Linear Precoding for MIMO With LDPC Coding and Reduced Complexity

Thomas Ketseoglou, *Senior Member, IEEE*, and Ender Ayanoglu, *Fellow, IEEE*

***Abstract*—In this paper, the problem of designing a linear precoder for Multiple-Input Multiple-Output (MIMO) systems employing Low-Density Parity-Check (LDPC) codes is addressed under the constraint of minimizing the dependence between the system's receiving branches, thus reducing the relevant transmitter and receiver complexities. Our approach constitutes an interesting generalization of Bit-Interleaved Coded Modulation with Multiple Beamforming (BICMB) which has shown many benefits in MIMO systems. We start with a Pareto optimal surface modeling of the system and show the difficulty involved in the corresponding optimization problem. We then propose an alternative, practical technique, called Per-Group Precoding (PGP), which groups together multiple input symbol streams and corresponding receiving branches in the "virtual" channel domain (after singular value decomposition of the original MIMO channel), and thus results in independent transmitting/receiving streams between groups. We show with numerical results that PGP offers almost optimal performance, albeit with significant reduction both in the precoder optimization and LDPC EXIT chart based decoding complexities.**

***Index Terms*—Per-group processing, maximal diversity precoder, near-capacity achieving channel coding, pareto optimal precoder, receiver independence factor.**

I. Introduction

THE concept of Multiple-Input Multiple-Output (MIMO) still represents a prevailing research direction in wireless communications due to its ever increasing capability to offer higher rate, more efficient communications, as measured by spectral utilization, and under low transmitting or receiving power. Within MIMO research, BICMB [1]–[3] has shown great potential for practical application, due to its excellent diversity gains and its simplicity. For example, BICMB in conjunction with convolutional coding offers maximum diversity and maximum spatial multiplexing simultaneously [1], thus it represents an optimal technique for this type of Forward Error Correction (FEC). In addition, there are many past works available which investigated with success linear precoding through exploitation of a unitary precoding matrix, mainly from a diversity maximization point of view [4], [5]. On the other hand, LDPC coding is the currently prevailing, near-capacity achieving error-correction technique that operates based on input to output mutual information and extrinsic information transfer (EXIT) charts [6], [7]. The problem of designing an optimal linear precoder toward maximizing the mutual information between the input and output was first considered in [8], [9] where the first optimal power allocation strategies are presented (e.g., Mercury Waterfilling (MWF)), together with general equations for the optimal precoder design. In addition, [10] also considered precoders for mutual information maximization and showed that the left eigenvectors of the optimal precoder can be set equal to the right eigenvectors of the channel. Finally, in [11], a mutual information maximizing precoder for a parallel layer MIMO detection system is presented reducing the performance gap between maximum likelihood and parallel layer detection.

Recently, globally optimal linear precoding techniques were presented [12], [13] for finite alphabet inputs, capable of achieving mutual information rates much higher than the previously presented MWF [8] techniques, by introducing input symbol correlation through a unitary input transformation matrix in conjunction with channel weight adjustment (power allocation). These mutual information maximizing globally optimal precoders are more appropriate for LDPC codes which are very popular currently, than e.g., Maximal Diversity Precoders (MDP) [12]. However, the gains presented in [12], [14], [15] are achieved at the expense of significantly increased system complexity, even for small modulation constellation size M (e.g., $M = 2, 4$). In addition, the interesting design of [12] requires a significant computational complexity increase at the receiver, even in its simplified implementation for M-ary Phase Shift Keying (MPSK) systems [14], due to e.g., the dependence present between receiving branches. This increase could be prohibitive if the receiver is the mobile destination, or if the number of receiving branches is high.

In this paper, we propose linear precoding techniques which offer high mutual information between input and output in a MIMO system with Quadrature Amplitude Modulation (QAM) and also offer semi-independence among the receiving branches, thus highly simplifying the Maximum A Posteriori probability (MAP) detector operation, and hence significantly reducing the receiver complexity. First, we investigate the theoretical aspects of jointly maximizing the mutual information rate and the independence of the receiving branches. This is done within a Pareto optimal surface context. We show that this problem becomes very involved due to the non-concavity of

Manuscript received March 18, 2014; revised August 2, 2014 and October 8, 2014; accepted December 6, 2014. Date of publication December 25, 2014; date of current version April 7, 2015. This paper was presented in part at the Asilomar Conference on Signals, Systems & Computers, November 2013. The associate editor coordinating the review of this paper and approving it for publication was A. Ghrayeb.

