matrix. By pairing these subchannels, it is obvious that in each pair, one of the subchannels is stronger than the other. It is

therefore intuitive that, the codewords be chosen so that the minimum Euclidean distance between the received codewords along the stronger subchannel component is larger than the minimum Euclidean distance along the weaker subchannel component. By doing so, the code design can make use of the total constrained transmit power to achieve a minimum received codeword Euclidean distance greater than that achieved with rotated constellations used in X-Codes. Y-Codes are designed based on this intuition, with the codewords forming a subset of a 2-D real skewed lattice [see Fig. 1(c)]. It can be seen that for the same rate (i.e., same number of equi-probable codewords), same transmit power constraint $P_T = 1$ and subchannel gains ($\lambda_1 = 1, \lambda_2 = 1/4$), Y-Codes achieve a greater minimum Euclidean distance between the received codewords when compared to X-Codes. Also, through simulations, we show that in ill-conditioned channels, Y-Codes have better error performance when compared to X-Codes. Y-Codes are parameterized with two parameters related to power allocated to the two subchannels. These parameters are computed so as to minimize the average error probability. The MLD complexity is the same as that of the scalar subchannels in linear precoders [8], [9] and is less than that of the X-Codes, while the performance of Y-Codes is better than that of X-Codes for ill-conditioned channel pairs.

3) *X-, Y-Precoders*: The X- and Y-Precoders employ the same pairing structure as that in X-, Y-Codes. However, the code generator matrix for each pair of subchannels is adaptively chosen for each channel realization. Through simulations it is observed that the average error performance of X- and Y-Precoders is marginally better than that of X- and Y-Codes.

Through average error probability analysis we show that, indeed, pairing of MIMO subchannels results in significant improvement in the overall diversity gain. The analytical results are also supported by numerical simulation. The simulation results have been reported in Section VI, from where it is clear that pairing of subchannels does indeed result in a higher diversity gain, when compared to the simple SVD precoding scheme (e.g., in Fig. 8, the error probability slope of the proposed X-,Y-Precoders is higher than the first order slope (no diversity gain) achieved by the linear precoders). Further, in Section VI, the error probability performance of X-,Y-Codes/Precoders has been shown to be better when compared to that of the other precoders reported in literature.

Pairing of good and bad (in terms of achievable diversity gain) subchannels has also been proposed in [12]. Despite having the same pairing structure, the proposed X- and Y-Codes/Precoders differ significantly from the E-dmin precoder proposed in [12] due to the fact that *i)* The encoder matrices for each pair are real-valued in case of X-,Y-Codes, as compared to being complex valued in the E-dmin precoder. This results in the ML detection for each pair to be over a 4-D real search space in case of the E-dmin precoder, as compared to only a 2-D real search for the proposed X- and Y-Codes, *ii)* The E-dmin precoder proposed in [12] was optimally designed only for 4-QAM. In [12],

---

[1]Pairing of two subchannels refers to joint coding of information symbols across the two subchannels.
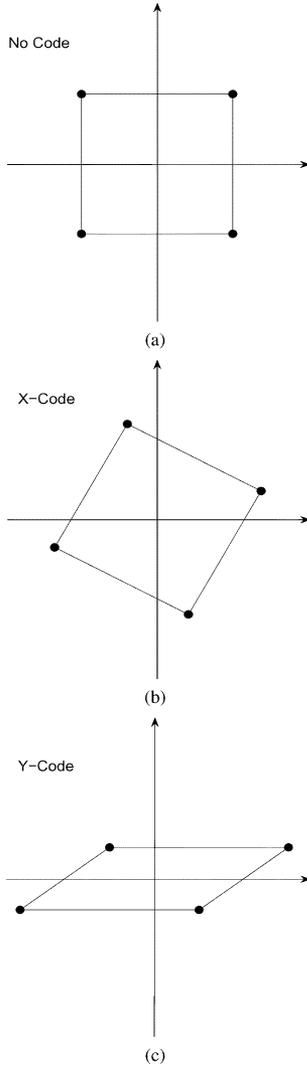
Fig. 1. Signal space of four 2-D codewords used to jointly code across two subchannels (horizontal and vertical). The average transmit power constraint is $P_T = 1$. The codewords are represented by solid dots.
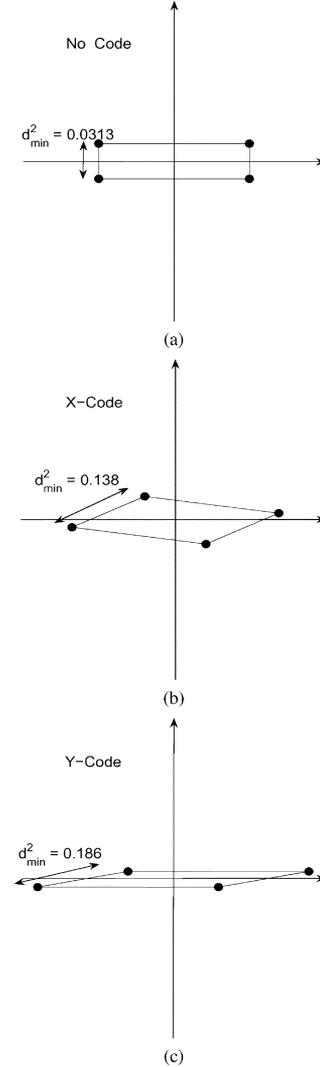

Fig. 2. Signal space of the received 2-D codewords. The gains of the hortizontal and vertical subchannels are $\lambda_1 = 1$ and $\lambda_2 = 1/4$ respectively. $d^2_{\min}$ is the minimum Euclidean distance between any two received codewords. The code parameters for X and Y-Code are optimized w.r.t. maximizing $d^2_{\min}$.

the optimal precoder design for higher order QAM could not be performed due to prohibitive complexity. In contrast, the proposed X- and Y-Codes are not designed for a specific modulation alphabet size, and are therefore more general than the E-dmin precoder, and *iii*) Through simulations it is observed that, with 4-QAM as the modulation alphabet, Y-Precoders have a similar bit error probability performance as the optimally designed E-dmin precoder. With higher order modulation alphabets (i.e., which achieve a rate higher than what is achieved with 4-QAM), Y-Precoders have a bit error probability performance significantly better than the E-dmin precoder.

The rest of the paper is organized as follows. Section II introduces the system model and SVD precoding. In Section III, we present the pairing of subchannels as a general coding strategy to achieve higher diversity order in fading channels. In Section IV, we propose the X-Codes and the X-Precoders. We show that ML decoding can be achieved with n_r 2-D real ML decoders. We also analyze the error probability performance and present the design of optimal X-Codes and X-Precoders.

In Section V, we propose the Y-Codes and Y-Precoders. We show that they have very low decoding complexity. We analyze the error probability performance and derive expressions for the optimal Y-Codes and Y-Precoders. Section VI shows the simulation results and comparisons with other precoders. Section VII discusses the complexity of the X-, Y-Codes/Precoders in comparison with other precoders. Conclusions are drawn in Section VIII.

Notations: Superscripts $^T, ^\dagger$, and * denote transposition, Hermitian transposition, and complex conjugation, respectively. The $n \times n$ identity matrix is denoted by $\mathbf{I}_n$, and the zero matrix is denoted by $\mathbf{0}$. $\mathbb{E}[\cdot]$ is the expectation operator, $\|\cdot\|$ denotes the Euclidean norm, and $|\cdot|$ denotes the absolute value of a complex number. The set of complex numbers, real numbers, nonnegative real numbers, and integers are denoted by $\mathbb{C}, \mathbb{R}, \mathbb{R}^+$ and $\mathbb{Z}$ respectively. Furthermore, $\lfloor c \rfloor$ denotes the largest integer not greater than c. Finally, we let $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary parts of a complex argument.

II. System Model and SVD Precoding

We consider a slow fading $n_t \times n_r$ MIMO ($n_r \leq n_t$), where the channel state information (CSI) is known perfectly at both the transmitter and receiver. Let $\mathbf{x} = (x_1, \ldots, x_{n_t})^T$ be the vector of symbols transmitted by the n_t transmit antennas in one channel use, and let $\mathbf{H} = \{h_{ij}\}, i = 1, \ldots, n_r, j = 1, \ldots, n_t$, be the $n_r \times n_t$ channel coefficient matrix, with h_{ij} as the complex channel gain between the j-th transmit antenna and the i-th receive antenna. The standard Rayleigh flat fading model is assumed with $h_{ij} \sim \mathcal{N}_c(0,1)$, i.e., i.i.d. complex Gaussian random variables with zero mean and unit variance. Rayleigh fading is one of the most common fading statistic used for the performance analysis of fading wireless channels. Nevertheless, improving diversity gain by pairing subchannels can be applied to any fading channel irrespective of its statistic. The *received vector* from the n_r receive antennas is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{n}$ is a spatially uncorrelated Gaussian noise vector such that $\mathbb{E}[\mathbf{n}\mathbf{n}^\dagger] = N_0 \mathbf{I}_{n_r}$.

Let the number of information symbols transmitted per channel use be $n_s (n_s \leq n_r)$. For every channel use, b information bits are first mapped to the *information symbol vector* $\mathbf{u} = (u_1, \ldots, u_{n_s})^T \in \mathbb{C}^{n_s}$, which is then mapped to the *data symbol vector* $\mathbf{z} = (z_1, \ldots, z_{n_s})^T \in \mathbb{C}^{n_s}$ using a $n_s \times n_s$ *encoding matrix* $\mathbf{G}$

$$\mathbf{z} = \mathbf{G}\mathbf{u} + \mathbf{u}^0 \tag{2}$$

where $\mathbf{u}^0 \in \mathbb{C}^{n_s}$ is a displacement vector used to reduce the average transmitted power.

Let $\mathbf{T}$ be the $n_t \times n_s$ precoding matrix which is applied to the data symbol vector to yield the transmitted vector

$$\mathbf{x} = \mathbf{T}\mathbf{z}. \tag{3}$$

It is obvious that the error performance is dependent on the precoding scheme (i.e., choice of $\mathbf{T}, \mathbf{G}$ and $\mathbf{u}^0$). Therefore for optimal error performance, $\mathbf{T}, \mathbf{G}$ and $\mathbf{u}^0$ are generally derived from the knowledge of $\mathbf{H}$ available at the transmitter. The transmission power constraint is given by

$$\mathbb{E}[||\mathbf{x}||^2] = P_T \tag{4}$$

and we define the SNR as

$$\gamma \triangleq \frac{P_T}{N_0}.$$

For the precoding schemes discussed in this paper, the rate and diversity gains are defined as follows. The rate R is defined as the number of information bits transmitted every channel use (bits-per-channel-use or bpcu). Since exactly b bits are transmitted every channel use, it is obvious that $R = b$ bpcu. For defining the achieved diversity gain/order δ, let $P(\mathbf{H}, \gamma)$ be the word error probability of $\mathbf{u}$ for a given channel realization $\mathbf{H}$ and a given SNR. Further, the average word error probability, i.e., word error probability of $\mathbf{u}$ averaged over the channel fading

statistics is $P(\gamma) = \mathbb{E}_{\mathbf{H}}[P(\mathbf{H}, \gamma)]$. The diversity gain/order is defined as

$$\delta \triangleq \lim_{\gamma \to \infty} \frac{-\log P(\gamma)}{\log \gamma}. \tag{5}$$

Note that this is the classical definition of diversity order, where the rate R is fixed for increasing SNR. This definition of rate and diversity is therefore different from that of Zheng and Tse [3][2].

Remark 1: Since we consider *slow fading* MIMO channels, transmissions are subject to block fading, and therefore diversity gain is the relevant metric to be considered. In case of *fast fading* MIMO channels, ergodic capacity is the relevant metric. In [21], we have demonstrated the superiority of X-Codes based precoder, in achieving a higher capacity than Mercury/waterfilling when information symbols belong to a discrete alphabet. ∎

The proposed X-, Y-Codes can be used to improve the error probability performance of the SVD precoding technique, which is based on the singular value decomposition of the channel matrix $\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}$, where $\mathbf{U} \in \mathbb{C}^{n_r \times n_r}$, $\mathbf{\Lambda} \in \mathbb{C}^{n_r \times n_r}, \mathbf{V} \in \mathbb{C}^{n_r \times n_t}$, such that $\mathbf{U}\mathbf{U}^\dagger = \mathbf{V}\mathbf{V}^\dagger = \mathbf{I}_{n_r}$ and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_{n_r})$ is the diagonal matrix of *singular values*, with $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_{n_r} \geq 0$ [1], [2].

Let $\tilde{\mathbf{V}} \in \mathbb{C}^{n_s \times n_t}$ be the submatrix with the first n_s rows of $\mathbf{V}$. The standard SVD precoder uses

$$\mathbf{T} = \tilde{\mathbf{V}}^\dagger$$
$$\mathbf{G} = \mathbf{I}_{n_s}$$
$$\mathbf{u}^0 = \mathbf{0} \tag{6}$$

and the receiver gets

$$\mathbf{y} = \mathbf{H}\mathbf{T}\mathbf{u} + \mathbf{n}. \tag{7}$$

Let $\tilde{\mathbf{U}} \in \mathbb{C}^{n_r \times n_s}$ be the submatrix with the first n_s columns of $\mathbf{U}$. The receiver computes

$$\mathbf{r} = \tilde{\mathbf{U}}^\dagger \mathbf{y} = \tilde{\mathbf{\Lambda}}\mathbf{u} + \mathbf{w} \tag{8}$$

where $\mathbf{w} \in \mathbb{C}^{n_s}$ is still an uncorrelated Gaussian noise vector with $\mathbb{E}[\mathbf{w}\mathbf{w}^\dagger] = N_0 \mathbf{I}_{n_s}$, $\tilde{\mathbf{\Lambda}} \triangleq \text{diag}(\lambda_1, \lambda_2, \ldots \lambda_{n_s})$, and $\mathbf{r} = (r_1, \ldots, r_{n_s})^T$. The SVD precoder therefore transforms the MIMO channel into n_s parallel subchannels/streams

$$r_i = \lambda_i u_i + w_i \quad i = 1, \ldots, n_s \tag{9}$$

with nonnegative fading coefficients λ_i.

The channel gain of the k-th subchannel is the k-th singular value of the channel matrix denoted by $\lambda_k, k = 1, 2, \ldots n_s$. Due to the ordering of the singular values during SVD decomposition, it is assumed that $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_{n_s}$. For the SVD precoder, it is also known that the diversity order achieved by the k-th stream alone (i.e., asymptotic slope of the averaged error probability for the information symbol u_k w.r.t. γ) is dependent upon how the probability density function (p.d.f.) of λ_k behaves

---

[2]Since the rate R is fixed with increasing γ, this actually corresponds to the point on the diversity multiplexing gain tradeoff curve with the multiplexing gain as zero.

around $\lambda_k = 0$ [14], [15]. In both [14] and [15], it is shown that if the p.d.f. of λ_k is $p(\lambda_k) = c_k \lambda_k^{(d_k-1)} + o(\lambda_k^{(d_k-1)})$ for $\lambda_k \to 0^+$, then d_k would be the diversity order of the k-th stream[3] For an i.i.d. Rayleigh faded $n_t \times n_r$ MIMO channel, the p.d.f. of the k-th singular value (around $\lambda_k = 0$) is $p(\lambda_k) = c_k \lambda_k^{(n_r-k+1)(n_t-k+1)-1} + o(\lambda_k^{(n_r-k+1)(n_t-k+1)-1})$ [14]. Therefore, the diversity order achieved by the k-th stream is $d_k = (n_r - k + 1)(n_t - k + 1)$. Hence, the lowest diversity order is achieved by the $k = n_s$-th stream. Similar results are also reported in [16].

When viewed as a single transmission system rather than multiple subchannels/streams, the overall[4] average word error probability of the information symbol vector $\mathbf{u}$ is dominated by the weakest subchannel [17], [18]. Due to the ordering of the singular values, it follows that the n_s-th subchannel is the weakest. Hence the overall diversity order achieved by the SVD precoder is $(n_r - n_s + 1)(n_t - n_s + 1)$. Further it is known that, the theoretical limit on the achievable overall diversity order is $n_r n_t$. The SVD precoding scheme can achieve this limit only when $n_s = 1$. This would, however, imply that, in order to achieve a target rate of R bpcu, the only transmitted symbol (since $n_s = 1$) must belong to some discrete signal set with 2^R complex symbols and an average symbol energy of P_T. In contrast, with $n_s = \min(n_r, n_t)$, the overall diversity order is much lower than the theoretical limit[5], but at the same time each information symbol is constrained to belong to some signal set with only $2^{R/n_s}$ complex symbols with an average energy of P_T/n_s. For the SVD precoding scheme, even though full MIMO diversity is achieved with $n_s = 1$, it is expected that at moderate SNR, the error probability performance achieved with $n_s = \min(n_t, n_r)$ is better than the error probability performance achieved with $n_s = 1$. A simple example with square QAM modulation symbols can illustrate this intuition. The error probability at moderate to high SNR is dependent upon the minimum Euclidean distance between the received codewords, which is related to the minimum Euclidean distance between the transmitted codewords. We therefore compare the minimum Euclidean distance between the transmitted codewords, for both the transmission scenarios (i.e., $n_s = 1$ and $n_s = \min(n_r, n_t)$), for a given rate R and average transmit power P_T. With $n_s = 1$, only one square complex 2^R-QAM information symbol is transmitted, and therefore the minimum Euclidean distance is $6P_T/(2^R-1)$. On the other hand, with $n_s = \min(n_r, n_t) = n_r, n_r$ square complex $2^{R/n_r}$-QAM symbols are transmitted per channel use, and the minimum Euclidean distance between the transmitted codewords is $6P_T/(n(2^{R/n_r} - 1))$. For $R > 0$ and $n_r \geq 1, (2^R - 1) \geq n(2^{R/n_r} - 1)$, and therefore it follows that $6P_T/(n(2^{\frac{R}{n_r}} - 1)) \geq 6P_T/(2^R - 1)$. Hence, at moderate SNR, the SVD precoding scheme with $n_s = \min(n_r, n_t)$ is expected to have an error probability performance better than the SVD precoding scheme with $n_s = 1$. Through simulations we have observed that, indeed at *moderate SNR*, the SVD precoding scheme with $n_s = \min(n_r, n_t)$ achieves a better error probability performance compared to the SVD precoding scheme

with $n_s = 1$. Based on the above discussion, it can be conjectured that, the SVD precoder is not the best precoder in terms of being both power efficient and achieving high diversity gain at the same time.