T. Ketseoglou is with the Electrical and Computer Engineering Department, California State Polytechnic University, Pomona, CA 91768 USA (e-mail: tketseoglou@csupomona.edu).

E. Ayanoglu is with with the Center for Pervasive Communications and Computing, Department of Electrical Engineering and Computer Science, University of California, Irvine, CA 92697-2625 USA (e-mail: ayanoglu@uci.edu).

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

Digital Object Identifier 10.1109/TWC.2014.2382125

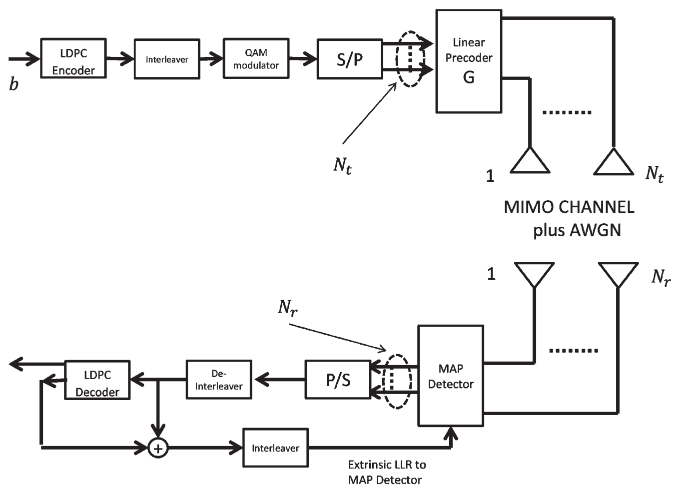


Fig. 1. Original MIMO System model at baseband with Linear Precoder $\mathbf{G}$, N_t transmitting antennas, N_r receiving antennas, and LDPC encoder and decoder. A generic coding rate R is assumed for the LDPC encoder. A similar system can be built for the equivalent channel presented.

the functions involved and thus more practical solutions should be sought. We then proceed to propose a new, interesting technique that groups together "similar" small numbers of multiple streams of input data and receiving branches and then it applies optimized precoding on each group. We carefully look at the group selection strategy, and we show that the best selection is based on selecting input and output groups based on maximum separation of their singular values. The proposed technique is named Per Group Precoding (PGP). PGP offers very good performance with significantly reduced complexity both at the precoder design and receiver levels, due to the independence among different groups and it can be successfully applied even to QAM constellations with $M \geq 16$.

II. Linear Precoder Optimization With Reduced Complexity

The N_t transmit antenna, N_r receive antenna MIMO model (Fig. 1) is described by the following equation

$$\mathbf{y} = \mathbf{HGx} + \mathbf{n}, \tag{1}$$

where $\mathbf{y}$ is the $N_r \times 1$ received vector, $\mathbf{H}$ is the $N_r \times N_t$ MIMO channel matrix comprising independent complex Gaussian components of mean zero and variance one and assumed quasi-static [1], $\mathbf{G}$ is the precoder matrix of size $N_t \times N_t$, $\mathbf{x}$ is the $N_t \times 1$ data vector with independent, identically distributed components of (normalized) power one (thus with covariance matrix $\mathbf{K}_X = \mathbf{I}_{N_t}$), each of which is in the QAM constellation, and $\mathbf{n}$ represents the circularly symmetric complex Additive White Gaussian Noise (AWGN) of size $N_r \times 1$, with mean zero and covariance matrix $\mathbf{K}_n = \sigma_n^2 \mathbf{I}_{N_r}$, where $\mathbf{I}_{N_r}$ is the $N_r \times N_r$ identity matrix, and $\sigma_n^2 = \frac{1}{SNR}$, SNR being the (coded) symbol signal-to-noise ratio. The precoding matrix $\mathbf{G}$ needs to satisfy the following power constraint

$$\mathrm{tr}(\mathbf{GG}^h) = N_t, \tag{2}$$

where $\mathrm{tr}(\mathbf{A})$, $\mathbf{A}^h$ denote the trace, and the Hermitian transpose of matrix $\mathbf{A}$, respectively.