We next formally discuss as to how to compare different precoding schemes, and the same shall be used throughout this paper. Let $\mathcal{P}_A$ and $\mathcal{P}_B$ be two precoding schemes. If the diversity order achieved by these two precoding schemes is different, then it is obvious that the precoding scheme which achieves a higher diversity order will obviously have a lower error probability asymptotically as $\gamma \to \infty$. Therefore at high SNR, the non trivial scenario is when both the precoding schemes achieve the same diversity order. For a given fixed target rate of R bpcu, and an overall diversity order of $\delta = d$, denoted by the pair (R, d) and achievable by both precoders, let the asymptotic coding gain in the error performance of $\mathcal{P}_A$ w.r.t. that of $\mathcal{P}_B$ be defined as

$$\gamma_{gap}(R, d) \triangleq \lim_{\gamma_A, \gamma_B \to \infty \,|\, P_e(\mathcal{P}_A, \gamma_A) = P_e(\mathcal{P}_B, \gamma_B)} \gamma_B/\gamma_A. \quad (10)$$

In (10), $P_e(\mathcal{P}_A, \gamma_A)$ refers to the word error probability of the precoder $\mathcal{P}_A$ at a SNR of γ_A. A similar definition holds true for $P_e(\mathcal{P}_B, \gamma_B)$. If $\gamma_{gap}(R, d) > 1$, the precoder $\mathcal{P}_A$ is said to be *better* than $\mathcal{P}_B$ for the given (R, d). This also means that the precoding scheme $\mathcal{P}_A$ is more *power efficient* than scheme $\mathcal{P}_B$ for the given (R, d). For a given d achievable by both the precoders, if $\gamma_{gap}(R, d) > 1$ for all possible values of R, then precoder $\mathcal{P}_A$ is said to be *universally better* than precoder $\mathcal{P}_B$ for the given diversity order d. For a given diversity order d, we can then define the *best* precoder to be the one which is *universally better* than all the other precoders with the same diversity order of d.

It would be of theoretical interest to find the *best* precoder for a given achievable diversity order d. Though in theory, the maximum possible achievable diversity order is $n_r n_t$, it is likely that the precoders achieving $d = n_r n_t$, would also require highly complex ML detection at the receiver. With Rayleigh fading, at SNR values of interest, we have observed that even for moderate values of (n_r, n_t), the error probability slope corresponding to the maximum diversity order $d = n_r n_t$ is only marginally better than that of a precoding scheme which achieves a diversity order slightly less than $n_r n_t$. For example, with $n_r = n_t = 4$, it is observed that the error probability slopes for the first and second subchannels (with gains λ_1 and λ_2) are almost similar at SNR values of interest. Therefore from a practical standpoint, it would be of interest to design precoding schemes which have a *low complexity* ML detector, can achieve *sufficiently high diversity* order, and which are almost as power efficient as the *best* precoder. In this paper, we present two precoders, X- and Y-Codes, both of which are shown to achieve high diversity order with low complexity ML detection. Y-Codes have an even lower ML detection complexity and better error probability performance than X-Codes.

III. PAIRING GOOD AND BAD SUBCHANNELS

Without loss of generality, we consider the SVD precoding scheme with even n_r and $n_s = n_r$. In this section, we motivate the pairing of subchannels as a low complexity technique to improve the overall diversity order. This pairing is inspired

---

[3]Any function $f(x)$ in a single variable x is said to be $o(g(x))$ i.e., $f(x) = o(g(x))$ if $\frac{f(x)}{g(x)} \to 0$ as $x \to 0$.

[4]In this paper, the word "overall" used in which ever context, applies to the whole information symbol vector $\mathbf{u}$.

[5]Note that when $n_r = n_t = n_s$, the diversity order achieved is only 1.

from the use of rotated constellations in SISO fading channels to achieve modulation coding diversity [19], [20]. The idea is to jointly code over a set of m information symbols, and transmit the coded information symbols over m different channel realizations (in frequency or time). This coding scheme guarantees a nonzero minimum distance between the transmitted codewords along any of the m component channels even in case of deep fades. Since the additive noise is Gaussian, a ML detection error would only happen when the minimum Euclidean distance between the received codewords is small. Due to a nonzero minimum Euclidean distance between the received codewords along any component channel, the minimum Euclidean distance between the received codewords is small only when all the component channels experience deep fade. Since the event of all the m component channels undergoing deep fade is less probable than a single channel undergoing deep fade, it can be concluded that the joint coding scheme with ML detection would result in improvement of the diversity order. Note that with the joint coding scheme, an m-fold diversity gain is fully achieved with ML detection whose complexity increases rapidly with m [19].

In order to keep the ML detection complexity low, we restrict to $m = 2$, and perform joint coding over pairs of subchannels of the MIMO channel. In particular, a pair of information symbols is jointly coded, and one of the two coded symbol is be transmitted on the stronger subchannel whereas the other coded symbol is be transmitted on the weaker subchannel.

With a MIMO channel, since the subchannel gains are *not i.i.d.*, the system is different from the SISO scenario discussed in [19], [20]. With MIMO subchannels, due to the ordering of the singular values, in any given pair of subchannels, one of the subchannels is always stronger than the other one. Due to this fact, an error will always happen if there is a deep fade in the stronger channel (since this automatically implies that the weaker channel is also in deep fade). This then implies that the maximum possible diversity order that can be achieved, when coding over a pair of MIMO subchannels, is indeed the diversity order achieved by transmitting only on the stronger subchannel[6]. Therefore when pairing MIMO subchannels, as long as the minimum distance between the transmitted 2-D codewords is nonzero along the stronger subchannel component, the joint coding scheme is guaranteed to achieve the maximum possible diversity. This is different from the case of SISO Rayleigh fading channels, where in order to achieve maximal diversity, the minimum distance between the codewords must be nonzero along all component channels [19], [20].

The pairing of subchannels is achieved as follows. The matrix $\mathbf{G} \in \mathbb{C}^{n_r \times n_r}$ is used to pair different subchannels in order to improve the overall diversity order. The precoding matrix $\mathbf{T} \in \mathbb{C}^{n_t \times n_r}$ and the transmitted vector $\mathbf{x}$ are given by

$$\mathbf{T} = \mathbf{V}^\dagger, \quad \mathbf{x} = \mathbf{V}^\dagger(\mathbf{G}\mathbf{u} + \mathbf{u}^0). \tag{11}$$

Let the list of pairings be $(i_k, j_k) \in [1, n_r] \times [1, n_r], k \in [1, n_r/2]$ and $i_k < j_k$. On the k-th pair, consisting of subchannels i_k and j_k, the information symbols u_{i_k} and u_{j_k} are jointly coded using a 2×2 matrix $\mathbf{A}_k$. In order to reduce the

---

[6]In the case of i.i.d. SISO channels, it is possible to achieve a diversity order greater than the diversity order of any of the component channels.

ML decoding complexity, we restrict the entries of $\mathbf{A}_k$ to be real valued. Each $\mathbf{A}_k \triangleq \{a_{k,i,j}\}, i, j \in [1, 2]$, is a submatrix of the code matrix $\mathbf{G}$ as follows:

$$\begin{aligned} g_{i_k, i_k} = a_{k,1,1} \quad & g_{i_k, j_k} = a_{k,1,2} \\ g_{j_k, i_k} = a_{k,2,1} \quad & g_{j_k, j_k} = a_{k,2,2} \end{aligned} \tag{12}$$

where $g_{i,j}$ is the entry of $\mathbf{G}$ in the i-th row and j-th column.

Both the proposed X- and Y-Codes achieve diversity improvement by jointly coding over a pair of subchannels. The only difference is in the structure of the linear code generator matrix $\mathbf{A}_k$ for the k-th pair. In the case of X-Codes, 2-D real rotation matrices are used, whereas for Y-Codes, the code generator matrix has a upper left triangular structure. Also, there are finitely many ways to pair the subchannels, and as we shall show later, one pairing which is optimal in terms of the achievable overall diversity, is to pair the k-th and the $(n_r - k + 1)$-th subchannel. When this pairing is represented in matrix form, the code matrix $\mathbf{G}$ has a cross-form structure, and hence the name X-Codes. With Y-Codes, the right bottom entries of the code generator matrices $\mathbf{A}_k$ for each pair is 0, and hence $\mathbf{G}$ appears like the letter "Y".

For example, with $n_r = 6$, the X-Code structure is given by

$$\mathbf{G} = \begin{bmatrix} a_{1,1,1} & & & & & a_{1,1,2} \\ & a_{2,1,1} & & & a_{2,1,2} & \\ & & a_{3,1,1} & a_{3,1,2} & & \\ & & a_{3,2,1} & a_{3,2,2} & & \\ & a_{2,2,1} & & & a_{2,2,2} & \\ a_{1,2,1} & & & & & a_{1,2,2} \end{bmatrix} \tag{13}$$

and the Y-Code structure is given by

$$\mathbf{G} = \begin{bmatrix} a_{1,1,1} & & & & & a_{1,1,2} \\ & a_{2,1,1} & & & a_{2,1,2} & \\ & & a_{3,1,1} & a_{3,1,2} & & \\ & & a_{3,2,1} & & & \\ & a_{2,2,1} & & & & \\ a_{1,2,1} & & & & & \end{bmatrix}. \tag{14}$$

Let $\mathbf{u}_k \triangleq [u_{i_k}, u_{j_k}]^T$ denote the k-th information pair. Due to the transmit power constraint in (4), and uniform transmit power allocation between the $n_r/2$ pairs, the encoder matrices $\mathbf{A}_k$ must satisfy

$$\mathbb{E}\left[\left\|\mathbf{A}_k\mathbf{u}_k + \mathbf{u}_k^0\right\|^2\right] = 2P_T/n_r, \quad k = 1, 2, \ldots n_r/2. \tag{15}$$

The expectation in (15) is over the distribution of the information symbol vector $\mathbf{u}_k$ and $\mathbf{u}_k^0$ is the subvector of the displacement vector $\mathbf{u}^0$ for the k-th pair. The matrices $\mathbf{A}_k$ for X- and Y-Codes can be either fixed *a priori* or can change with every channel realization. The latter case leads to the X- and Y-Precoders.

A. ML Decoding

Given the received vector $\mathbf{y}$, the receiver computes

$$\mathbf{r} = \mathbf{U}^\dagger\mathbf{y} - \mathbf{\Lambda}\mathbf{u}^0. \tag{16}$$

Since $\mathbf{u}^0$ is a deterministic function of the channel state, it is known to both the transmitter and receiver. Using (1) and (11), we can rewrite (16) as

$$\mathbf{r} = \mathbf{\Lambda G u} + \mathbf{w} = \mathbf{M u} + \mathbf{w} \qquad (17)$$

where $\mathbf{M} \triangleq \mathbf{\Lambda G}$ is the effective channel matrix and $\mathbf{w} \triangleq \mathbf{U}^\dagger \mathbf{n}$ is a noise vector with the same statistics as $\mathbf{n}$. Further, we let

$$\mathbf{r}_k \triangleq [r_{i_k}, r_{j_k}]^T$$
$$\mathbf{w}_k \triangleq [w_{i_k}, w_{j_k}]^T .$$

Let $\mathbf{M}_k \in \mathbb{R}^{2 \times 2}$ denote the 2×2 submatrix of $\mathbf{M}$ consisting of entries in the i_k and j_k rows and columns. Then (17) can be equivalently written as

$$\mathbf{r}_k = \mathbf{M}_k \mathbf{u}_k + \mathbf{w}_k, \quad k = 1, \dots, n_r/2. \qquad (18)$$

Also, let $\Re(\mathbf{u}_k), \Im(\mathbf{u}_k) \in \mathcal{S}_k$, where $\mathcal{S}_k$ is a finite signal set in the 2-D real space.

The rate R is then given by

$$R = 2 \sum_{k=1}^{n_r/2} \log_2(|\mathcal{S}_k|). \qquad (19)$$

Also, let $\mathcal{S}_R \triangleq \mathcal{S}_1 \times \mathcal{S}_2 \cdots \times \mathcal{S}_{n_r/2}$ be the Cartesian product of the finite signal sets $\mathcal{S}_k, \quad k = 1, 2 \dots n_r/2$, then $\Re(\mathbf{u}), \Im(\mathbf{u}) \in \mathcal{S}_R$.

From (18), it is also clear that the ML decoder for $\mathbf{u}$ reduces to independent ML decoders for each $\mathbf{u}_k$. Further, the ML decoding for the k-th pair can be separated into independent ML decoding of the real and imaginary components of $\mathbf{u}_k$, i.e.,

$$\Re(\hat{\mathbf{u}}_k) = \arg \min_{\Re(\mathbf{u}_k) \in \mathcal{S}_k} \|\Re(\mathbf{r}_k) - \mathbf{M}_k \Re(\mathbf{u}_k)\|^2 \qquad (20)$$
$$\text{and}$$
$$\Im(\hat{\mathbf{u}}_k) = \arg \min_{\Im(\mathbf{u}_k) \in \mathcal{S}_k} \|\Im(\mathbf{r}_k) - \mathbf{M}_k \Im(\mathbf{u}_k)\|^2 \qquad (21)$$

where $\hat{\mathbf{u}}_k = (\hat{u}_{k,1}, \hat{u}_{k,2})^T$ is the output of the ML detector for the k-th pair.

Further, let $\tilde{\mathbf{u}} = (\tilde{u}_1, \tilde{u}_2, \dots \tilde{u}_{n_r})^T$ denote the detected information symbol vector $\mathbf{u}$. The entries of $\tilde{\mathbf{u}}$ are composed of the $n_r/2$ ML detector outputs $\hat{\mathbf{u}}_k, \quad k = 1, 2, \dots n_r/2$, as follows.

$$\tilde{u}_{i_k} = \hat{u}_{k,1}, \ \tilde{u}_{j_k} = \hat{u}_{k,2} \quad k = 1, 2, \dots n_r/2. \qquad (22)$$

B. Performance Analysis

For a given channel realization $\mathbf{H}$, the word error probability (WEP) for the k-th transmitted information symbol pair is given by

$$P_k(\mathbf{H}) = \frac{1}{|\mathcal{S}_k|^2} \sum_{\Re(\mathbf{u}_k), \Im(\mathbf{u}_k) \in \mathcal{S}_k} \text{Prob}(\hat{\mathbf{u}}_k \neq \mathbf{u}_k \mid \mathbf{u}_k, \mathbf{H}).$$
$$(23)$$

The overall average word error probability for the information symbol vector $\mathbf{u}$ is given by

$$P(\mathbf{H}) = \frac{1}{|\mathcal{S}_R|^2} \sum_{\Re(\mathbf{u}), \Im(\mathbf{u}) \in \mathcal{S}_R} \text{Prob}(\tilde{\mathbf{u}} \neq \mathbf{u} \mid \mathbf{u}, \mathbf{H}). \qquad (24)$$

For a given channel realization $\mathbf{H}$, the transmitted information vector $\mathbf{u}$ is not in error if and only if none of the pairs of information symbols $\mathbf{u}_k, \quad k = 1, 2, \dots n_r/2$ are in error. Further, since the additive receiver noise for each pair is independent, we have

$$P(\mathbf{H}) = 1 - \prod_{k=1}^{n_r/2} (1 - P_k(\mathbf{H})). \qquad (25)$$

The overall word error probability for the information symbol vector $\mathbf{u}$, averaged over the channel fading statistics, is given by

$$P \triangleq \mathbb{E}_{\mathbf{H}}[P(\mathbf{H})]. \qquad (26)$$

Similarly, the average word error probability for the k-th pair is given by

$$P_k \triangleq \mathbb{E}_{\mathbf{H}}[P_k(\mathbf{H})]. \qquad (27)$$

From (20) and (21), we see that for a given channel realization $\mathbf{H}$, the WEPs for the real and the imaginary components of the k-th pair are the same. Therefore, without loss of generality we can analyze the WEP only for the real component, which is given by

$$P_k'(\mathbf{H}) = \frac{1}{|\mathcal{S}_k|} \sum_{\Re(\mathbf{u}_k) \in \mathcal{S}_k} \text{Prob}(\Re(\hat{\mathbf{u}}_k) \neq \Re(\mathbf{u}_k) | \Re(\mathbf{u}_k), \mathbf{H}).$$
$$(28)$$

Since the additive receiver noise on the real and imaginary components of each pair are i.i.d., it follows that $P_k(\mathbf{H}) = 1 - (1 - P_k'(\mathbf{H}))^2$. Let $P_k'(\Re(\mathbf{u}_k)) \triangleq \mathbb{E}_{\mathbf{H}}[\text{Prob}(\Re(\hat{\mathbf{u}}_k) \neq \Re(\mathbf{u}_k) | \Re(\mathbf{u}_k), \mathbf{H})]$, then the average word error probability of the real component of $\mathbf{u}_k$ is then given by

$$P_k' \triangleq \mathbb{E}_{\mathbf{H}}[P_k'(\mathbf{H})] = \frac{1}{|\mathcal{S}_k|} \sum_{\Re(\mathbf{u}_k)} P_k'(\Re(\mathbf{u}_k)) \qquad (29)$$

where $P_k'(\Re(\mathbf{u}_k))$ has to be evaluated differently for X-, Y-Codes and X-, Y-Precoders. To explain this difference we need the following definitions.