An equivalent model, assuming without loss of generality $N_{tv} = N_{rv} = N_t$ in the virtual domain (after singular value decomposition (SVD),[1] where N_{tv}, N_{rv} represent the transmitting and receiving antennas in the virtual domain, respectively, called herein the "virtual" channel can be easily built based on [12] as follows

$$\mathbf{y} = \Sigma_H \Sigma_G \mathbf{V}_G^h \mathbf{x} + \mathbf{n}, \tag{3}$$

where Σ_H and Σ_G are diagonal matrices containing *non-zero* singular values of $\mathbf{H}$, $\mathbf{G}$, respectively, padded with zeroes if necessary for dimension consistency, and $\mathbf{V}_G$ is the matrix of the right singular vectors of $\mathbf{G}$. Thus, all matrices in (3) are of size $N_t \times N_t$, while the vectors are of size $N_t \times 1$. When LDPC coding with sufficient blocklength (see below) is employed in this MIMO system, the overall utilization in $b/s/Hz$ is determined by the mutual information between the transmitting branches $\mathbf{x}$ and the receiving ones, $\mathbf{y}$ [6], [7]. It is shown [12] that the mutual information between $\mathbf{x}$ and $\mathbf{y}$, $I(\mathbf{x};\mathbf{y})$, is only a function of $\mathbf{W} = \mathbf{V}_G \Sigma_H^2 \Sigma_G^2 \mathbf{V}_G^h$, an important property toward mutual information maximization. The optimal precoder $\mathbf{G}$ is found by solving

$$\begin{aligned} &\underset{\mathbf{G}}{\text{maximize}} \quad I(\mathbf{x};\mathbf{y}) \\ &\text{subject to} \quad \mathrm{tr}(\mathbf{GG}^h) = N_t. \end{aligned} \tag{4}$$

In (4), the constraint present is due to the total (normalized) average MIMO input power which needs to be kept equal to N_t. The average MIMO input power is given as $P_{MIMO} = \mathbb{E}(\mathrm{tr}(\mathbf{Gxx}^h\mathbf{G}^h)) = \mathrm{tr}(\mathbf{GG}^h)$, due to the assumptions made on $\mathbf{x}$. The solution to (4) results in exponential complexity at both transmitter and receiver, as shown in Section II.

When an appropriate LDPC code[2] of sufficient blocklength N_b is employed in the described MIMO system, in conjunction with Gray coding and interleaving employed at the transmitter, followed by MAP detector at the receiver, the system offers very low bit error rate (BER) (e.g., $BER < 10^{-4}$) [6], [7], provided that the coding rate of the LDPC code satisfies the condition

$$R < \frac{I(\mathbf{x},\mathbf{y})}{N_t \log_2(M)}. \tag{5}$$

This is due to the fact that LDPC codes are near-capacity achieving codes. For example, based on the published results in [12], a blocklength $N_b \geq 2400$ would suffice toward meeting (5) closely for a 2 × 2 MIMO system. Thus, designing precoders for high input-output mutual information is more appropriate for LDPC systems than other type of precoders, e.g., MDP. Based on this fact, we focus on this type of precoder designs without special attention on the LDPC code design details.

[1] Due to SVD, the original $N_r \times N_t$ system described by (1) is equivalent to a size $N_t \times N_t$ virtual MIMO system.

[2] Other types of near-capacity achieving channel coding, e.g., Turbo coding could also be employed in our MIMO precoding schemes, as well. However, as LDPC codes represent one form of the currently prevailing channel coding techniques, we decided to focus on this type of channel coding herein.

A. Pareto Surface Precoder Optimization

In this paper, we are interested in precoders which are capable of maximizing jointly $I(\mathbf{x};\mathbf{y})$ and a measure of independence between the elements of $\mathbf{y}$ in (3). Our rationale is that these types of precoders lead to simpler designs with reduced complexity both at the transmitter and the receiver. To see this, we use the fact that higher values of the degree of independence between the elements of $\mathbf{y}$ are facilitated by independent partitions of the MIMO channel input-output space into independent groups. An input-output MIMO Independent Group Partition (IGP) with N_g groups, $\mathcal{P}$, is defined as follows: $\mathcal{P} = \{\mathcal{S}_1, \cdots, \mathcal{S}_{N_g}\}$, where each group in the partition, $\mathcal{S}_i$ with $1 \le i \le N_g$, comprises N_{gi} unique, non-intersecting input-output vector pairs, $(\mathbf{x}_{\mathcal{S}_i}, \mathbf{y}_{\mathcal{S}_i})$, which satisfy $\cup_{i=1}^{N_g} \mathcal{N}(\mathbf{x}_{\mathcal{S}_i}) = \cup_{i=1}^{N_g} \mathcal{N}(\mathbf{y}_{\mathcal{S}_i}) = \{1, 2, \cdots, N_t\}$ and $\mathcal{N}(\mathbf{x}_{\mathcal{S}_i}) \cap_{j \ne i} \mathcal{N}(\mathbf{x}_{\mathcal{S}_j}) = \mathcal{N}(\mathbf{y}_{\mathcal{S}_i}) \cap_{j \ne i} \mathcal{N}(\mathbf{y}_{\mathcal{S}_j}) = \emptyset$, where $\mathcal{N}(\cdot)$ represents the index set (nodes) present in the argument set (within the parenthesis). In addition, IGP requires that each input node, n, in a partition set $\mathcal{S}_i$ is connected only to the corresponding outputs in $\mathcal{S}_i$, in other words, the elements of $\mathbf{V}_G$ corresponding to all other outputs are set to zero, i.e., $\mathbf{V}_G(n, j) = 0$, for output nodes j with $j \notin \mathcal{N}(\mathbf{y}_{\mathcal{S}_i})$, and where $\mathbf{V}_G(n, j)$ represents the ith row, jth column entry of matrix $\mathbf{V}_G$ of (3). All the IGP schemes considered herein use square group structure, i.e., with same number of input and output elements in each group. In other words, $|\mathcal{N}(\mathbf{x}_{\mathcal{S}_i})| = |\mathcal{N}(\mathbf{y}_{\mathcal{S}_i})| = N_{vi}$, with $\sum_{i=1}^{N_g} N_{vi} = N_t$.