For a given channel realization, and therefore deterministic values of λ_{i_k} and λ_{j_k} for the k-th pair, we let $P_k'(\Re(\mathbf{u}_k), \lambda_{i_k}, \lambda_{j_k}, \mathbf{A}_k)$ denote the error probability of ML detection for the real component of the k-th channel, given that the information symbol $\mathbf{u}_k$ was transmitted on the k-th pair. For X-, Y-Codes, the matrices $\mathbf{A}_k$ are fixed *a priori* and are not functions of the deterministic value of subchannel gains, and therefore, $P_k'(\Re(\mathbf{u}_k))$ is given by

$$P_k'(\Re(\mathbf{u}_k)) = \mathbb{E}_{(\lambda_{i_k}, \lambda_{j_k})}[P_k'(\Re(\mathbf{u}_k), \lambda_{i_k}, \lambda_{j_k}, \mathbf{A}_k)]. \qquad (30)$$

We observe that $P'_k(\Re(\mathbf{u}_k))$ is actually a function of $\mathbf{A}_k$ and therefore the optimal error performance is obtained by minimizing (29) over $\mathbf{A}_k$. Thus, the optimal matrix for the k-th pair is given by

$$\mathbf{A}_k^{opt} = \arg\min_{\mathbf{A}_k} \sum_{\Re(\mathbf{u}_k)} \mathbb{E}_{(\lambda_{i_k}, \lambda_{j_k})}[P'_k(\Re(\mathbf{u}_k), \lambda_{i_k}, \lambda_{j_k}, \mathbf{A}_k)].$$

(31)

The minimization in (31) is constrained over matrices $\mathbf{A}_k$ which satisfy (15). The corresponding optimal average word error probability for the real component of the k-th pair is given by

$$P_k^{opt} = \frac{\sum_{\Re(\mathbf{u}_k)} \mathbb{E}_{(\lambda_{i_k}, \lambda_{j_k})} \left[P'_k \left(\Re(\mathbf{u}_k), \lambda_{i_k}, \lambda_{j_k}, \mathbf{A}_k^{opt} \right) \right]}{|\mathcal{S}_k|}.$$

(32)

For the X-, Y-Precoder, the matrices $\mathbf{A}_k$ are chosen adaptively every time the channel changes. For optimal error performance, the matrices $\mathbf{A}_k$ are chosen so as to minimize the error probability $P'_k(\mathbf{H})$ for a given channel realization $\mathbf{H}$. The optimal encoding matrix for the k-th pair is then given by

$$\mathbf{A}_k^{opt}(\lambda_{i_k}, \lambda_{j_k}) = \arg\min_{\mathbf{A}_k} \sum_{\Re(\mathbf{u}_k)} P'_k(\Re(\mathbf{u}_k), \lambda_{i_k}, \lambda_{j_k}, \mathbf{A}_k).$$

(33)

The minimization in (33) is constrained over matrices $\mathbf{A}_k$ which satisfy (15). Therefore, with X- and Y-Precoders, the optimal average word error probability for the real component of the k-th pair is given by (34), as shown at the bottom of the page. Comparing (34) and (32), we immediately observe that the optimal error performance of X-, Y-Precoders is better than that of X-, Y-Codes.

Our next goal is to derive an analytic expression for P'_k. We shall only discuss the derivation for X-, Y-Codes, since the error performance of X-, Y-Precoders is better than X-, Y-Codes and therefore they achieve at least as much diversity order as X-, Y-Codes.

Getting an exact analytic expression for P'_k is difficult, and therefore we try to get tight upper bounds using the union bound.

Let $\{\Re(\mathbf{u}_k) \to \Re(\mathbf{v}_k)\}$ denote the pairwise error event that, given $\mathbf{u}_k$ was transmitted on the k-th pair, the real part of the ML detector for the k-th pair decodes in favor of some other vector $\Re(\mathbf{v}_k)$. Further, let us denote the corresponding pairwise error probability (PEP) by $P'_k(\Re(\mathbf{u}_k) \to \Re(\mathbf{v}_k))$. Using the union bounding technique, $P'_k(\Re(\mathbf{u}_k))$ can be upper bounded by the sum of all the possible pairwise error probabilities. From (29), it is clear that this upper bound on $P'_k(\Re(\mathbf{u}_k))$ induces an upper bound on P'_k, which is given by

$$P'_k \leq \frac{1}{|\mathcal{S}_k|} \sum_{\Re(\mathbf{u}_k)} \sum_{\Re(\mathbf{v}_k) \neq \Re(\mathbf{u}_k)} P'_k(\Re(\mathbf{u}_k) \to \Re(\mathbf{v}_k)). \quad (35)$$

Due to Gaussian noise, this can be further written as shown in (36), found at the bottom of the page, where

$$d_k^2(\Re(\mathbf{u}_k), \Re(\mathbf{v}_k), \mathbf{A}_k) \triangleq \|\mathbf{M}_k(\Re(\mathbf{u}_k) - \Re(\mathbf{v}_k))\|^2$$

(37)

and

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt. \quad (38)$$

The expectation in (36) is over the joint distribution of the channel gains $(\lambda_{i_k}, \lambda_{j_k})$. The joint p.d.f. of the ordered eigenvalues of $\mathbf{H}^\dagger \mathbf{H}$ is given by the well-known Wishart distribution [13]. However, evaluating the expectation over $(\lambda_{i_k}, \lambda_{j_k})$ in (36) is still a difficult problem except for trivial cases (like $n_r = n_t = 2$). We therefore try to lower bound $d_k^2(\Re(\mathbf{u}_k), \Re(\mathbf{v}_k), \mathbf{A}_k)$ with a quantity depending only on λ_{i_k}. Since $\lambda_{i_k} \geq \lambda_{j_k} \geq 0$, using the definition of $\mathbf{M}$ and $\mathbf{M}_k$, we have

$$d_k^2(\Re(\mathbf{u}_k), \Re(\mathbf{v}_k), \mathbf{A}_k) \geq \lambda_{i_k}^2 \tilde{d}_k^2(\Re(\mathbf{u}_k), \Re(\mathbf{v}_k), \mathbf{A}_k) \quad (39)$$

where we define the *generalized pairwise distance* between the vectors $\Re(\mathbf{u}_k)$ and $\Re(\mathbf{v}_k)$ as

$$\tilde{d}_k^2(\Re(\mathbf{u}_k), \Re(\mathbf{v}_k), \mathbf{A}_k) \triangleq e_{k,1}^2 \quad (40)$$

and

$$\mathbf{e}_k \triangleq \mathbf{A}_k(\Re(\mathbf{u}_k) - \Re(\mathbf{v}_k)) = [e_{k,1}, e_{k,2}] \quad (41)$$

and we let $e_{k,1}$ denote the first component of the 2-D vector $\mathbf{e}_k$. We further define the *generalized minimum distance* as follows:

$$g_k(\mathbf{A}_k) = \min_{\Re(\mathbf{u}_k) \neq \Re(\mathbf{v}_k)} \tilde{d}_k^2(\Re(\mathbf{u}_k), \Re(\mathbf{v}_k), \mathbf{A}_k). \quad (42)$$

The following theorem gives an upper bound to P'_k based on the union bounding technique discussed above.

$$P_k^{opt} = \frac{\sum_{\Re(\mathbf{u}_k)} \mathbb{E}_{(\lambda_{i_k}, \lambda_{j_k})} \left[P'_k \left(\Re(\mathbf{u}_k), \lambda_{i_k}, \lambda_{j_k}, \mathbf{A}_k^{opt}(\lambda_{i_k}, \lambda_{j_k}) \right) \right]}{|\mathcal{S}_k|}$$

(34)

$$P'_k \leq \frac{\sum_{\Re(\mathbf{u}_k)} \sum_{\Re(\mathbf{v}_k) \neq \Re(\mathbf{u}_k)} \mathbb{E}\left[Q\left(\sqrt{\frac{d_k^2(\Re(\mathbf{u}_k), \Re(\mathbf{v}_k), \mathbf{A}_k)}{2N_0}} \right) \right]}{|\mathcal{S}_k|}$$

(36)

Theorem 1: An upper bound to P_k' as $\gamma \to \infty$ is given by

$$P_k' \leq c_k(|\mathcal{S}_k| - 1)\left[\frac{2P_T}{g_k(\mathbf{A}_k)}\right]^{\delta_k}\gamma^{-\delta_k} + o(\gamma^{-\delta_k}) \qquad (43)$$

where

$$\delta_k \triangleq (n_t - i_k + 1)(n_r - i_k + 1)$$
$$c_k \triangleq \frac{C(i_k)}{2^{\delta_k}}((2\delta_k - 1) \ldots 5 \cdot 3 \cdot 1).$$

and the coefficients $C(m), 1 \leq m \leq \min(n_r, n_t)$ are defined in [15].

Proof: See Appendix A. ∎

The diversity order achievable by the k-th pair is given by

$$d_k \triangleq \lim_{\gamma \to \infty} \frac{-\log P_k}{\log \gamma}. \qquad (44)$$

As $\gamma \to \infty$, $P_k(\mathbf{H}) = 1 - (1 - P_k'(\mathbf{H}))^2 \leq 2P_k'(\mathbf{H})$. Therefore, $P_k \leq 2P_k'$. Using this fact and (43), the diversity order achievable by the k-th pair is lower bounded as

$$d_k \geq \delta_k. \qquad (45)$$

Let the *overall diversity order* be defined as

$$\delta_{ord} \triangleq \lim_{\gamma \to \infty} \frac{-\log P}{\log \gamma}. \qquad (46)$$

The following theorem gives a lower bound on the overall achievable diversity order.

Theorem 2: A lower bound on the overall achievable diversity order is given by

$$\delta_{ord} \geq \min_k \delta_k. \qquad (47)$$

Proof: See Appendix B. ∎

Remark 2: A similar fact has been stated without proof in [17], where it is mentioned that with multiple subchannels/streams, the overall error probability at high SNR is dominated by the error probability of the subchannel having the lowest diversity gain. It is then concluded that the overall diversity order is equal to the diversity order of the subchannel having the lowest diversity gain. ∎

The bound on the overall diversity order δ_{ord}, given by (47), also holds for the X-, Y-Precoders. This is so because, for each channel realization $\mathbf{H}$, X- and Y-Precoders could choose the encoding matrices to be the same as the matrices designed for X-,Y-Codes.

C. Design of Optimal Pairing

From the lower bound on δ_{ord} (given by (47)) it is clear that the following pairing of subchannels

$$i_k = k, \quad j_k = n_r - k + 1 \qquad (48)$$

achieves the following best lower bound:

$$\delta_{ord} \geq \left(\frac{n_r}{2} + 1\right)\left(n_t - \frac{n_r}{2} + 1\right). \qquad (49)$$

Remark 3: Note that this corresponds to a cross-form generator matrix $\mathbf{G}$, and is not the only pairing for the best lower bound. Also, we note that the overall diversity order improves significantly, when compared to the case of no pairing. As an example, with $n_r = n_t = n_s$, the overall diversity order achieved with the proposed pairing structure is $(n_r/2 + 1)^2$ as compared to an overall diversity order of only 1, when no pairing of subchannels is performed. It can be shown that, if only n_s (n_s even) out of the n_r subchannels are used for transmission, the lower bound on the overall achievable diversity order with the proposed pairing structure is $(n_r - \frac{n_s}{2} + 1)(n_t - \frac{n_s}{2} + 1)$. ∎

For X- and Y-Codes, although it is hard to compute $\mathbf{A}_k^{opt}$ in (31), we can compute the best $\mathbf{A}_k$, denoted by $\mathbf{A}_k^*$, which minimizes the upper bound on P_k' in (43). We then have

$$\mathbf{A}_k^* = \arg \max_{\mathbf{A}_k | \mathbb{E}[\|\mathbf{A}_k\mathbf{u}_k + \mathbf{u}_k^0\|^2] = \frac{2P_T}{n_r}} \frac{g_k(\mathbf{A}_k)}{2P_T}. \qquad (50)$$

Using (43), (48), and (50), we obtain

$$P_k' \leq c_k(|\mathcal{S}_k| - 1)\left[\frac{2P_T}{g_k(\mathbf{A}_k^*)}\right]^{\delta_k^*}\gamma^{-\delta_k^*} + o(\gamma^{-\delta_k^*}) \qquad (51)$$

where $\delta_k^* \triangleq (n_t - k + 1)(n_r - k + 1)$.

IV. X-CODES AND X-PRECODERS

A. X-Codes and X-Precoders: Encoding and Decoding

For X-Codes, each symbol in $\mathbf{u}$ takes values from a regular M^2-QAM constellation, which consists of the Cartesian product of two M-PAM constellations $\mathcal{S} \triangleq \{\tau(2i - (M - 1)) | i = 0, 1, \ldots (M - 1)\}$ used on the real or the imaginary components of two subchannels (i.e., $\mathcal{S}_k = \mathcal{S}^2$ for $k = 1, 2, \ldots, n_r/2$). The scaling factor τ is defined as

$$\tau \triangleq \sqrt{\frac{3E_s}{2(M^2 - 1)}}$$

and

$$E_s = \frac{P_T}{n_r}$$

is the average symbol energy for each information symbol in the vector $\mathbf{u}$. Gray mapping is used to map the bits separately to the real and imaginary component of the symbols in $\mathbf{u}$. We impose an orthogonality constraint on each $\mathbf{A}_k$ (in (12)) and conveniently parameterize it with a single angle θ_k

$$\mathbf{A}_k = \begin{bmatrix} \cos(\theta_k) & \sin(\theta_k) \\ -\sin(\theta_k) & \cos(\theta_k) \end{bmatrix} \qquad (52)$$

where $k = 1, \ldots n_r/2$. We notice that 1) since $\mathbf{A}_k$ is orthogonal, $\mathbf{G}$ is also orthogonal; 2) for X-Codes we fix the angles θ_k *a priori*, whereas for the X-Precoders we change the angles for each channel realization; 3) we can fix $\mathbf{u}^0$ in (2) to be the zero
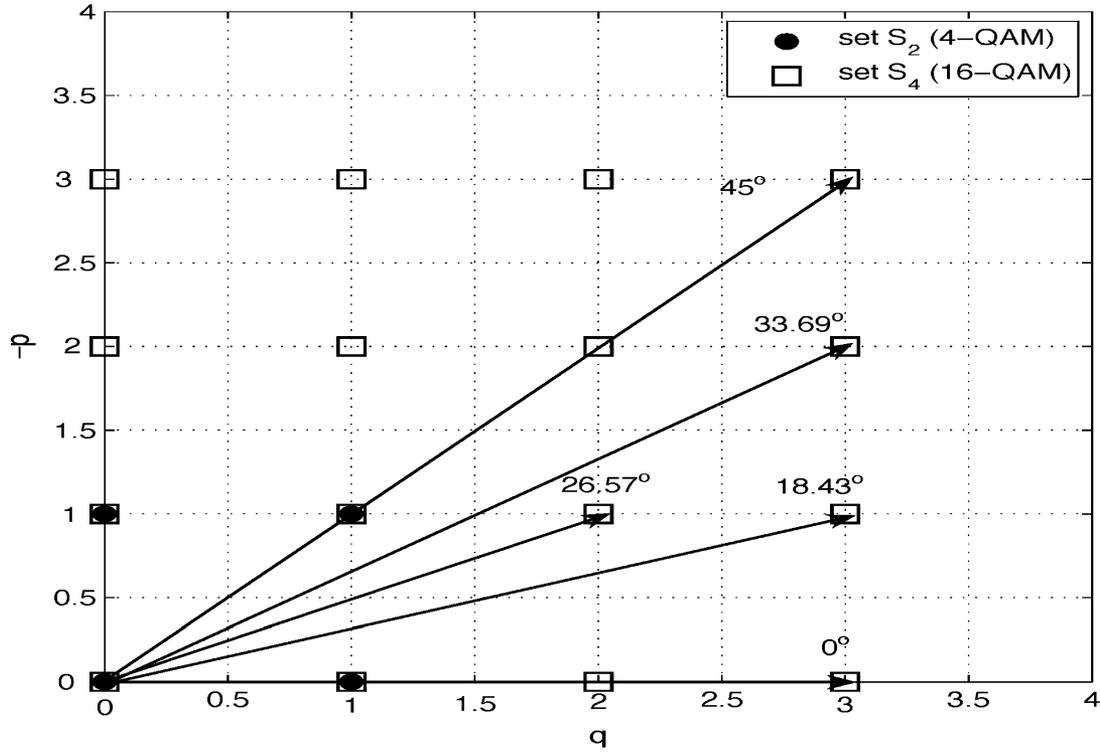
Fig. 3. One quadrant of the set $\mathbb{S}_M$ for $M = 2, 4$ (4,16-QAM modulation). The critical angles where performance degrades severely are shown to coincide with $\tan^{-1}(-p/q)$.

vector, since to the orthogonality of $\mathbf{G}$ preserves the QAM shape of the signal set.

From (20) and (21), it is obvious that two 2-D real ML decoders are needed for each pair. Since there are $n_r/2$ pairs, the total decoding complexity is n_r 2-D real ML decoders. For X-Codes, the matrices $\mathbf{M}_k$ in (20) and (21) are given by

$$\mathbf{M}_k = \begin{bmatrix} \lambda_{i_k} \cos(\theta_k) & \lambda_{i_k} \sin(\theta_k) \\ -\lambda_{j_k} \sin(\theta_k) & \lambda_{j_k} \cos(\theta_k) \end{bmatrix}. \qquad (53)$$

B. Optimal Design of X-Codes

In order to find the best angle θ_k^* for the k-th pair, we attempt to maximize the generalized minimum distance $g_k(\mathbf{A}_k)$ (defined in (42)) under the transmit power constraints.

For X-Codes, the difference vector between the real components of any two information vectors $\Re(\mathbf{u}_k)$ and $\Re(\mathbf{v}_k)$ for the k-th pair is given by

$$\Re(\mathbf{u}_k) - \Re(\mathbf{v}_k) = \sqrt{\frac{6E_s}{M^2 - 1}}(p, q)^T, (p, q) \in \mathbb{S}_M \qquad (54)$$

where

$$\mathbb{S}_M \triangleq \{(p, q) \mid |p|, |q| \in [0, (M - 1)], \quad (p, q) \neq (0, 0)\}.$$

The set $\mathbb{S}_M$ for $M = 2$ (4-QAM) and $M = 4$ (16-QAM) is shown in Fig. 3. Using (54) in (40), the generalized pairwise distance between $\Re(\mathbf{u}_k)$ and $\Re(\mathbf{v}_k)$ is given by

$$\tilde{d}_k^2(\Re(\mathbf{u}_k), \Re(\mathbf{v}_k), \mathbf{A}_k) = \frac{6P_T}{n_r(M^2 - 1)} \times (p \cos(\theta_k) + q \sin(\theta_k))^2. \qquad (55)$$

Since $\mathbf{A}_k$ is parameterizable with a single angle θ_k, we shall rename the generalized minimum distance in (42) by

$$g_k(\theta_k, M) \triangleq g_k(\mathbf{A}_k)$$
$$= \frac{6P_T \min_{(p,q) \in \mathbb{S}_M} (p^2 + q^2) \cos^2(\theta_k - \varphi_{p,q})}{n_r(M^2 - 1)} \qquad (56)$$

where

$$\varphi_{p,q} \triangleq \tan^{-1}\left(\frac{q}{p}\right).$$

Using (50), the best θ_k, denoted by θ_k^*, is given by

$$\theta_k^* = \arg \max_{\theta_k \in [0, \frac{\pi}{4}]} \min_{(p,q) \in \mathbb{S}_M} (p^2 + q^2) \cos^2(\theta_k - \varphi_{p,q}). \qquad (57)$$

Following (51), the best achievable upper bound for P_k' is given by

$$P_k' \leq (M^2 - 1)c_k \left[\frac{(M^2 - 1)n_r}{3g_k(\theta_k^*, M)} \right]^{\delta_k^*} \gamma^{-\delta_k^*} + o(\gamma^{-\delta_k^*}). \qquad (58)$$

Remark 4: It is easily shown by the symmetry of the set $\mathbb{S}_M$ that it suffices to consider $\theta_k \in [0, \frac{\pi}{4}]$ for the maximization in (57). The min-max optimization problem does not have explicit analytical solutions except for small values of M, for example $M = 2$. But since the encoder matrices are fixed *a priori*, these computations can be performed off-line only once. ∎
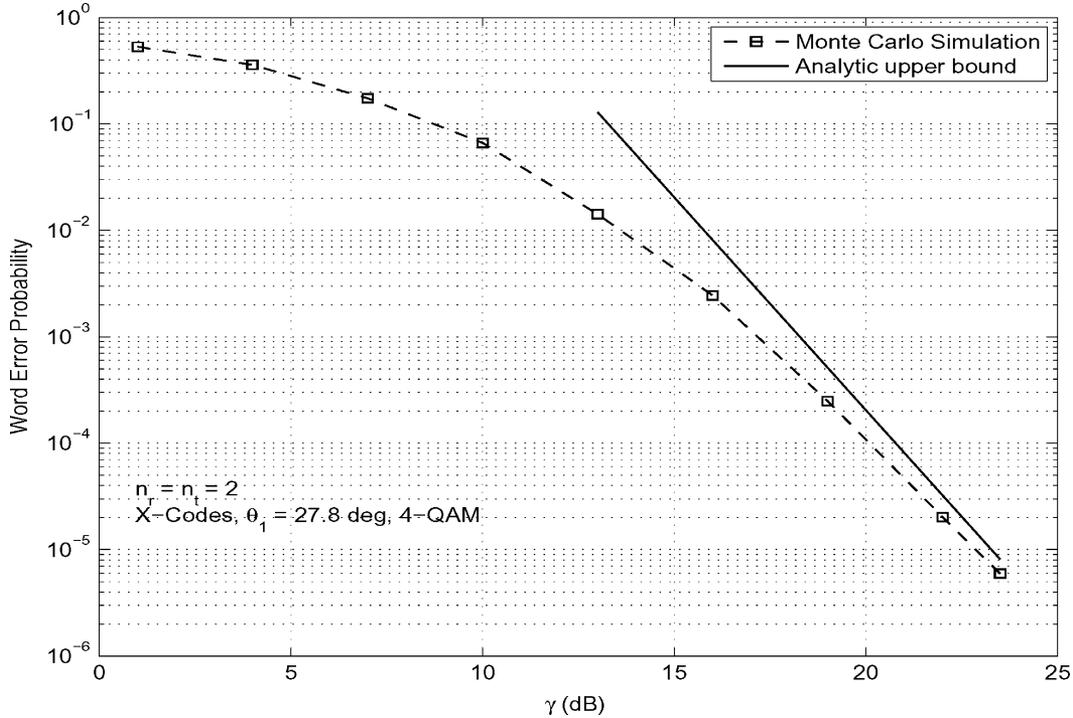
Fig. 4.   Union bound for word error probability. $n_r = n_t = 2$ and $M = 2$ (4-QAM) modulation.

For small MIMO systems, such as 2×2, it is possible to get a tighter upper bound by evaluating the expectation in (36) over both singular values. P_1' is then upper bounded as

$$P_1' \le \sum_{(p,q) \in \mathbb{S}_M} \frac{(70/81)(M^2-1)^4}{M^2(p\cos(\theta_1) + q\sin(\theta_1))^6(p^2+q^2)} \gamma^{-4} + o(\gamma^{-4}) \quad (59)$$

where θ_1 is the angle used for the only pair. For larger MIMO systems, it is preferable to use the inequality in (39) involving only one singular value, since evaluating the expectation containing two singular values becomes very tedious. In Fig. 4, we compare the word error probability of a 2×2 MIMO system with that given by (59), and observe that the union bound is indeed tight at high SNR.

In Fig. 5, we plot the variation of the upper bound to the WEP in (59) w.r.t. the angle θ_1 for the 2×2 MIMO system with 4-QAM and 16-QAM modulation. We observe that WEP is indeed sensitive to the rotation angle. With 4-QAM modulation, the WEP worsens as the angle approaches either 0 or 45 degrees. With 16-QAM modulation, the performance is even more sensitive to the rotation angle. Moreover, in addition to 0 and 45 degrees, we observe that the performance is poor, also when the angles are chosen near 18.5, 26.6 and 33.7 degrees, corresponding to $\varphi_{3,1}$, $\varphi_{2,1}$, and $\varphi_{3,2}$, respectively. We explain this as follows. From (36), it is clear that the error performance at high SNR is determined by the minimum value of the distance

$$d_k^2(p, q, \theta_k) \triangleq \|\mathbf{M}_k(\Re(\mathbf{u}_k) - \Re(\mathbf{v}_k))\|^2 \quad (60)$$

and we obtain $d_k^2(p, q, \theta_k)$ as

$$(p^2 + q^2)\left(\lambda_{i_k}^2 \cos^2(\theta_k - \varphi_{p,q}) + \lambda_{j_k}^2 \sin^2(\theta_k - \varphi_{p,q})\right)$$

when (p, q) runs over the set $\mathbb{S}_M$. If $\theta_k - \varphi_{p,q} = \pi/2$, i.e., $\theta_k = \tan^{-1}(-p/q)$ for some $(p, q) \in \mathbb{S}_M$, then the minimum distance is independent of λ_{i_k} and depends only upon λ_{j_k}. This implies a loss of diversity order since the diversity order of the square fading coefficient $\lambda_{j_k}^2$ is less than that of $\lambda_{i_k}^2$. For the case of $n_r = n_t = 2$, this would mean a reduction of diversity order from 4 to 1. The set $\mathbb{S}_M$ and the critical angles are illustrated in Fig. 3.

C. Optimal Design of X-Precoder

For X-Precoders, the optimal rotation angle is tedious to compute due to lack of exact expressions for the word error probability $P_k(\mathbf{H})$. Just like X-Codes, the union bound to $P_k'(\mathbf{H})$ is given by

$$P_k'(\mathbf{H}) \le \frac{1}{|\mathcal{S}_k|}$$
$$\times \sum_{\Re(\mathbf{u}_k)} \sum_{\Re(\mathbf{v}_k) \ne \Re(\mathbf{u}_k)} Q\left(\sqrt{\frac{d_k^2(\Re(\mathbf{u}_k), \Re(\mathbf{v}_k), \mathbf{A}_k)}{2N_0}}\right). \quad (61)$$

However, unlike the analysis for X-Codes, we do not further upper bound this union bound by using (39), since by doing so we would have lost information about λ_{j_k}. Instead, in the pairwise sum in (61), we look for the term with the highest contribution to the union bound and try to minimize it. The best angle for the k-th pair is then given by

$$\tilde{\theta}_k(\lambda_{i_k}, \lambda_{j_k}) = \arg \max_{\theta_k \in [0, \frac{\pi}{4}]} \min_{(p,q) \in \mathbb{S}_M} d_k^2(p, q, \theta_k) \quad (62)$$

where $d_k^2(p, q, \theta_k)$ is given by (60). Just like X-Codes, it can be shown that for the maximization in (62), it suffices to consider the range $[0, \frac{\pi}{4}]$ for θ_k, instead of the entire range $[0, 2\pi]$. The
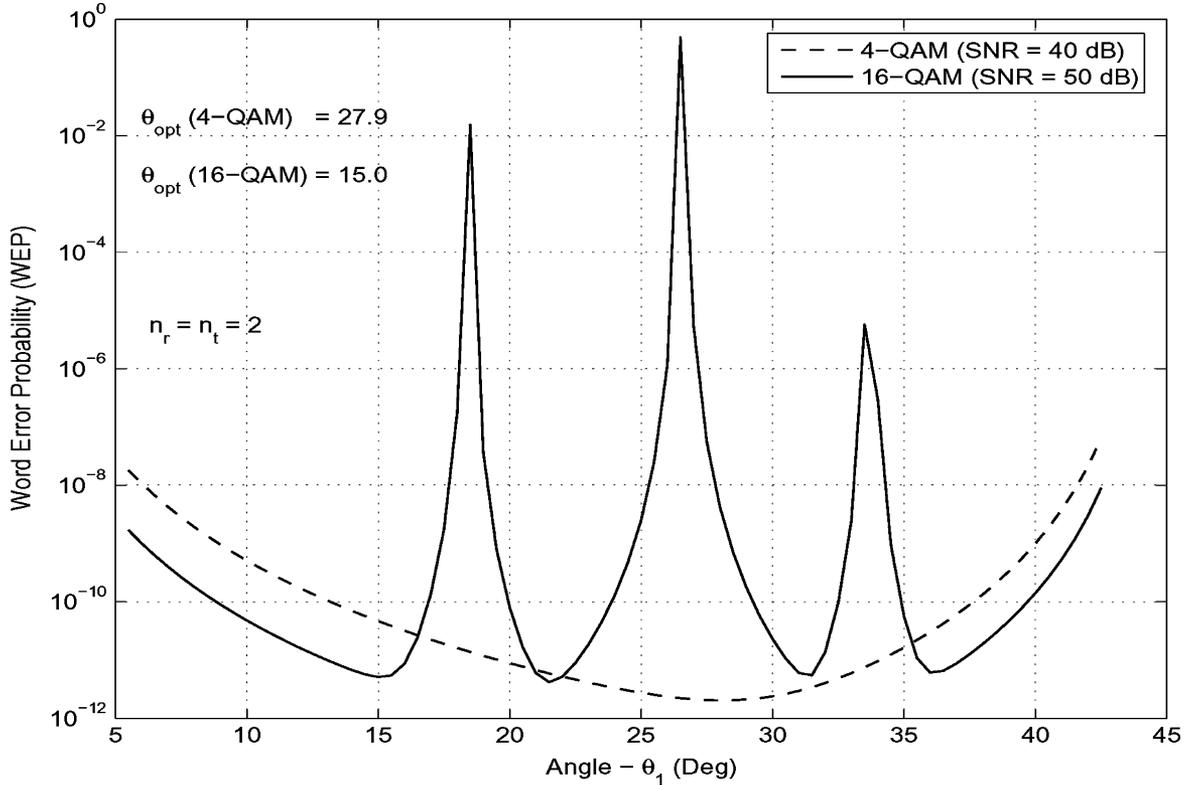
Fig. 5. Sensitivity of word error probability w.r.t θ_1. $n_r = n_t = 2$ and $M = 2, 4$ (4,16-QAM) modulation.

optimization problem in (62) is analytically tractable only for small values of M. Also, the minimization over $(p, q) \in \mathbb{S}_M$ need not be over the full set containing $|\mathbb{S}_M| = 4M(M-1)$ elements. In fact, it can be shown that the number of elements to be searched is at most $(M^2 - 3M + 6)/2$. For example, for $M = 4$ (16-QAM), we need to search only five elements instead of the full set of 48 elements.

Theorem 3: For $M = 2$ (4-QAM), the exact $\tilde{\theta}_k(\lambda_{i_k}, \lambda_{j_k})$ is given by

$$\begin{cases} \pi/4 & \beta_k \leq \sqrt{3} \\ \tan^{-1}\left[(\beta_k^2 - 1) - \sqrt{(\beta_k^2 - 1)^2 - \beta_k^2}\right] & \beta_k > \sqrt{3} \end{cases}$$
(63)

where $\beta_k \triangleq \frac{\lambda_{i_k}}{\lambda_{j_k}}$ is the condition number for the k-th pair.

Proof: See Appendix C. ∎

Further let

$$\tilde{d}_{k,\min}^2(\lambda_{i_k}, \lambda_{j_k}) \triangleq \max_{\theta_k \in [0, \frac{\pi}{4}]} \min_{(p,q) \in \mathbb{S}_M} d_k^2(p, q, \theta_k)$$
(64)

then using (64) and (61), the truncated union bound to P_k' is given by

$$P_k' \leq (M^2 - 1)\mathbb{E}\left[Q\left(\sqrt{\frac{\tilde{d}_{k,\min}^2(\lambda_{i_k}, \lambda_{j_k})}{2N_0}}\right)\right].$$
(65)

The expectation in (65) is over the joint distribution of $(\lambda_{i_k}, \lambda_{j_k})$ and is difficult to compute analytically. We therefore use Monte-Carlo simulations to evaluate the exact error probability P_k'.

V. Y-CODES AND Y-PRECODER

A. Motivation

As we will see in Section VI (see Fig. 6), the bit error probability performance of X-Codes is better than the one of the other precoders when the condition number for a pair of subchannels is small. However, the bit error probability performance of X-Codes degrades with increasing condition number. Since, typical fading channels are ill conditioned[7] with high probability, it is necessary to design precoders which are robust to ill-conditioned channels. Also, ML detection for X-Codes involves a 2-D search, which is slightly more complex than the linear precoders reported in [7], [8] and [9], for which ML detection involves only a 1-D search.

Therefore, we have two important problems to be solved: *i)* improvement in error performance for ill conditioned channels and *ii)* reduction in ML detection complexity. We firstly address the issue of performance improvement in ill conditioned channels. Towards this end, we ask ourselves the following question: "For a given transmit power constraint P_T and rate R, what is the best possible code design in terms of achieving the minimum average bit/symbol/word error rate?". It is not easy to find the best possible code in closed form, but based upon analysis we

[7]A $n_t \times n_r$ MIMO channel ($n_r \leq n_t$) is said to be ill conditioned if its condition number is large, i.e., $\lambda_1/\lambda_{n_r} \gg 1$.
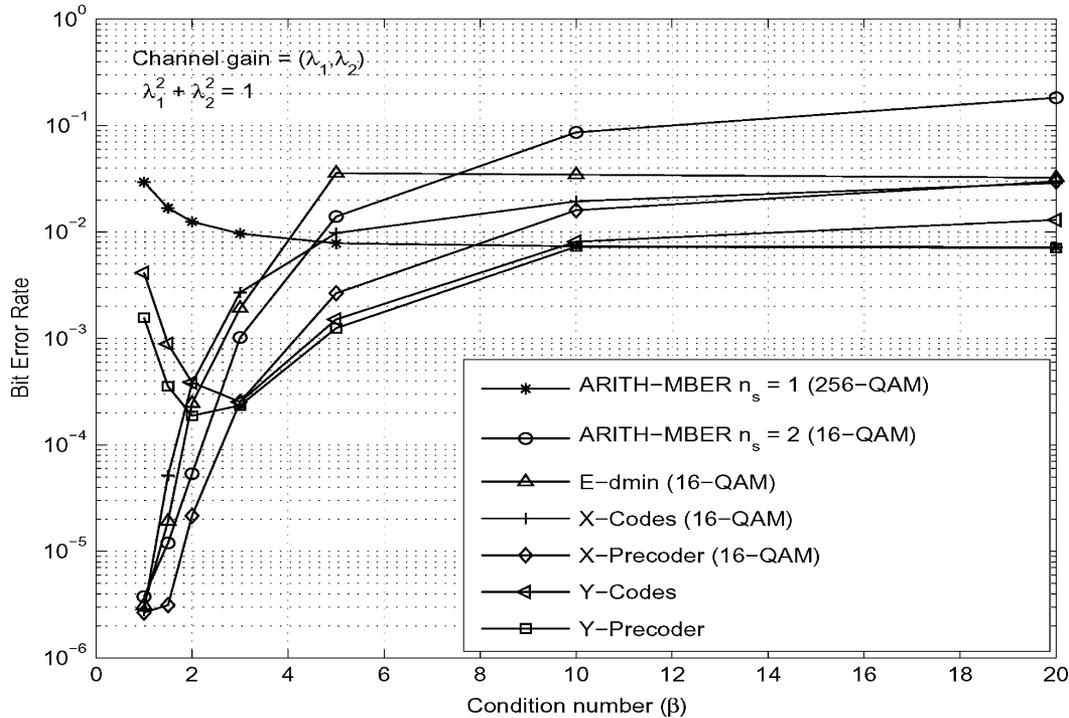
Fig. 6.   Effect of the channel condition number on error performance of various precoders for a 2×2 system with target rate $R = 8$ bpcu.

can definitely gain insight into the properties that a good code must have.

It is observed that the error performance at high SNR is dependent on the minimum value of the pairwise distance $d_k^2(\Re(\mathbf{u}_k), \Re(\mathbf{v}_k), \mathbf{A}_k)$ [see (37)] over all possible information vectors $\mathbf{u}_k \neq \mathbf{v}_k$. Using the definition, we have

$$d_k^2(\Re(\mathbf{u}_k), \Re(\mathbf{v}_k), \mathbf{A}_k) = \|\mathbf{M}_k(\Re(\mathbf{u}_k) - \Re(\mathbf{v}_k))\|^2$$
$$= \lambda_{i_k}^2 e_{k,1}^2 + \lambda_{j_k}^2 e_{k,2}^2 \qquad (66)$$

where $\mathbf{e}_k \triangleq \mathbf{A}_k(\Re(\mathbf{u}_k) - \Re(\mathbf{v}_k))$.

Let $\beta_k = \lambda_{i_k}/\lambda_{j_k}$ be the condition number for the k-th pair as defined in Theorem 3, then we have $\beta_k \geq 1$, since $\lambda_{i_k} \geq \lambda_{j_k}$. For the special case of $\beta_k = 1, d_k^2(\Re(\mathbf{u}_k), \Re(\mathbf{v}_k), \mathbf{A}_k)$ is proportional to $\|\mathbf{e}_k\|^2$, which is the Euclidean distance between the code vectors $\Re(\mathbf{u}_k)$ and $\Re(\mathbf{v}_k)$. Therefore, for $\beta_k = 1$, the design of good codes is independent of the subchannel gains $(\lambda_{i_k}, \lambda_{j_k})$. However, the design of good codes becomes harder for values of $\beta_k > 1$. We immediately observe that, since $\lambda_{i_k} > \lambda_{j_k}$, the effective Euclidean distance in (66) gives more weight to the term $e_{k,1}^2$, which is the squared difference of the vectors along the stronger subchannel component. Since the total transmit power is constrained, codes should be designed such that the minimum possible separation between any two code vectors is larger along the stronger subchannel as compared to the minimum possible separation along the weaker subchannel.

Hence, it is obvious that X-Codes (based on 2-D rotation matrices) may not be a good code design for ill conditioned subchannels, where $\beta_k \gg 1$. This is because, with rotated QAM

constellations, the codewords achieve the same nonzero minimum distance along both the stronger as well as the weaker subchannel. Specially in cases where the condition number is large, it is intuitive that a code design which achieves maximal nonzero minimum distance along the stronger subchannel and zero minimal distance along the weaker subchannel, would perform better than the best rotated constellation. This has been illustrated in Figs. (1) and (2), where it can be seen that, when compared to X-Codes, Y-Codes achieve a larger minimum distance between the received codewords.