Thus, the IGP structure clearly defines N_g independent MIMO input-output groups, i.e., N_g independent, smaller dimension MIMO systems. In the sequel we will also need the notion of an Output only Independent Group Partition (OIGP). Let $p(\cdot)$ be the pdf of the quantity within the parenthesis, and let $\Pr(\cdot)$ be the probability of the quantity within the parenthesis. An OIGP MIMO system is defined as a partition of the output vector $\mathbf{y}$ into N_g independent groups, i.e., with $p(\mathbf{y}) = \prod_{l=1}^{N_g} p(\mathbf{y}_{\mathcal{S}_l})$. An interesting fact is that an AWGN MIMO system as defined by (3) is equivalent to an IGP with the same number of groups, N_g, as we show below. Now we present two lemmas that are essential to the overall paper understanding, with their proofs presented in the Appendices.

Lemma 2.1: Any MIMO OIGP with N_g groups is equivalent to a MIMO IGP with the same number of groups.

Based on this lemma, we see that seeking for higher degree of independence in the output of the MIMO system is equivalent to searching for an IGP.

Now, let $\mathcal{S}(l)(1 \le l \le N_t)$, be the mapping from the lth input $\mathbf{x}_l$ to its corresponding group, and let $|\mathcal{S}(l)|$ be the number of elements in this group.

Lemma 2.2: Any MIMO OIGP with N_g groups results in an equivalent expression of the symbol MAP probability, $\Pr(\mathbf{x}_i|\mathbf{y})$, as follows

$$\Pr(\mathbf{x}_i|\mathbf{y}) = \Pr\left(\mathbf{x}_i|\mathbf{y}_{\mathcal{S}(i)}\right)$$

where we have used $\Pr(\cdot)$ since $\mathbf{x}_i$ takes values in a discrete set.

Due to this lemma, the MAP detector with OIGP needs to only look at the MIMO system defined by $\mathcal{S}(i)$ to calculate the MAP probability, thus simplifying the overall MAP detector complexity as only $N_{vi} < N_t$ input and output nodes are present in the MAP detector of $\mathbf{x}_i$. Then each group can be separately precoded and decoded at the receiver thus simplifying the overall system complexity. This reduction in the receiver complexity is evaluated in more detail in Section III.

A widely accepted measure of independence among the elements of a random vector is the difference between the joint entropy and the sum of all marginal entropies, which herein is called Receiver Independence Factor (RIF)

$$RIF \doteq H(\mathbf{y}) - H_I(\mathbf{y}) = H(\mathbf{y}) - \sum_{i=1}^{N_r} H_i(\mathbf{y}_i), \tag{6}$$

where $H(\cdot)$ represents the entropy of the argument in the parenthesis, $\mathbf{y}_i$ is the ith receiving branch, and we defined $H_I(\mathbf{y}) \doteq \sum_{i=1}^{N_r} H_i(\mathbf{y}_i)$. Thus, to exploit a high RIF jointly with a high mutual information, one should aim at jointly maximizing over $\mathbf{G}$ the vector $[H(\mathbf{y})RIF]^T$ over the (proper) cone $\mathbb{R}_+^2$, satisfying the power constraint $\mathrm{tr}(\mathbf{G}\mathbf{G}^h) = N_t$, resulting in a multicriterion problem [16]. A scalarization of this problem simplifies the problem complexity dramatically and results in a well-accepted method to recover optimal points of the original problem,[3] where $f_0(\mathbf{x})$ is the objective function. In addition, scalarization and Pareto optimal point finding is the natural way to study multicriterion problems such as the one we are dealing with here. For problems over $\mathbb{R}_+^2$, i.e., with $f_0(\mathbf{x}) = [f_{01}(\mathbf{x}) f_{02}(\mathbf{x})]^T$, scalarization leads to maximizing over $\mathbf{x}$ the Pareto surface scalar function $f_{01}(\mathbf{x}) + \lambda_0 f_{02}(\mathbf{x})$ with $\lambda_0 \ge 0$. The Pareto optimization objective function [16] for our RIF MIMO precoder should then be