This insight leads us to design codes, which have a zero minimum distance along the weaker subchannel so that, under a fixed transmit power constraint, more separation between the codewords can be achieved along the stronger subchannel. A simple design is to have the code vectors belong to a subset of some skewed 2-D lattice, an example of which is shown in Fig. 7 (there are totally eight code vectors represented by small filled circles). Since the code vectors belong to a lattice, they can be expressed as a linear transformation of a subset of $\mathbb{Z}^2$. The simple structure of the code results in a very simple ML detector for each subchannel pair, which has a detection complexity of the same order as that of a 1-dimensional scalar channel (like the linear precoders in [8] and [9]). It is also noted that, for a code with M vectors, the transmitted code vector assumes only two possible amplitude values along the weaker subchannel component, and $M/2$ different values along the stronger subchannel component. This is in fact a simple rate allocation scheme, where only 1 information bit is transmitted through the weaker subchannel, and the remaining bits are transmitted through the stronger subchannel. More complex
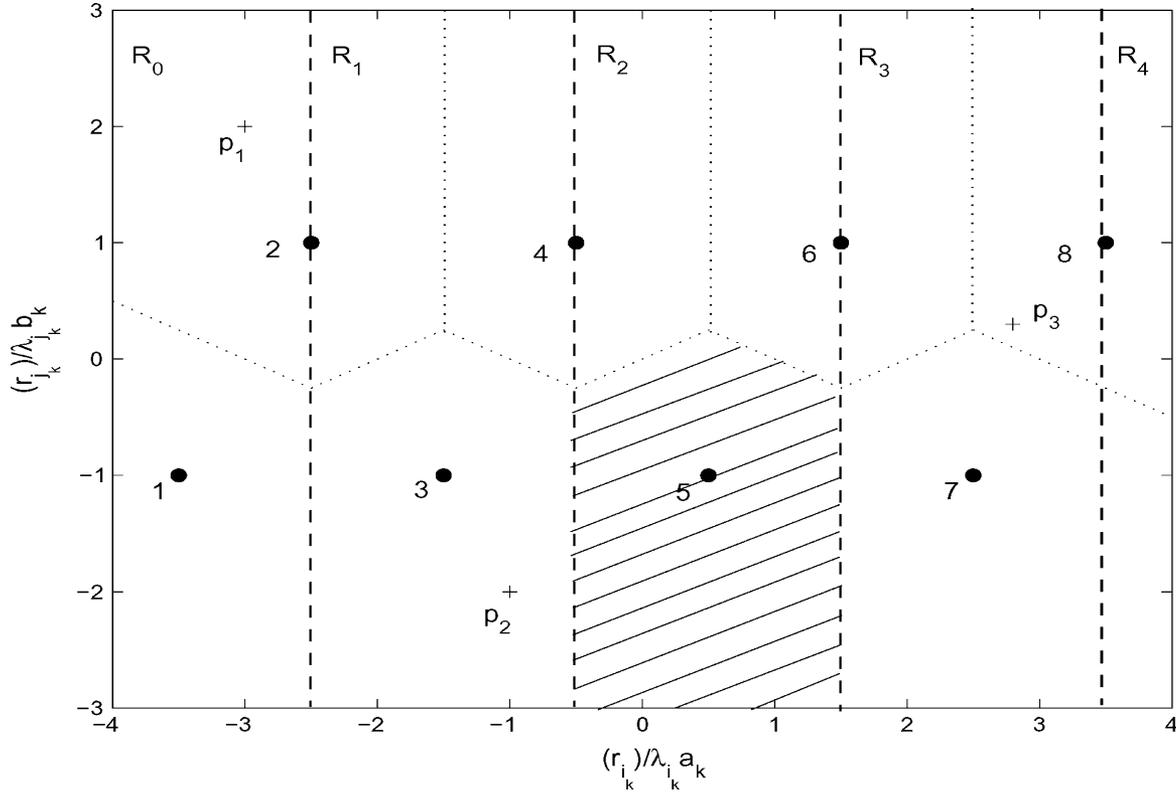
Fig. 7.   Received signal space for the real component of the k-th pair. With Y-Codes ($M = 8$), we have 5 regions separated by vertical dashed lines. The scaled codebook vectors are represented by small filled circles along with their corresponding codebook index number. Dotted lines demarcate the boundary between the ML decision regions.

rate allocation schemes are possible, but would result in more complex ML detectors.

B. *Y-Codes and Y-Precoders: Encoding*

For Y-Codes and Y-Precoders, the matrices $\mathbf{A}_k$ have the structure

$$\mathbf{A}_k = \begin{bmatrix} a_k & 2a_k \\ 2b_k & 0 \end{bmatrix} \tag{67}$$

where $a_k, b_k \in \mathbb{R}^+$. For Y-Codes/Precoders, the set $\mathcal{S}_k$ is given by the Cartesian product

$$\mathcal{S}_k \triangleq \{0, 1\} \times \left\{ 0, \ldots, \frac{M}{2} - 1 \right\}. \tag{68}$$

For example, with $M = 4$, the set $\mathcal{S}_k$ is given by

$$\mathcal{S}_k = \{(0,0)^T, (0,1)^T, (1,0)^T, (1,1)^T\}. \tag{69}$$

The real and imaginary components of the displacement vector for the k-th pair, $\mathbf{u}_k^0$ are given by

$$\Re(\mathbf{u}_k^0) = \Im(\mathbf{u}_k^0) = \left[-\frac{(M-1)a_k}{2}, -b_k \right]^T. \tag{70}$$

We consider the 2-D codebook of cardinality M generated by applying $\mathbf{A}_k$ to the elements of $\mathcal{S}_k$ and adding the displacement vector. The M code vectors of the 2-D codebook are given by

$$Y_k(v) = \left[a_k \left((v-1) - \frac{M-1}{2} \right), b_k(-1)^v \right]^T \tag{71}$$

where $v = 1, \ldots, M$.

With the codebook notation, v refers to the index of the code vector $Y_k(v)$ in the codebook. Further, let the codebook indices of the vectors, to be transmitted on the real and imaginary components of the k-th pair be v_k^I and v_k^Q respectively. The components of the data symbol vector $\mathbf{z}$ are then given by

$$\begin{aligned} (\Re(z_{i_k}), \Re(z_{j_k}))^T &= Y_k\left(v_k^I\right) \\ (\Im(z_{i_k}), \Im(z_{j_k}))^T &= Y_k\left(v_k^Q\right) \end{aligned} \tag{72}$$

where $k = 1, 2, \ldots, n_r/2$.

Due to the transmit power constraint in (15), a_k and b_k must satisfy

$$b_k^2 + a_k^2 \frac{M^2 - 1}{12} = \frac{P_T}{n_r}. \tag{73}$$

The only difference between Y-Codes and Y-Precoders is that, for Y-Codes, the parameters a_k and b_k are fixed *a priori*, whereas, for the Y-Precoders, these are chosen every time the channel changes.

C. Y-Codes and Y-Precoders: ML Decoding

Using our codebook notation in (71), the ML decoding rule in (20) and (21), can be equivalently written as

$$\hat{v}_k^I = \arg \min_{v \in \{0,\ldots,(M-1)\}} \|\Re(\mathbf{r}_k) - \Lambda_k Y_k(v)\|^2$$

$$\hat{v}_k^Q = \arg \min_{v \in \{0,\ldots,(M-1)\}} \|\Im(\mathbf{r}_k) - \Lambda_k Y_k(v)\|^2 \quad (74)$$

where $\hat{v}_k^I$ and $\hat{v}_k^Q$ are ML estimates of the codeword indices transmitted on the real and imaginary components for the k-th pair.

We next discuss a low complexity solution for the detection problem in (74). The algorithm is the same for all pairs, and the same for both the real and imaginary components of each pair. Therefore, we only discuss the algorithm for the real component of the k-th pair. We first partition the 2-D received signal space ($\mathbb{R}^2$) into ($\frac{M}{2} + 1$) regions as follows:

$$R_0 : \left\{ [x,y]^T \in \mathbb{R}^2 \Big| -\infty \leq \left(\frac{x}{\lambda_{i_k} a_k} + \frac{M-1}{2} \right) \leq 1 \right\}$$

$$R_{\frac{M}{2}} : \left\{ [x,y]^T \in \mathbb{R}^2 \Big| (M-1) \right.$$
$$\left. \leq \left(\frac{x}{\lambda_{i_k} a_k} + \frac{M-1}{2} \right) \leq +\infty \right\}$$

$$R_i : \left\{ [x,y]^T \in \mathbb{R}^2 \Big| (2i-1) \right.$$
$$\left. \leq \left(\frac{x}{\lambda_{i_k} a_k} + \frac{M-1}{2} \right) \leq (2i+1) \right\} \quad (75)$$

where $i \in [1, M/2 - 1]$.

In Fig. 7, we illustrate the five regions with $M = 8$ for the real component of the k-th pair. The first step of the decoding algorithm is to find the region to which the received vector belongs. Let

$$t_k = \left\lfloor \frac{\Re(r_{i_k})}{2\lambda_{i_k} a_k} + \frac{M+1}{4} \right\rfloor.$$

The received vector belongs to the region R_{ζ_k}, where ζ_k is explicitly given by

$$\zeta_k = \begin{cases} 0 & t_k \leq 0 \\ \frac{M}{2} & t_k \geq \frac{M}{2} \\ t_k & \text{otherwise.} \end{cases} \quad (76)$$

For example, in Fig. 7, the received vectors $\mathbf{p}_1, \mathbf{p}_2$, and $\mathbf{p}_3$ belong to R_0, R_1, and R_3, respectively. It can be shown that, once the received vector is decoded to the region R_{ζ_k}, the ML code vector is one among a reduced set of at most 3 code vectors. Therefore, at most 3 Euclidean distances need to be computed to solve the ML detection problem in (74), as compared to computing all the M Euclidean distances in case of a brute force search. For example, in Fig. 7, for the received vector $\mathbf{p}_3 \in R_3$, the ML code vector is among $Y_k(6), Y_k(7)$ or $Y_k(8)$.

However, once we know the region of the received vector, it is possible to directly find the ML code vector even without computing the three Euclidean distances. This involves just checking at most 3 linear relations between the two components of the received vector. Therefore, the ML decoding complexity

of Y-Codes is the same as that of a scalar channel. For example, in Fig. 7, the received vector $\mathbf{p}_3$ is to the right of the perpendicular bisector between $Y_k(6)$ and $Y_k(8)$. The vector $\mathbf{p}_3$ is also above the perpendicular bisector between $Y_k(7)$ and $Y_k(8)$. From these two checks, it can be easily concluded that the ML code vector is $Y_k(8)$. Due to the structure of the codebook, the ML decision regions can be very easily outlined. In Fig. 7, the dotted lines demarcate the boundary of the ML decision regions. The hatched region illustrates the ML decision region of $Y_k(5)$.

D. Optimal Design of Y-Codes

Given the optimal pairing in (48), the next step towards designing optimal Y-Codes is to find the optimal value of (a_k, b_k), which minimizes the average error probability P_k' for the k-th pair. For Y-Codes, once chosen, (a_k, b_k) are fixed and do not change with every channel realization. Since the ML decision regions are known precisely, it is possible to calculate the exact error probability P_k'. With our codebook notation, we identify code vectors by their index in the codebook. For the k-th pair, the average word error probability of the real component can therefore be expressed in terms of the code vector indices as

$$P_k' = \frac{1}{M} \sum_v P_k'(v) \quad (77)$$

where $P_k'(v) \triangleq \mathbb{E}_{\mathbf{H}}[\text{Prob}(\hat{v}_k^I \neq v \mid Y_k(v), \mathbf{H})]$, $P_k'(v)$ is the average probability of error when the code vector $Y_k(v)$ is transmitted and it is given by

$$P_k'(v) = \begin{cases} \mathbb{E}[g_1(a_k, b_k)] & 3 \leq v \leq (M-2) \\ \mathbb{E}[g_2(a_k, b_k)] & v = 1, M \\ \mathbb{E}[g_1(a_k, b_k) - g_3(a_k, b_k)] & v = 2, (M-1) \end{cases} \quad (78)$$

where all the expectations are over the joint distribution of $(\lambda_{i_k}, \lambda_{j_k})$. Let

$$\Psi_k(x) \triangleq \frac{\sqrt{2}(2a_k \lambda_{i_k} x - a_k^2 \lambda_{i_k}^2 - 4b_k^2 \lambda_{j_k}^2)}{4 b_k \lambda_{j_k} \sqrt{N_0}}$$

$$\Phi_k(x) \triangleq \frac{-\sqrt{2}(2a_k \lambda_{i_k} x + a_k^2 \lambda_{i_k}^2 + 4b_k^2 \lambda_{j_k}^2)}{4 b_k \lambda_{j_k} \sqrt{N_0}}.$$

The functions $g_1(a_k, b_k), g_2(a_k, b_k)$ and $g_3(a_k, b_k)$ are given by

$$g_1(a_k, b_k) \triangleq 1 - \int_0^{\lambda_{i_k} a_k} \frac{2 e^{\frac{-x^2}{N_0}}}{\sqrt{\pi N_0}} Q(\Psi_k(x)) dx$$

$$g_2(a_k, b_k) \triangleq 1 - \int_{-\infty}^{\lambda_{i_k} a_k} \frac{e^{\frac{-x^2}{N_0}}}{\sqrt{\pi N_0}} Q(\Psi_k(x)) dx$$

$$g_3(a_k, b_k) \triangleq \int_{-\infty}^{-\lambda_{i_k} a_k} \frac{e^{\frac{-x^2}{N_0}}}{\sqrt{\pi N_0}} Q(\Phi_k(x)) dx. \quad (79)$$

To compute the optimal (a_k, b_k), we have to minimize P_k' w.r.t. (a_k, b_k) subject to the transmit power constraint in (73). However, it is difficult to get closed form expressions for the optimal (a_k, b_k) due to the intractability of the integrals in (79). This difficulty is further compounded due to the evaluation of

expectation over the joint distribution of $(\lambda_{i_k}, \lambda_{j_k})$. However, since (a_k, b_k) are fixed *a priori*, it is always possible to approximately compute the optimal (a_k, b_k) off-line, using Monte-Carlo techniques.

E. Optimal Design of Y-Precoder

For the Y-Precoder, finding the optimal (a_k, b_k) which minimizes $P_k'(\mathbf{H})$ for a given channel realization $\mathbf{H}$ is again difficult due to the intractability of the integrals in (79). In the case of Y-Codes these could be computed off-line since (a_k, b_k) are fixed *a priori*. However, for Y-Precoders these cannot be computed off-line, since the optimal (a_k, b_k) have to be computed every time the channel changes.

To reduce the complexity of computing (a_k, b_k) adaptively as the channel changes, we resort to simpler bounds on the error probability $P_k'(\mathbf{H})$ for which the optimal (a_k, b_k) require minimal computation. One such bound is given by the union bounding technique, which was also used in the analysis of the error performance for X-Precoders. For Y-Precoders, the union bound to $P_k'(\mathbf{H})$ is given by

$$P_k'(\mathbf{H}) \leq (M-1)Q\left(\sqrt{\frac{d_{k,\min}^2(a_k, b_k)}{2N_0}}\right) \quad (80)$$

where $d_{k,\min}^2(a_k, b_k)$ is given by

$$\min_{v \neq w} \left(\lambda_{i_k}^2 a_k^2 (v-w)^2 + \lambda_{j_k}^2 b_k^2 ((-1)^v - (-1)^w)^2\right) \quad (81)$$

where v and w are distinct indices of the codebook. A good choice of (a_k, b_k) is obtained by minimizing the union bound for $P_k'(\mathbf{H})$ in (80) w.r.t (a_k, b_k), subject to the transmit power constraints in (73). From (80), this choice of (a_k, b_k), denoted by (a_k^*, b_k^*), maximizes $d_{k,\min}^2(a_k, b_k)$ for the fixed channel gain of $(\lambda_{i_k}, \lambda_{j_k})$. The next theorem gives the closed form analytical expression for (a_k^*, b_k^*).

Theorem 4: A good choice of (a_k, b_k) defined as

$$(a_k^*, b_k^*)$$
$$\triangleq \arg \max_{\{(a_k, b_k) \in (\mathbb{R}^+)^2 \mid b_k^2 + a_k^2 \frac{M^2-1}{12} = \frac{P_T}{n_r}\}} d_{k,\min}^2(a_k, b_k) \quad (82)$$

is given by (83), as shown at the bottom of the page, where $M' = \frac{M^2-1}{9}$.

The corresponding optimal value of $d_{k,\min}^2(a_k, b_k)$ is given by

$$d_{k,\min}^2(a_k^*, b_k^*) = \begin{cases} \frac{12P_T \lambda_{i_k}^2}{n_r(M^2-1)} & \beta_k^2 \geq \frac{M^2-1}{3} \\ \frac{16P_T \lambda_{i_k}^2}{n_r\left(3\beta_k^2 + \frac{(M^2-1)}{3}\right)} & \beta_k^2 < \frac{M^2-1}{3}. \end{cases} \quad (84)$$

Proof: See Appendix D. ∎

If we now look back at the codebook for Y-Precoders, we notice that there is power allocation on the two subchannels through the parameters a_k and b_k, which can be chosen optimally based upon the knowledge of channel gains. From (83), we observe that the Y-Precoders use only the stronger subchannel, when channel condition number is high ($\beta_k^2 \geq \frac{M^2-1}{3}$). For good channel condition ($\beta_k^2 < \frac{M^2-1}{3}$), power is distributed between the two channels depending on the channel condition. This adaptive power allocation between the stronger and the weaker subchannels, enables Y-Precoders to achieve an error performance better than X-Precoders in ill conditioned channels. Y-Codes/Y-Precoders also have a fixed-rate allocation between the two subchannels of a pair, since out of the $\log_2(M)$ bits, one bit can be used to decide whether the vector in the codebook is at even index (corresponding to the second component being equal to $+b_k$) or at odd index (corresponding to the second component being equal to $-b_k$). The remaining bits are then used to appropriately choose among the vectors at even or odd indices. Therefore, in a way, the proposed Y-Codes always transmits 1 bit of information on the weaker subchannel and $\log_2(M) - 1$ bits on the stronger subchannel. This rate allocation may not be the best and therefore even better code books can be constructed, but may result in higher ML detection complexity. The ML decoding complexity for the proposed Y-Codes/Y-Precoders is low and is independent of M.

VI. SIMULATION RESULTS

In this section, we compare the error probability performance of X-, Y-Codes and X-, Y-Precoders with other precoders proposed in literature. For all the simulations, we assume $n_r = n_t$. Throughout this section, the subchannel pairing used for X-,Y-Codes/Precoders is given by (48).

For X- and Y-Codes, the encoding matrices for each pair, $\mathbf{A}_k, k = 1, 2, \ldots, n_r/2$ are fixed *a priori*. In case of X-Codes, the fixed rotation angles are given by (57). For Y-Codes, the value of (a_k, b_k) for the k-th pair of subchannels, is computed, by minimizing P_k' in (77) w.r.t. (a_k, b_k). Note that the expectation in (78), is carried out using Monte-Carlo techniques, assuming a Rayleigh flat fading MIMO channel.

For X-,Y-Precoders, the encoding matrices for each pair, $\mathbf{A}_k, k = 1, 2, \ldots, n_r/2$ are not fixed, and are computed every time the channel changes. For X-Precoders, the rotation angle for each of the $n_r/2$ subchannel pairs is given by (62). For Y-Precoders, (a_k, b_k) for the k-th pair of subchannels is given by Theorem 4.

Error performance comparison of the proposed X-,Y-Codes/ Precoders is made with 1) the E-dmin (equal dmin precoder proposed in [12]), 2) the Arithmetic mean BER precoder (ARITH-MBER) proposed in [8], 3) the Equal Error linear precoder (EE) (based upon optimizing the minimum eigenvalue

$$(a_k^*, b_k^*) = \begin{cases} \left(\sqrt{\frac{12P_T}{n_r(M^2-1)}}, 0\right) & \beta_k^2 \geq \frac{M^2-1}{3} \\ \left(\sqrt{\frac{4P_T}{3n_r(\beta_k^2+M')}}, \beta_k\sqrt{\frac{P_T}{n_r(\beta_k^2+M')}}\right) & \beta_k^2 < \frac{M^2-1}{3} \end{cases} \quad (83)$$

for a given transmit power constraint [9]), 4) the TH precoder (based upon the idea of Tomlinson-Harashima precoding applied in the MIMO context [7]), and 5) the channel inversion (CI) precoder, also known as the Zero Forcing precoder [4].

A. Effect of Channel Condition Number on Error Performance

In Fig. 6, we plot the bit error probability/rate (BER) of all precoding schemes for a 2×2 MIMO at $\gamma = 26$ dB, as a function of the condition number $\beta = \lambda_1 / \lambda_2$. We fix the total channel gain to be 1, i.e., $\lambda_1^2 + \lambda_2^2 = 1$, and the target rate to be $R = 8$ bpcu. For a given β, the BER of all precoders is evaluated through Monte-Carlo simulations, for a fixed 2×2 MIMO channel with channel matrix, $\mathbf{H} = \mathrm{diag}(\beta/\sqrt{\beta^2+1}, 1/\sqrt{\beta^2+1})$.

We briefly discuss the precoding schemes which are compared to the proposed X-,Y-Codes/Precoders. ARITH-MBER transmits $n_s \leq 2$ symbols, each from a QAM modulation alphabet. For the ARITH-MBER precoder, when $n_s = 1$, every channel use, a 256-QAM modulation symbol (16-PAM on the real and imaginary component) is transmitted on the stronger subchannel and the weaker subchannel is not used for transmission (note that for a 2×2 system, there are only two subchannels, i.e., only 1 pair of subchannels). When $n_s = 2$, two 16-QAM modulation symbols are transmitted, one on each subchannel. E-dmin is a precoding scheme in which the complex linear precoding matrix is adapted to each channel realization, but both the subchannels are always used for transmission (i.e., $n_s = 2$). With E-dmin, the modulation alphabet used is 16-QAM.

In Fig. 6, we notice that schemes which are fixed and do not adapt their transmission with the varying channel, have good BER performance for small values of β. The BER performance is, however, poor with increasing β. BER performance of X-Codes is also seen to deteriorate with increasing β. The only exceptions are Y-Codes and ARITH-MBER ($n_s = 1$). For ARITH-MBER with $n_s = 1$, only the stronger subchannel is used for transmission, and therefore its BER improves with increasing β, because, with $\lambda_1^2 + \lambda_2^2 = 1, \lambda_1$ (i.e., the stronger subchannel gain) increases with increasing β. Compared to X-Codes/X-Precoders, the BER performance of Y-Codes/Y-Precoders is more stable with increasing β due to the fact that the codebook is designed in order to maximize the minimum codeword distance along the stronger subchannel component without caring about the separation on the weaker subchannel component.

It is also observed that, the Y-Precoders/Y-Codes perform better than X-Precoders/X-Codes for large β, and hence for channels which are ill conditioned, Y-Precoders/Y-Codes have a better error performance compared to X-Precoders/X-Codes. We shall see later that, indeed for the Rayleigh fading channel, Y-Precoders/Y-Codes perform better than X-Precoders/X-Codes. This justifies the fact that precoders for ill conditioned channels should be designed to achieve more separation in the minimum Euclidean distance (between received codewords) along the component corresponding to the stronger subchannel. It can also be seen that, despite the suboptimal

choice of encoding matrices $\mathbf{A}_k$ (based upon the union bound to the error probability of $\mathbf{u}_k$), X-Precoders perform better than X-Codes, and Y-Precoders perform better than Y-Codes.

B. Diversity Order Comparison

We next discuss the overall diversity order achieved by the various precoding schemes with a Rayleigh flat fading MIMO channel. Let the number of subchannels used for transmission be n_s ($n_s \leq n_r$). The diversity order achieved by the linear precoders (EE and ARITH-MBER) and THP is $(n_r - n_s + 1)(n_t - n_s + 1)$ and $(n_t - n_s + 1)$ respectively, whereas the diversity order achieved by E-dmin and X-, Y-Codes/Precoders is $(n_r - \frac{n_s}{2} + 1)(n_t - \frac{n_s}{2} + 1)$. The CI scheme achieves infinite diversity, but it suffers from power enhancement at the transmitter. Among all the other schemes (except CI), we observe that E-dmin and X-, Y-Codes/Precoders have the best overall diversity order.

The subsequent simulation results assume a MIMO channel with Rayleigh flat fading statistics. All BER performance reported are averaged over the channel fading statistics through Monte-Carlo simulations.

C. Comparison of BER Performance $n_s = \min(n_r, n_t)$

In Fig. 8, we plot the BER performance of all precoders for $n_r = n_t = n_s = 2, 4$ and a target rate of $R = 2n_s$ bpcu. The proposed X, Y-Precoders and E-dmin have the best error performance. The increased diversity order achieved by the pairing scheme is obvious from the higher diversity slope of the error probability for the X, Y-Precoders compared to a diversity slope of order 1 for the linear precoder ARITH-MBER and THP. This observation supports the analytical result on the improvement of diversity order of the error performance by pairing of MIMO subchannels (see Sections III-B, III-C). The performance of CI is inferior due to enhanced transmit power requirement arising from the bad conditioning of the channel. It is observed that the proposed Y-Precoders perform the best for $n_r = n_t = 2$, with E-dmin only 0.5 dB away at a BER of 10^{-3}. For $n_r = n_t = 4$, E-dmin performs better than Y-Precoders by 0.4 dB at a BER of 10^{-3}. However, E-dmin has this performance gain at a higher encoding and decoding complexity compared to the Y-Precoder.

D. Comparison of BER Performance for $n_r = n_t = 2, 4, n_s \leq \min(n_r, n_t)$

In this section, we report the observations made on the simulated BER performance of various precoding schemes for a given target rate of R bpcu and achievable overall diversity gain $\delta_{ord} = (n_r/2 + 1)^2$. This diversity gain is achieved by X-,Y-Precoder and E-dmin with $n_s = n_r$ information symbols transmitted every channel use, and by the linear precoders (EE and ARITH-MBER) with $n_s = n_r/2$ information symbols transmitted every channel use. Therefore, for the same target rate R, the information symbols transmitted by the linear precoders must belong to a modulation alphabet with $2^{2n_r/R}$ elements (we select n_r and R such that n_r/R is an integer), as compared to a $2^{n_r/R}$ element modulation alphabet set required for X-,Y-Precoder and E-dmin.
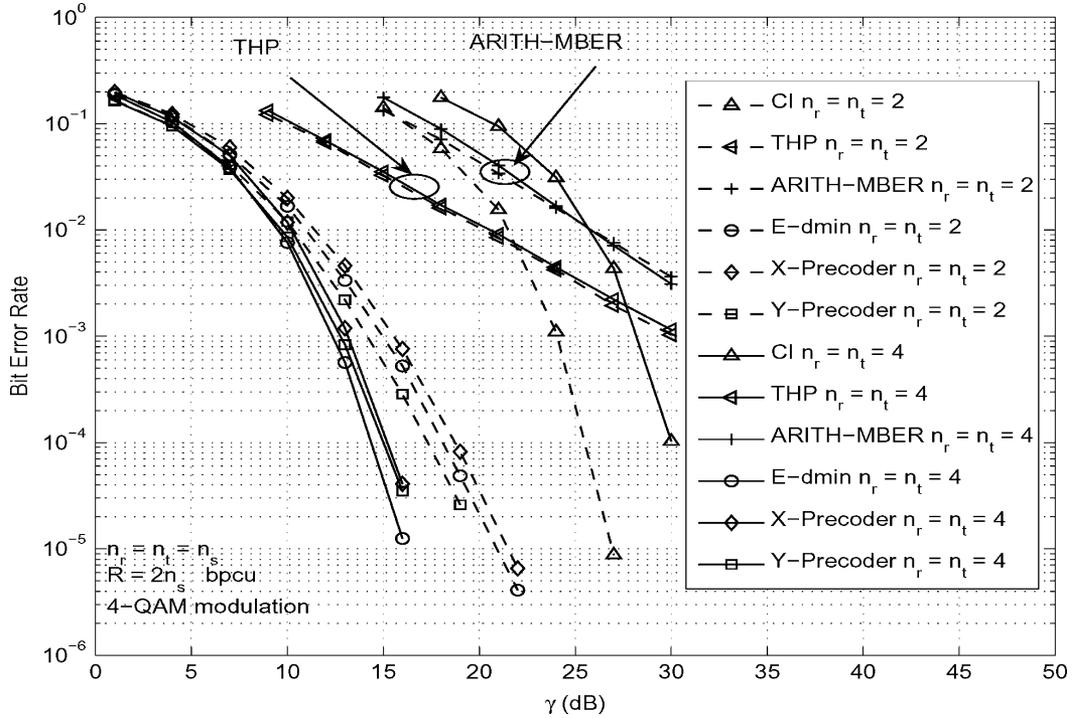
Fig. 8. BER comparison between various precoders for $n_r = n_t = n_s = 2, 4$ and $M = 2$ (4-QAM) modulation. Target rate is $R = 2n_s$ bpcu.
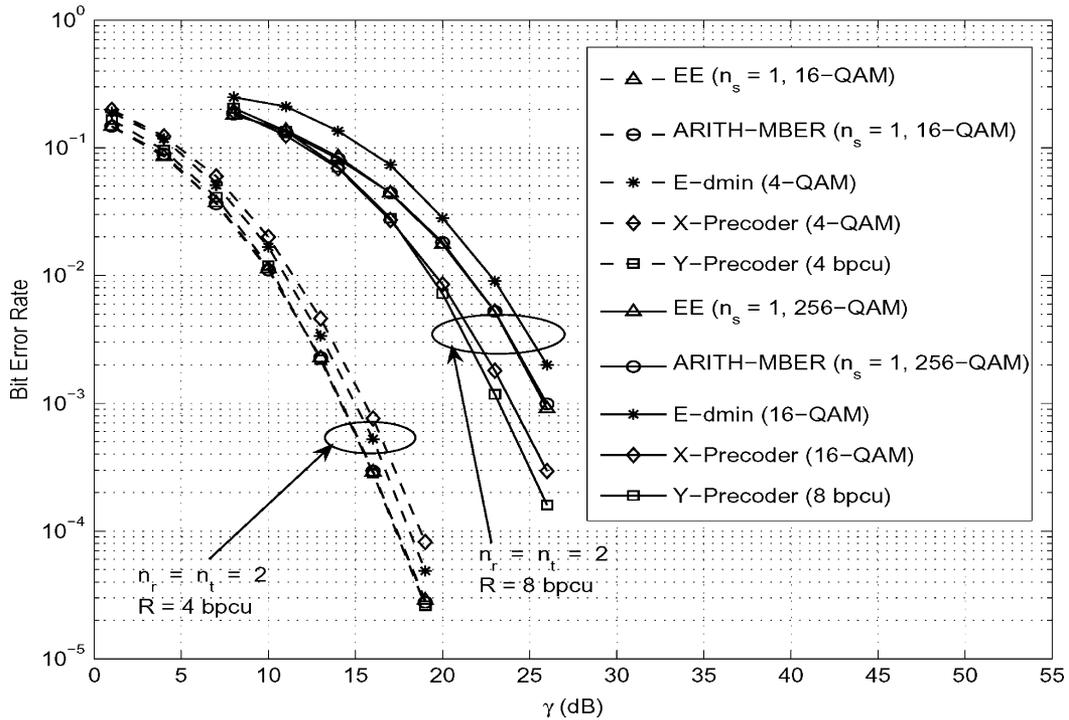

Fig. 9. BER comparison between various precoders for $n_r = n_t = 2$ and target rate $R = 4, 8$ bpcu.

In Fig. 9, we plot the BER for $n_r = n_t = 2$, and a target rate of $R = 4, 8$ bpcu. It is observed that the best performance is achieved by the proposed Y-Precoder. For a target rate of $R = 4$ bpcu, ARITH-MBER (16-QAM) also has a similar performance. However, for a rate of $R = 8$ bpcu, the performance of ARITH-MBER (256-QAM) is worse than that of Y-Precoders by about 2.8 dB at a BER of 10^{-3}. This is because, to achieve a diversity order same as that achieved by X-,Y-Precoder and E-dmin, linear precoders use only half of the available subchannels for transmission (i.e., $n_s = \min(n_r, n_t)/2 = n_r/2$). Hence, to achieve the same target rate R, they have to use higher order QAM, which results in loss of power efficiency. Based along the lines of the discussion in Section II and observations made in Fig. 9, it can be said that for a 2×2 Rayleigh faded MIMO channel, with a target rate of $R = 8$ bpcu and an overall diversity
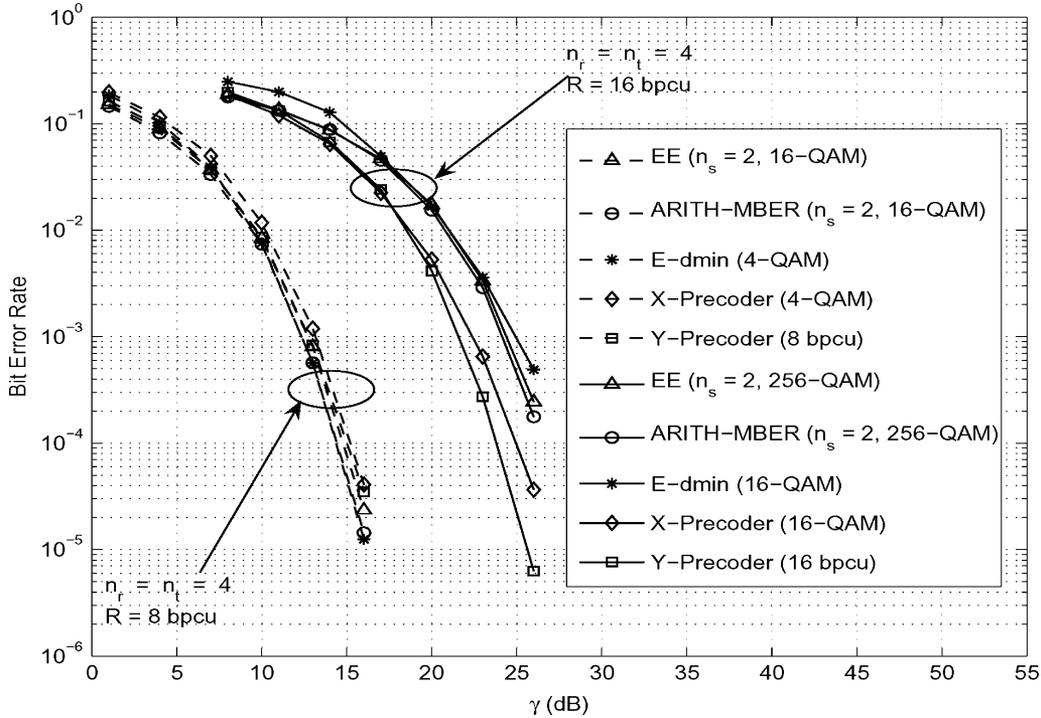
Fig. 10. BER comparison between various precoders for $n_r = n_t = 4$ and target rate $R = 8, 16$ bpcu.

order of $\delta_{ord} = 4$, Y-Precoders are *better* compared to the other considered precoders.

In Fig. 10, we plot the BER for $n_r = n_t = 4$, and a target rate of $R = 8, 16$ bpcu. For a rate of $R = 8$ bpcu, E-dmin and ARITH-MBER (16-QAM) have the best error performance. Y-Precoders perform only about 0.5 dB away at a BER of 10^{-3}. However, for a target rate of $R = 16$ bpcu, Y-Precoders perform the best. ARITH-MBER ($n_s = 2$) with 256-QAM modulation on both subchannels performs 2.6 dB worse than Y-Precoders at a BER of 10^{-3}. E-dmin performs the worst and is about 3.5 dB away from Y-Precoders at a BER of 10^{-3}. E-dmin has poor performance since the precoder proposed in [12] has been optimized for 4-QAM, and therefore it does not perform that well when the target rate is higher than $2n_s$ bpcu.

E. X-Codes Versus Y-Codes

In Fig. 11, we compare the BER performance of the proposed X and Y-Codes for a $n_r = n_t = 2$ system with a target rate of $R = 4, 8$ bpcu. It is observed that Y-Codes have a significant performance gain over X-Codes. For a target rate of both $R = 4$ and $R = 8$ bpcu, Y-Codes perform better than X-Codes by about 1.5 dB at a BER of 10^{-3}. This is primarily due to the novel skewed lattice constellation structure of the proposed Y-Codes (as compared to the simple rotation encoder for X-Codes), which ensures that the minimal distance between received codewords does not become too small when channel is ill conditioned. In Fig. 12, we compare the BER performance of the proposed X and Y-Codes for a $n_r = n_t = 4$ system with a rate of $R = 8, 16$ bpcu. Y-Codes again perform better than X-Codes by about 0.7 dB for a rate of $R = 8$ bpcu, and by about 1.5 dB for a rate of $R = 16$ bpcu.