$$I(\mathbf{x};\mathbf{y}) + \lambda_0 \left(H(\mathbf{y}) - \sum_{i=1}^{N_r} H_i(\mathbf{y}_i) \right), \tag{7}$$

where $\lambda_0 \ge 0$. However, due to the MIMO model in (1), since the noise is complex Gaussian with independent components,

$$I(\mathbf{x};\mathbf{y}) = H(\mathbf{y}) - H(\mathbf{y}|\mathbf{x}) = H(\mathbf{y}) - N_r \log_2\left(\pi e \sigma_n^2\right), \tag{8}$$

where $\log_2(\cdot)$ represents the logarithm with base 2. Thus, by substituting (8) into (9), the following equivalent problem results, since $-N_r \log_2(\pi e \sigma_n^2)$ is constant

$$\begin{aligned} &\underset{\mathbf{G}}{\text{maximize}} && (\lambda_0 + 1) H(\mathbf{y}) - \lambda_0 H_I(\mathbf{y}) \\ &\text{subject to} && \mathrm{tr}(\mathbf{G}\mathbf{G}^h) = N_t, \end{aligned} \tag{9}$$

where $\lambda_0 \ge 0$. Substituting $\lambda \doteq \lambda_0/(\lambda_0 + 1)$, the following equivalent optimization problem needs to be solved to find the Pareto optimal precoder $\mathbf{G}$:

$$\begin{aligned} &\underset{\mathbf{G}}{\text{maximize}} && H(\mathbf{y}) - \lambda \sum_{i=1}^{N_r} H_i(\mathbf{y}_i) \\ &\text{subject to} && \mathrm{tr}(\mathbf{G}\mathbf{G}^h) = N_t \\ &\text{and} && 0 \le \lambda \le 1, \end{aligned} \tag{10}$$

[3] A point x_0 in the multicriterion problem is optimal iff $f_0(\mathbf{x}_0) \preceq_{\mathbb{R}_+^2} f_0(\mathbf{x})$ for feasible $\mathbf{x}$ points leads to $\mathbf{x} = \mathbf{x}_0$.

called the "original problem." Using the "equivalent model" of (3), the following problem needs to be solved instead:

$$\begin{aligned}&\underset{\mathbf{W}=\mathbf{V}_G^h\Sigma_G^2\Sigma_H^2\mathbf{V}_G}{\text{maximize}} && H(\mathbf{y})-\lambda\sum_{i=1}^{N_t}H_i(\mathbf{y}_i)\\ &\text{subject to} && \text{tr}(\Sigma_G^2)=N_t\\ &\text{and} && 0\le\lambda\le 1,\end{aligned} \tag{11}$$

called the "equivalent problem" where the reception model of (3) is employed. Our initial goal is to search for precoder solutions to the problem in (11), for a general λ, to better understand this type of optimal precoder. Then, by increasing λ, we should be capable to achieve a precoder solution with partitioned branches (due to high independence). This will result in a much simpler problem as each sub-precoder (corresponding to different partition) will be easier to determine. However, as our analysis below shows, the problem in (11) is not amenable to a solution. To see this we need a series of lemmas and theorems regarding the parameters of the problem as presented below.

Let $\mathcal{S}$ represent a subset of the receiving node set $\{1,2,\cdots,N_r\}$ and let $\mathbf{y}_\mathcal{S}$ be the restriction of $\mathbf{y}$ to $\mathcal{S}$. We need to introduce here the following two Minimum Mean Square Error (MMSE) matrices which are instrumental in the results presented herein, due to the intimate connection between MMSE matrices and the gradient of $H(\mathbf{y})$ with respect to different parameters, e.g., $\mathbf{W}$, in complex Gaussian MIMO channels with finite alphabet inputs [12], [17].