F. X-, Y-Codes vs. X-, Y-Precoders

In this section, we discuss the performance gain achieved by adaptively choosing the encoder matrices $\mathbf{A}_k$ for each channel realization, as compared to having them fixed *a priori*.

In Fig. 11, we compare the performance of the X-, Y-Precoders with that of X-, Y-Codes for $n_r = n_t = 2$ with $R = 4, 8$ bpcu. For $R = 4$ bpcu, X-, Y-Precoders perform marginally better than X-, Y-Codes (by only about 0.2 dB at BER of 10^{-3}). However, for $R = 8$ bpcu, the X-Precoder performs better than X-Codes by about 1.0 dB, whereas Y-Precoders perform better than Y-Codes by about 0.2 dB at a BER of 10^{-3}. Therefore, adapting the encoder matrices with channel realization is beneficial for X-Codes. However, compared to the performance gain of X-Precoders over X-Codes, it is observed that Y-Precoders do not have as much gain in performance over Y-Codes.

For $n_r = n_t = 4$, it is observed from Fig. 12 that for $R = 8$ bpcu, X-, Y-Precoders have almost similar performance as X-, Y-Codes. However, for $R = 16$ bpcu, X-Precoders perform better than X-Codes by about 0.7 dB, whereas Y-Precoders perform better than Y-Codes by about 0.3 dB at a BER of 10^{-3}.

The performance gain of X-Precoders over X-Codes is much more significant as compared to the performance gain of Y-Precoders over Y-Codes. Also, for X-Precoders, this performance gain is significant only with higher order QAM. This is due to the fact that the error performance is much more sensitive to the rotation angle for higher order QAM (see Fig. 5), and therefore adjusting the rotation angle with respect to the varying channel is expected to result in performance improvement.

On the other hand Y-Precoders are only marginally better than Y-Codes irrespective of the transmission rate R. This is attributed to the fact that for the Y-Precoders we optimize an upper bound to the probability of error $P_k'(\mathbf{H})$, rather than the
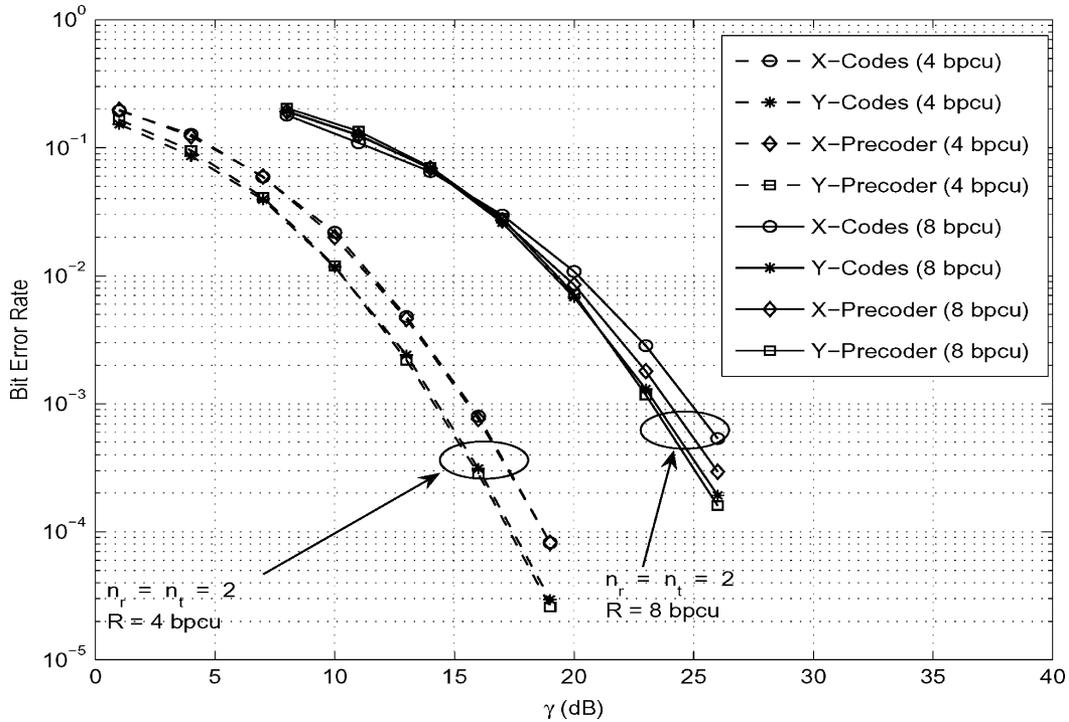
Fig. 11. BER comparison between the proposed X-Codes and Y-Codes for $n_r = n_t = 2$ with rate $R = 4, 8$ bpcu.
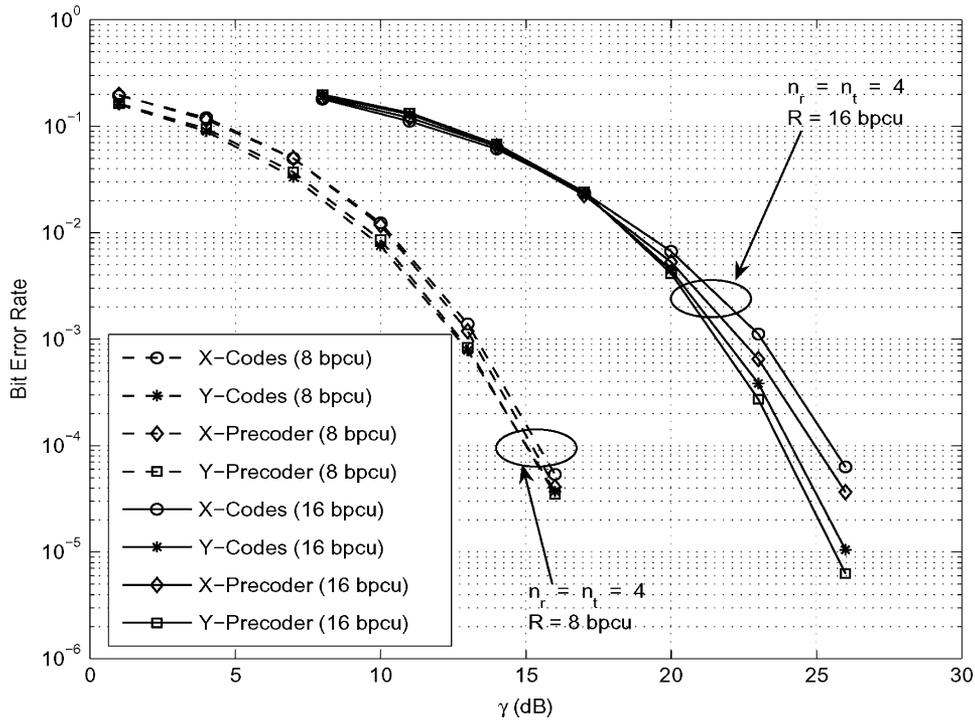

Fig. 12. BER comparison between the proposed X-Codes and Y-Codes for $n_r = n_t = 4$ with rate $R = 8, 16$ bpcu.

exact error probability. We do this, because of the high encoding complexity which in turn is due to the analytical intractability of optimizing the exact error probability expressions. This leads to a suboptimal choice of the encoder matrices, and therefore a suboptimal error performance.

This fact is also obvious from Fig. 13, where we plot the exact optimal average word error probability for Y-Precoders in comparison with the average word error probability of the proposed

suboptimal Y-Precoder. The exact optimal word error probability (i.e., error probability with the optimal choice of encoder matrices as given by (34)) is computed through Monte Carlo techniques. The exact optimal average word error probability is better than the average word error probability of the proposed suboptimal Y-Precoder by about 1.8 dB for a $n_r = n_t = n_s = 2$ system, and is better by about 1.0 dB for a $n_r = n_t = n_s = 4$ system at a word error probability of 10^{-1} and a target rate of
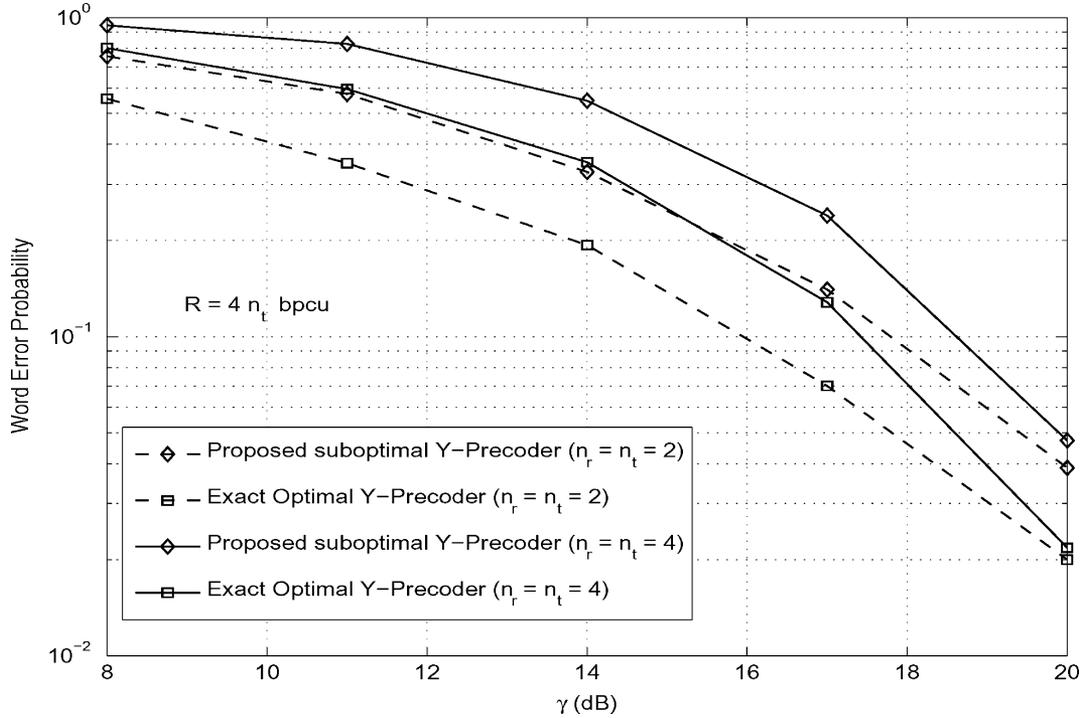
Fig. 13. Word error probability comparison between the proposed suboptimal Y-Precoders and exact optimal Y-Precoders for $n_r = n_t = 2, 4$ with rate $R = 4n_t$ bpcu.

$R = 4n_t$ bpcu. This therefore suggests the existence of better Y-Precoders compared to what has been proposed in this paper.

VII. COMPLEXITY

In this section, we discuss the computational complexity of X-, Y-Codes and compare it with other precoding schemes. In terms of the encoding complexity, all the schemes have a similar order of complexity, $O(n_r n_t)$, which is due to the transmit preprocessing filter. In terms of detection complexity, Y-Codes, CI, THP and ARITH-MBER/EE have a similar detection complexity, which is equivalent to symbol by symbol ML detection. ML detection for X-Codes employs search over a finite subset of some 2-D real search space. E-dmin has the worst ML detection complexity, due to a 4-D real search space. Y-Codes have a detection complexity even lower than X-Codes, since the search is essentially one dimensional (just checking at most three linear relations between the two components of the received signal vector, see Section V-C).

VIII. CONCLUSION

We proposed X-, Y-Codes/Precoders which can achieve high rate and high diversity at low complexity by pairing the MIMO subchannels prior to SVD precoding. It is observed that indeed pairing of channels can significantly improve the overall diversity. Among all possible pairings, pairing the k-th sub-channel with the $(n_r - k + 1)$-th subchannel was found to be optimal in terms of achieving the best overall diversity order. One way of pairing the subchannels is by using rotation based encoding, as was proposed for X-Codes/Precoders. The proposed X-Codes/Precoders have good performance for well conditioned channels. For ill-conditioned channels, we then pro-

posed Y-Codes/Precoders. It is shown by simulation and analysis that Y-Codes/Precoders achieve *better* error performance at very low complexity, when compared to other precoders in the literature. In practice, in order to improve the overall error performance, it is possible to adaptively switch between X- and Y-Codes/Precoders depending on the channel condition.

APPENDIX A
PROOF OF THEOREM 1

Towards proving Theorem 1, we shall find the following Lemma useful (see [15, Prop. 1] for the proof).

Lemma 1: Consider a real scalar channel modeled by $y = \sqrt{\alpha}x + n$, where $x = \pm\sqrt{E_s}$, $n \sim \mathcal{N}(0, \sigma^2)$. Let $F(\alpha) = K\alpha^k + o(\alpha^k)$, for $\alpha \to 0^+$ be the cdf (cumulative density function) of α, where K is a constant and k is a positive integer. Let $\gamma = E_s/\sigma^2$ be the SNR. Then the probability of error is given by $P(\gamma) = \mathbb{E}_\alpha[Q(\sqrt{\alpha\gamma})]$, and the asymptotic error probability for $\gamma \to \infty$ is given by

$$P = \frac{K}{2}((2k-1) \cdot (2k-3)\ldots 5 \cdot 3 \cdot 1)\gamma^{-k} + o(\gamma^{-k}).$$

∎

The proof of Theorem 1 is as follows. Since $Q(\cdot)$ is a monotonically decreasing function with increasing argument, we can further upper bound (36) using (39) and (42) as follows:

$$P_k' \leq (|\mathcal{S}_k| - 1)\mathbb{E}\left[Q\left(\sqrt{\frac{\lambda_{i_k}^2 g_k(\mathbf{A}_k)}{2N_0}}\right)\right]. \quad (85)$$

For a Rayleigh faded channel, the marginal pdf of the s-th eigen-value λ_s^2 (for $\lambda_s^2 \to 0$) is given by [14]

$$f(\lambda_s^2) = C(s)(\lambda_s^2)^{N_t(s)N_r(s)-1} + o\left((\lambda_s^2)^{N_t(s)N_r(s)-1}\right) \tag{86}$$

where $N_t(s) \triangleq (n_t - s + 1)$, $N_r(s) \triangleq (n_r - s + 1)$ and $C(s)$ is a constant given in [14]. Using the pdf in (86), the cdf $F_s(u) = P(\lambda_s^2 \le u)$ (for $u \to 0^+$) is given by

$$F_s(u) = \frac{C(s)}{N_t(s)N_r(s)} u^{N_t(s)N_r(s)} + o\left(u^{N_t(s)N_r(s)}\right). \tag{87}$$

Using Lemma 1 and (87), the bound in (85) can be further written as

$$P_k' \le (|\mathcal{S}_k| - 1)c_k \left(\frac{2P_T}{g_k(\mathbf{A}_k)}\right)^{\delta_k} \gamma^{-\delta_k} + o(\gamma^{-\delta_k}) \tag{88}$$

where $\delta_k \triangleq (n_t - i_k + 1)(n_r - i_k + 1)$ and $c_k \triangleq \frac{C(i_k)}{2\delta_k}((2\delta_k - 1) \cdot \ldots \cdot 5 \cdot 3 \cdot 1)$. $\blacksquare$

APPENDIX B
PROOF OF THEOREM 2

The following lemma is useful towards proving Theorem 2.

Lemma 2: Given a set of $n \ge 2$ probabilities $\{p_1, p_2, \ldots, p_n\}$ $(0 < p_i \le 1, i = 1, 2, \ldots n)$, it is true that

$$\left(1 - \prod_{k=1}^{n}(1 - p_k)\right) < 2^{n-2} \sum_{k=1}^{n} p_k. \tag{89}$$

Proof: Expanding the product in $(1 - \prod_{k=1}^{n}(1 - p_k))$, we have

$$1 - \prod_{k=1}^{n}(1 - p_k) = \sum_{k=1}^{n} p_k + \sum_{i=2}^{n}(-1)^{i+1} S_i^n(p_1, p_2, \ldots, p_n) \tag{90}$$

where $S_i^n(p_1, p_2, \ldots, p_n)$ is a symmetric multinomial in the variables $(p_1, p_2, \ldots, p_n)$, and is defined as

$$S_i^n(p_1, p_2, \ldots, p_n) = \sum_{1 \le j_1 < j_2 < \cdots < j_i \le n} \prod_{l=1}^{i} p_{j_l}. \tag{91}$$

In (91), the number of summands in the expansion of $S_i^n(p_1, p_2, \ldots, p_n)$ is $\binom{n}{i}$. Also, each such summand is a product of probabilities, and therefore the value of each such summand is bounded from above by 1. Therefore, an upper bound to $S_i^n(p_1, p_2, \ldots, p_n)$ is given by

$$S_i^n(p_1, p_2, \ldots, p_n) \le \binom{n}{i}. \tag{92}$$

Taking only the positive terms in

$$\sum_{i=2}^{n}(-1)^{i+1} S_i^n(p_1, p_2, \ldots, p_n)$$

into consideration, we have

$$1 - \prod_{k=1}^{n}(1 - p_k) < \sum_{k=1}^{n} p_k + \sum_{i=3, i \text{ odd}}^{n} S_i^n(p_1, p_2, \ldots, p_n). \tag{93}$$

Further, using the definition of S_i^n in (91), each symmetric multinomial can be upper bounded as

$$\begin{aligned}
&S_i(p_1, p_2, \ldots, p_n) \\
&= \sum_{t=1}^{n} \sum_{j_1 = t < j_2 < \cdots < j_i \le n} \left(p_t \prod_{l=2}^{i} p_{j_l}\right) \\
&= \sum_{t=1}^{n-i+1} p_t S_{i-1}^{n-t}(p_{t+1}, \ldots, p_n).
\end{aligned} \tag{94}$$

Applying (92), since, for $t > 1$, $\binom{n-t}{i-1} < \binom{n-1}{i-1}$, we have

$$\begin{aligned}
S_i(p_1, p_2, \ldots, p_n) &\le \sum_{t=1}^{n-i+1} p_t \binom{n-t}{i-1} \\
&< \sum_{t=1}^{n-i+1} p_t \binom{n-1}{i-1}.
\end{aligned} \tag{95}$$

Since $i > 1$ and $(n - i + 1) < n$, we obtain

$$S_i(p_1, p_2, \ldots, p_n) < \binom{n-1}{i-1} \sum_{t=1}^{n} p_t. \tag{96}$$

Using (96) in (93), we have

$$1 - \prod_{k=1}^{n}(1 - p_k) < \sum_{k=1}^{n} p_k \left(1 + \sum_{i=3, i \text{ odd}}^{n} \binom{n-1}{i-1}\right). \tag{97}$$

Further

$$\begin{aligned}
1 + \sum_{i=3, i \text{ odd}}^{n} \binom{n-1}{i-1} &= \sum_{i=0, i \text{ even}}^{n-1} \binom{n-1}{i} \\
&= 2^{n-2}.
\end{aligned} \tag{98}$$

Using this equality in (97) we have

$$1 - \prod_{k=1}^{n}(1 - p_k) < 2^{n-2} \sum_{k=1}^{n} p_k. \tag{99}$$

This proves the lemma. $\blacksquare$

The proof of Theorem 2 now follows. For $n_r = 2$, there is only one pair, and therefore the overall diversity order is the

same as the diversity order of this pair. For $n_r > 2$, since $P(\mathbf{H}) = 1 - \prod_{k=1}^{n_r/2}(1 - P_k(\mathbf{H}))$, applying Lemma 2, we have

$$P(\mathbf{H}) < 2^{\frac{n_r}{2}-2} \sum_{k=1}^{\frac{n_r}{2}} P_k(\mathbf{H}). \tag{100}$$

Averaging over the channel fading statistics, we have

$$P < 2^{\frac{n_r}{2}-2} \sum_{k=1}^{\frac{n_r}{2}} P_k. \tag{101}$$

Based on (43), it can be argued that, for each k there exists some $\gamma_k^* > 0$, such that for $\gamma \geq \gamma_k^*$, it is true that $P_k < f_k \gamma^{-\delta_k}$, where f_k is some positive constant independent of γ. Further, let $\gamma^* \triangleq \max_k \gamma_k^*$. Let $\delta_{\min}$ and $k_{\min}$ be defined as

$$\delta_{\min} \triangleq \min_k \delta_k, \quad k_{\min} \triangleq \arg\min_k \delta_k. \tag{102}$$

Note that $k_{\min}$ is the index of the pair which has the lowest achievable diversity order among all the $n_r/2$ pairs. Since there are finitely many pairs, there does exist a critical SNR value, $\gamma_c \triangleq \max_k \left(\frac{f_k}{f_{k_{\min}}}\right)^{\frac{1}{(\delta_k - \delta_{\min})}}$, beyond which $k_{\min} = \arg\max_k f_k \gamma^{-\delta_k}$. Using this fact, and (101) it therefore follows that for $\gamma > \max(\gamma_c, \gamma^*)$

$$P < \frac{n_r}{2} f_{k_{\min}} 2^{\frac{n_r}{2}-2} \gamma^{-\delta_{\min}}. \tag{103}$$

From (103), it is clear that

$$\delta_{ord} \geq \min_k \delta_k.$$

APPENDIX C
PROOF OF THEOREM 3

Let $d(\theta_k, \lambda_{i_k}, \lambda_{j_k}) \triangleq \min_{(p,q) \in \mathbb{S}_2} d_k^2(p, q, \theta_k)$, where $d_k^2(p, q, \theta_k)$ is defined in (60). The objective is to find the optimal θ_k which maximizes $d(\theta_k, \lambda_{i_k}, \lambda_{j_k})$. The set $\mathbb{S}_2$ for 4-QAM contains exactly 8 elements. Also due to sign symmetry of this set (i.e., if $(p,q) \in \mathbb{S}_2$ then so do $(p,-q), (-p,q)$ and $(-p,-q)$), there are actually only 4 distances to be computed. For a given angle θ_k these distances are enumerated as follows:

$$d_1 = \lambda_{i_k}^2 \cos^2(\theta_k) + \lambda_{j_k}^2 \sin^2(\theta_k)$$
$$d_2 = \lambda_{i_k}^2 \sin^2(\theta_k) + \lambda_{j_k}^2 \cos^2(\theta_k)$$
$$d_3 = \lambda_{i_k}^2 (\cos(\theta_k) + \sin(\theta_k))^2$$
$$\quad + \lambda_{j_k}^2 (\cos(\theta_k) - \sin(\theta_k))^2$$
$$d_4 = \lambda_{i_k}^2 (\cos(\theta_k) - \sin(\theta_k))^2$$
$$\quad + \lambda_{j_k}^2 (\cos(\theta_k) + \sin(\theta_k))^2.$$

Therefore, $d(\theta_k, \lambda_{i_k}, \lambda_{j_k})$ can be expressed in terms of these distances as

$$d(\theta_k, \lambda_{i_k}, \lambda_{j_k}) = \min(d_1, d_2, d_3, d_4)$$
$$= \min(\min(d_1, d_2), \min(d_3, d_4)).$$

It can be shown that for the maximization in (62), it suffices to only consider the range of θ_k to be $(0, \pi/4]$. Due to the ordering of singular values in the SVD decomposition, $\lambda_{i_k} \geq \lambda_{j_k}$. Hence, it can be concluded that

$$d_2 = \min(d_1, d_2)$$
$$d_4 = \min(d_3, d_4)$$

and therefore

$$d(\theta_k, \lambda_{i_k}, \lambda_{j_k}) = \min(d_2, d_4).$$

Let

$$\beta_k \triangleq \frac{\lambda_{i_k}}{\lambda_{j_k}}.$$

Then, d_4 is the minimum if the following condition is satisfied.

$$\frac{1}{\beta_k^2} \leq h(\theta_k) \tag{104}$$

where

$$h(\theta_k) \triangleq 1 - \frac{1}{\sin(2\theta_k) + \sin^2(\theta_k)}.$$

It can be seen that over the interval $(0, \pi/4]$, $h(\theta_k)$ is a continuous and monotonically increasing function. The maximum value of $h(\theta_k)$ over this interval is $\frac{1}{3}$. Therefore, we now consider two situations depending upon whether β_k is greater than or less than $\sqrt{3}$.

If $\beta_k \leq \sqrt{3}$, then $\frac{1}{\beta_k^2} \geq \frac{1}{3}$. Since $h(\theta_k)$ is always less than $\frac{1}{3}$, we can conclude that the condition in (104) is never satisfied and therefore d_2 is the minimum. Further, since $\lambda_{i_k} \geq \lambda_{j_k}$, d_2 is a monotonically increasing function of θ_k and therefore the solution to (62) is $\pi/4$.

If $\beta_k \geq \sqrt{3}$, then $\frac{1}{\beta_k^2} \leq \frac{1}{3}$. Since $h(\theta_k)$ is a monotonically increasing function we observe that d_4 is the minimum when $\theta_k \geq \theta_k^*$ or else d_2 is the minimum. Here θ_k^* is such that $\theta_k^* \in (0, \pi/4]$ and $h(\theta_k^*) = \frac{1}{\beta_k^2}$. Further, it is observed that d_2 is a monotonically increasing function of θ_k whereas d_4 is monotonically decreasing. Also, $d_2 = d_4$ when $\theta_k = \theta_k^*$. Therefore, it can be concluded that $\min(d_2, d_4)$ is maximized when $\theta_k = \theta_k^*$. Hence for $\beta_k \geq \sqrt{3}$ the solution to (62) is θ_k^*. We now solve for θ_k^* from $h(\theta_k^*) = 1/\beta_k^2$, which yields a quadratic equation in $\tan(\theta_k^*)$, with one solution in $(0, \pi/4]$ given by

$$\theta_k^* = \tan^{-1}\left[(\beta_k^2 - 1) - \sqrt{(\beta_k^2 - 1)^2 - \beta_k^2}\right].$$

Combining, the optimal angles obtained for $\beta_k \leq \sqrt{3}$ and $\beta_k > \sqrt{3}$, we get the solution to the maximization problem in Theorem 3. ∎

APPENDIX D
PROOF OF THEOREM 4

We first get an expression for $d_{k,\min}^2(a_k, b_k)$ as defined in (81). For any code vector at index v which is even, the nearest distance to any other code vector with even index is $4\lambda_{i_k}^2 a_k^2$.

The nearest distance to any code vector at odd index is $\lambda_{i_k}^2 a_k^2 + 4\lambda_{j_k}^2 b_k^2$. The same holds true if v is odd. Hence $d_{k,\min}^2(a_k, b_k)$ is given by

$$d_{k,\min}^2(a_k, b_k) = \min\left(4\lambda_{i_k}^2 a_k^2, \lambda_{i_k}^2 a_k^2 + 4\lambda_{j_k}^2 b_k^2\right). \quad (105)$$

Therefore, our objective is to solve the following constrained min-max optimization problem:

$$(a_k^*, b_k^*) = \arg \max_{\{(a_k,b_k)\in\mathbb{R}^{2+} \mid b_k^2 + a_k^2 \frac{M^2-1}{12} = \frac{P_T}{n_r}\}} d_{k,\min}^2(a_k, b_k). \quad (106)$$

In (106), we let $T_1 \triangleq 4\lambda_{i_k}^2 a_k^2$, which is geometrically a straight line w.r.t. a_k^2 passing through the origin and attaining a maximum value of

$$\frac{48P_T\lambda_{i_k}^2}{n_r(M^2-1)}$$

at

$$a_k^2 = \frac{12P_T}{n_r(M^2-1)}.$$

This is because, the transmit power constraint bounds the value of a_k^2 as

$$a_k^2 \leq \frac{12P_T}{n_r(M^2-1)}.$$

In (106), we let $T_2 \triangleq (\lambda_{i_k}^2 a_k^2 + 4\lambda_{j_k}^2 b_k^2)$. Since

$$b_k^2 + a_k^2 \frac{(M^2-1)}{12} = \frac{p_T}{n_r}$$

we can express T_2 as

$$T_2 = \lambda_{j_k}^2\left(4\frac{P_T}{n_r} + a_k^2\left(\beta_k^2 - \frac{M^2-1}{3}\right)\right). \quad (107)$$

From (107), we observe that, if $\beta_k^2 \geq (M^2-1)/3$, then T_2 represents a straight line with positive slope, having a value of $4\lambda_{j_k}^2 P_T/n_r$ at $a_k = 0$ and attaining a maximum value of

$$\frac{12P_T\lambda_{i_k}^2}{n_r(M^2-1)}$$

at

$$a_k^2 = \frac{12P_T}{n_r(M^2-1)}.$$

This maximum value is less than the maximum attained by T_1. Since both T_1 and T_2 have positive slope the minimum among T_1 and T_2 is maximized at $a_k^{*2} = 12P_T/(n_r(M^2-1))$, which implies that $b_k^* = 0$. The value of $d_{k,\min}^2(a_k, b_k)$ at $a_k = a_k^*$ is the maximum value attained by T_2. Therefore, when the channel condition exceeds a certain threshold, it is optimal to allocate all power to the stronger subchannel only.

On the other hand, if $\beta_k^2 < \frac{M^2-1}{3}$, then T_2 represents a straight line with negative slope, whereas T_1 has positive slope, and therefore the minimum between them is maximized when they are both equal. Therefore, the optimal (a_k^*, b_k^*) must satisfy

$$4\lambda_{i_k}^2 a_k^{*2} = \lambda_{i_k}^2 a_k^{*2} + 4\lambda_{j_k}^2 b_k^{*2}. \quad (108)$$

Using the fact that $b_k^{*2} + a_k^{*2}\frac{M^2-1}{12} = \frac{P_T}{n_r}$, the optimal (a_k^*, b_k^*) is given by

$$(a_k^*, b_k^*) = \left(\frac{\sqrt{\frac{4P_T}{3n_r}}}{\sqrt{\beta_k^2 + \frac{M^2-1}{9}}}, \sqrt{\frac{P_T}{n_r}}\frac{\beta_k}{\sqrt{\beta_k^2 + \frac{M^2-1}{9}}}\right). \quad (109)$$

Using (109), the optimal value of $d_{k,\min}^2(a_k, b_k)$ for $\beta_k^2 < \frac{M^2-1}{3}$ is given by

$$d_{k,\min}^2(a_k^*, b_k^*) = \frac{16P_T\lambda_{i_k}^2}{n_r\left(3\beta_k^2 + \frac{(M^2-1)}{3}\right)}.$$

■

REFERENCES

[1] G. Raleigh and J. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 44, no. 3, pp. 357–366, Mar. 1998.
[2] R. Knopp and G. Caire, "Power control schemes for TDD systems with multiple transmit and receive antennas," in *Proc. IEEE Global Telecommunications Conf. (Globecom)*, Rio de Janeiro, Brazil, Nov. 1999, pp. 2326–2330.
[3] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.
[4] P. W. Baier, M. Meurer, T. Weber, and H. Troeger, "Joint transmission (JT), an alternative rationale for the downlink of time division CDMA using multi-element transmit antennas," in *Proc. IEEE Int. Symp. pread Spectrum Techniques and Applications (ISSSTA)*, Parsippany, NJ, Sep. 2000, pp. 1–5.
[5] H. Harashima and H. Miyakawa, "Matched transmission technique for channels with inter-symbol interference," *IEEE Trans. Commun.*, vol. COM-20, no. 4, pp. 774–780, Aug. 1972.
[6] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *Electron. Lett.*, vol. 7, pp. 138–139, 1971.
[7] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. B. Huber, "Space-time transmission using Tomlinson-Harashima precoding," in *Proc. Int. Seminar on Broadband Communications (IZS'02)*, Zurich, Switzerland, Feb. 2002.
[8] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE. Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.
[9] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1051–1064, May 2002.
[10] C. Windpassinger and R. F. H. Fischer, "Low-complexity near-maximum-likelihood detection and precoding for MIMO systems using lattice reduction," in *Proc. Information Theory Workshop (ITW'03)*, Paris, France, Mar. 31–Apr. 4 2003.
[11] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication—Part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.
[12] B. Vrigneau, J. Letessier, P. Rostaing, L. Collin, and G. Burel, "Extension of the MIMO precoder based on the minimum euclidean distance: A cross-form matrix," *IEEE J. Select. Topics Signal Process.*, vol. 2, no. 2, pp. 135–146, Apr. 2008.
[13] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Found. and Trends in Commun. and Inf. Theory*, vol. 1, no. 1, 2004.
[14] L. G. Ordonez, D. P. Palomar, A. P. Zamora, and J. R. Fonollosa, "High-SNR analytical performance of spatial multiplexing MIMO systems with CSI," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5447–5463, Nov. 2007.
[15] Z. Wang and G. B. Giannakis, "A simple and general parametrization quantifying performance in fading channels," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1389–1398, Aug. 2003.

[16] J. B. Andersen, "Array gain and capacity for known random channels with multiple element arrays at both ends," *IEEE J. Select. Areas Commun.*, vol. 18, no. 11, pp. 2172–2178, Nov. 2000.

[17] E. Sengul, E. Akay, and E. Ayanoglu, "Diversity analysis of single and multiple beamforming," *IEEE Trans. Commun.*, vol. 54, no. 6, pp. 990–993, Jun. 2006.

[18] H. Sampath, "Linear Precoding and Decoding for Multi-Input Multi-Output (MIMO) Wireless Channels," Ph.D. dissertation, Stanford Univ., Stanford, CA, Apr. 2001.

[19] J. Boutros, E. Viterbo, C. Rastello, and J. C. Belfiore, "Good lattice constellations for both Rayleigh fading and Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 502–518, Mar. 1996.

[20] J. Boutros and E. Viterbo, "Signal space diversity: A power and bandwidth efficient diversity technique for the Rayleigh fading channel," *IEEE. Trans. Inf. Theory*, vol. 44, no. 4, pp. 1453–1467, Jul. 1998.

[21] S. K. Mohammed, E. Viterbo, Y. Hong, and A. Chockalingam, "Precoding with X-codes to increase capacity with discrete input alphabets," in *Proc. IEEE Int. Symp. Information Theory (ISIT'10)*, Austin, TX, Jun. 2010.

Saif Khan Mohammed (S'10) received the B.Tech. degree in computer science and engineering from the Indian Institute of Technology, New Delhi, India, in 1998 and the Ph.D. degree from the Electrical Communication Engineering Department, Indian Institute of Science, Bangalore, India, in 2010.

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

Dr. Mohammed 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 electrical engineering in 1995, 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 in the Dipartimento di Elettronica, Politecnico di Torino. In 1997–1998, he was a Postdoctoral Research Fellow in the Information Sciences Research Center, AT&T Research, Florham Park, NJ. He became First Assistant Professor (1998) then Associate Professor (2005) in Dipartimento di Elettronica, Politecnico di Torino. In 2006, he became a Full Professor in DEIS at the University of Calabria, Calabria, Italy. Since September 2010, he has been a Full Professor in the ECSE Department, Monash University, Melbourne, Australia.

In 1993 he was Visiting Researcher in 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, France. In 2003, he was a Visiting Researcher at the Maths Department, EPFL, Lausanne, Switzerland. In 2004, he was Visiting Researcher at the Telecommunications Department, UNICAMP, Campinas, Brazil. In 2005, 2006, and 2009, he was a Visiting Researcher at the ITR of UniSA, Adelaide, Australia. In 2007, he was a Visiting Fellow at 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.

Dr. Viterbo is an ISI Highly Cited Researcher since 2009. He is the Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY, *European Transactions on Telecommunications*, and *Journal of Communications and Networks*, and Guest Editor for 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 in the Department of Electrical and Computer Systems Engineering, Monash University, Melbourne, Australia. She was previously with the Institute of Telecommunications Research, University of South Australia; at the Institute of Advanced Telecommunications, Swansea University, U.K.; and at the University of Calabria, Calabria, Italy. Her research interests include information and communication theory with applications to telecommunication engineering.

Dr. Hong is a Technical Program Committee Chair of AUSCTW'11, Melbourne, Australia. She was the Publicity Chair at the IEEE Information Theory Workshop 2009, Sicily, Italy. She was a Technical Program Committee member for various IEEE conferences such as IEEE ICC 2011, VTC 2011, PIMRC, and WCNC 2008. She 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; a Wireless Data Communication System Scholarship from UNSW; and a NICTA-ACoRN Earlier Career Researcher award for a paper presented at the Australian Communication Theory Workshop (AUSCTW), Adelaide, Australia, in 2007. She is a member of ACoRN.

Ananthanarayanan Chockalingam (S'92–M'93–SM'98) received the B.E. (Hons.) 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, India, in 1985, and the Ph.D. degree in electrical communication engineering (ECE) from the Indian Institute of Science (IISc), Bangalore, India, in 1993.

During 1986 to 1993, he worked with the Transmission R&D Division of the Indian Telephone Industries Limited, Bangalore. From December 1993 to May 1996, he was a Postdoctoral Fellow and an Assistant Project Scientist at the Department of Electrical and Computer Engineering, University of California, San Diego. From May 1996 to December 1998, he served 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, India, 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 ON SELECTED AREAS IN COMMUNICATION (*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 the Indian National Academy of Engineering.