$$\Phi_{\mathbf{xx}^h}(\mathbf{y}_\mathcal{S}) \doteq \mathbb{E}\left((\mathbf{x}-\mathbb{E}(\mathbf{x}|\mathbf{y}_\mathcal{S}))(\mathbf{x}-\mathbb{E}(\mathbf{x}|\mathbf{y}_\mathcal{S}))^h|\mathbf{y}_\mathcal{S}\right), \tag{12}$$

and

$$\Phi_{\mathbf{xx}^h,\mathcal{S}} \doteq \mathbb{E}(\Phi_{\mathbf{xx}^h}(\mathbf{y}_\mathcal{S})). \tag{13}$$

For the special case when all receiving branches are considered, the notations $\Phi_{\mathbf{xx}^h}(\mathbf{y})$ and $\Phi_{\mathbf{xx}^h}$ are used, for the two MMSE matrices, respectively. The following are results we prove in this paper.

Lemma 2.3: The functions $H_i(\mathbf{y}_i)$, $i=1,2,\cdots,N_t$ are concave functions of $\mathbf{W}$, with gradient $\nabla_\mathbf{W}H_i(\mathbf{y}_i)=\frac{1}{\sigma_n^2}\Phi_{\mathbf{xx}^h,\mathcal{S}=\{i\}}$. Here i in the subscript stands for selecting the ith receiving branch only.

Lemma 2.4: The function $H_I(\mathbf{y})$ is a concave function of $\mathbf{W}$ with gradient, $\nabla_\mathbf{W}H_I(\mathbf{y})=\frac{1}{\sigma_n^2}\sum_{i=1}^{N_r}\Phi_{\mathbf{xx}^h,\mathcal{S}=\{i\}}$.

Theorem 2.5: The optimal precoder of (10) satisfies

$$\mathbf{H}^h\mathbf{HG}\Phi_{\mathbf{xx}^h}-\lambda\sum_{i=1}^{N_r}\mathbf{H}_i^h\mathbf{H}_i\mathbf{G}\Phi_{\mathbf{xx}^h,\mathcal{S}=\{i\}}=\nu\mathbf{G}, \tag{14}$$

for some positive constant ν, and where $\mathbf{A}_i$ represents the ith row of a matrix $\mathbf{A}$.

The equation described in (14) fits within the class of Sylvester matrix equations [18] for which solutions can be found. However, in our case, the MMSE matrices $\Phi_{\mathbf{xx}^h}$ and $\Phi_{\mathbf{xx}^h,\mathcal{S}=\{i\}}(1\le i\le N_r)$ are also functions of the precoder matrix, $\mathbf{G}$. Thus, finding optimal $\mathbf{G}$ based on (14) becomes highly involved. In addition, (14) represents an interesting generalization of the optimal linear precoder equation which first appeared in [8] (for the $\lambda=0$ case).

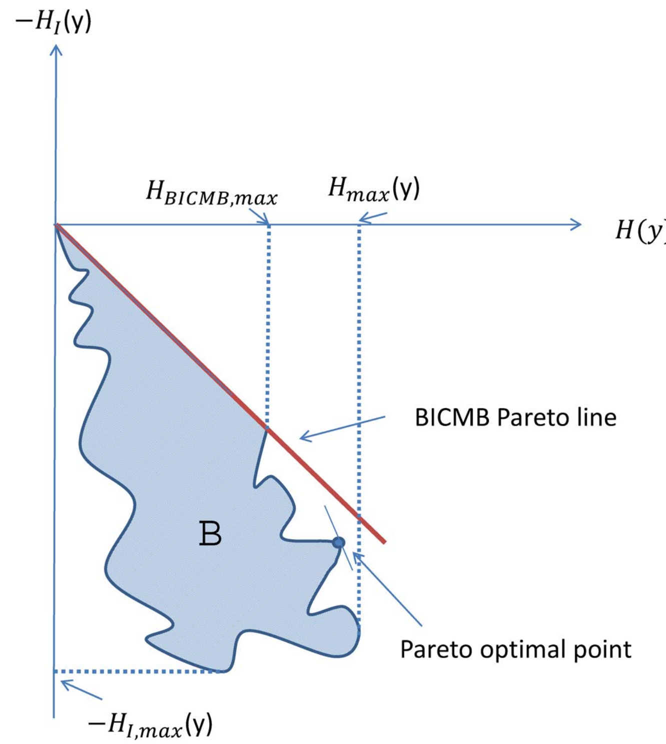


Fig. 2. Pareto optimal maximization. The BICMB solution is shown as the bisecting line segment of the $-\pi/2$ angle, together with its normal vector. A Pareto optimal point is shown as a circle. The maximum achievable mutual information under BICMB, $H_{BICMB,max}(\mathbf{y})$, the global maximum, $H_{max}(\mathbf{y})$, and the maximum under full independence, $H_{I,max}(\mathbf{y})$ are also shown.

Theorem 2.6: The case of ordinary BICMB is the solution of the equivalent problem (11) with $\lambda=1$, i.e., with $\mathbf{V}_G=\mathbf{I}$ being the optimal $\mathbf{V}_G$ and Σ_G an arbitrary non-negative element diagonal matrix satisfying $\text{tr}(\Sigma_G^2)=N_t$.

In Fig. 2 we present a general scenario for the proposed Pareto type optimization. The BICMB Pareto optimal solution is the bisecting line of the $-\pi/2$ angle of the coordinates. Clearly, the objective function value set, $\mathcal{B}$, defines in general a complicated non-convex boundary line on which, there might exist some additional Pareto optimal points, as shown.

Theorem 2.7: The objective in the equivalent problem in (11) is not a concave function of $\mathbf{W}$ when $\lambda=1$.

Thus, based on this theorem, the optimization problems described in (10), and (11) are in general more involved, due to the lack of concavity of the objective function with respect to $\mathbf{W}$.

The proofs of these lemmas and theorems are presented in the Appendices. As entropies are scalar functions of complex matrices in our model, the approach to proving most of these is based on differentiation theory of complex matrices as per [19] and invoking Hessian matrices and the Schur complement [16].

From the above presented theorems, we see that depending on the value of λ, the Pareto surface precoder optimization approach leads to either BICMB when $\lambda=1$, or to a very complex solution as described in Theorem 2.5. Further, as BICMB does not fully utilize the benefits of precoding as required by LDPC codes, its utility within the current context is limited. In general, although increasing λ facilitates solutions which offer partition in the input-output space, one needs to find specific solutions for each λ to be able to capitalize on this fact. Thus, although the

Pareto surface is the natural way to study this type of multicriterion optimization precoder, the required complexity is too high to achieve this. Armed with this result, we next investigate an alternative setup that leads to a different problem with a better solution for our purposes. However, in Section III we present some results which corroborate the concept of Pareto optimal precoder in the context described here. More specifically, we show that the joint search for optimal values of $H(\mathbf{y}), RIF$ leads to precoder designs which indeed outperform other designs in achieving almost optimal mutual information while at the same time they require significantly lower computational complexities at the transmitter and receiver. Thus, the Pareto optimal precoder represents a conceptually correct design choice.

B. A Different Optimization Problem

For any N_g group partition of the "virtual" receiving branches, $\mathbf{y}$ in the equivalent model (3), i.e.,

$$\mathcal{N}(\mathbf{y}_{\mathcal{S}_i})\cap_{j\neq i}\mathcal{N}(\mathbf{y}_{\mathcal{S}_j})=\emptyset \text{ and} \cup_{i=1}^{N_g}\mathcal{N}(\mathbf{y}_{\mathcal{S}_i})=\{1,2,\cdots,N_t\}, \quad (15)$$

where $\mathcal{N}(\cdot)$ represents the index set (nodes) present in the argument set (within the parenthesis), the following inequality is true for fixed $\mathbf{V}_G$, Σ_G:

$$H(\mathbf{y}) \leq \sum_{i=1}^{N_g} H_i(\mathbf{y}_{\mathcal{S}_i}) \leq \sum_{i=1}^{N_r} H_i(\mathbf{y}_i). \quad (16)$$

This is a very fundamental result in information theory [20] and equality holds if and only if the group outputs are independent. Thus, a general partitioning in larger groups brings the MIMO output entropy closer to the partition one (note that this applies to the MIMO input-output mutual information too, due to its irrelevant constant difference from the entropy value). Consider the following optimization problem:

$$\begin{aligned} &\underset{\mathbf{V}_G,\Sigma_G}{\text{maximize}} && H(\mathbf{y}) - \sum_{i=1}^{N_g} H_i(\mathbf{y}_{\mathcal{S}_i}) \\ &\text{subject to} && \mathcal{N}(\mathbf{y}_{\mathcal{S}_i})\cap_{j\neq i}\mathcal{N}(\mathbf{y}_{\mathcal{S}_j})=\emptyset \\ &\text{and} && \cup_{i=1}^{N_g}\mathcal{N}(\mathbf{y}_{\mathcal{S}_i})=\{1,2,\cdots,N_t\}. \end{aligned} \quad (17)$$

The solution of this problem leads to independence between the N_g output groups, i.e., $H(\mathbf{y})=\sum_{i=1}^{N_g} H_i(\mathbf{y}_{\mathcal{S}_i})$ (e.g. [20]), creating an OIGP, as explained above. Based on Lemma 1, an OIGP for this type of AWGN MIMO channel is equivalent to an IGP, thus inputs need to be also partitioned into the same number of groups, wherin each input group is only connected to one output group only. To take advantage of this, consider the following problem:

$$\begin{aligned} &\underset{\mathbf{V}_G,\Sigma_G}{\text{maximize}} && H(\mathbf{y}) \\ &\text{subject to} && H(\mathbf{y}) = \sum_{i=1}^{N_g} H_i(\mathbf{y}_{\mathcal{S}_i}) \\ &\text{and} && \operatorname{tr}\left(\Sigma_G^2\right) = N_t. \end{aligned} \quad (18)$$

This way, the previous result can be easily utilized to exploit this type of inter-group independence. This is the generalized PGP problem (G-PGP). A simpler version is obtained if we further specialize the power constraint in (18) to $\operatorname{tr}(\Sigma_{G_i}^2) = N_{t_i}$, for $i=1,2,\cdots,N_g$. The corresponding solution is called Per Group Precoding (PGP) and it is found as follows: For a particular variable selection method, let $\mathbf{x}_{s_i}$, $\mathbf{y}_{s_i}$ be the data variables and the receiving vector variables in the ith selection subset (group), respectively. Let us denote by N_{t_i}, N_{r_i}, N_g, the numbers of spatially multiplexed data streams, spatially multiplexed receiving antennas per group, and of PGP groups, respectively, then, $N_t=\sum_{i=1}^{N_g} N_{t_i}$ and $N_r=\sum_{i=1}^{N_g} N_{r_i}$. PGP solves the following N_g optimization sub-problems, one for each i (group) ($i=1,2,\cdots,N_g$):

$$\begin{aligned} &\underset{\mathbf{W}_{s_i}}{\text{maximize}} && I(\mathbf{x}_{s_i},\mathbf{y}_{s_i}) \\ &\text{subject to} && \mathbf{W}_{s_i}^h=\mathbf{W}_{s_i} \\ &\text{and} && \operatorname{tr}\left(\Sigma_{G_i}^2\right)=N_{t_i}. \end{aligned} \quad (19)$$

Theorem 2.8: The PGP solution is in the feasible region of the original problem (9).

This is due to the fact that the solution offered by PGP in the equivalent model constitutes a solution to the original problem, i.e., the block diagonal matrix

$$\mathbf{V} = \operatorname{diag}\left[\mathbf{V}_1,\cdots,\mathbf{V}_{N_g}\right] = \begin{bmatrix} \mathbf{V}_{P_1} & 0 & 0 & & & \cdots & 0 \\ 0 & \mathbf{V}_{P_2} & 0 & \cdots & & & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & & \cdots & 0 & \mathbf{V}_{P_{N_g}} \end{bmatrix}$$

is a unitary matrix, and the diagonal matrix

$$\Sigma_G = \operatorname{diag}\left[\Sigma_1,\cdots,\Sigma_{N_g}\right]$$

satisfies $\sum_{i=1}^{N_g}\operatorname{tr}(\Sigma_{G_i}^2)=N_t$. The complete proof is presented in the Appendix.

C. Computational Complexity Evaluation and Comparison

The original linear precoder (global) optimization problem as described by (9) is solved in [12]. Assuming $N_r=N_t$, as it is the case for the presented results herein, we can perform a basic calculation of the complexity involved at the transmitter and receiver sites of a MIMO system employing linear precoding techniques to increase the input-output mutual information. We are evaluating the computational complexity order O for the different transmitter and receiver processing stages employed. O represents the total number of calculations, including the number of real additions, and multiplications required for a task, expressed as a constant times the dominating calculation factor involved in the task. The computational complexity order is then this calculation dominating factor of the task.

1) Evaluation of Transmitter Complexity: Table I shows the complexity order for the different precoder stages at the transmitter. These are the gradient evaluations (GE), as required by the two backtracking line search algorithms, the two required evaluations of the objective function I (OE) per backtracking line search iteration, and the calculation of $d\Theta$ of the differential of the unitary matrix Θ (UE). Table I presents the corresponding

TABLE I
TRANSMITTER COMPUTATIONAL COMPLEXITY ORDER OF GLOBALLY OPTIMAL LINEAR PRECODER

Attribute	*Computational Complexity Order,* $\mathcal{O}$
GE	$2M^{2N_t}$
OE	$2M^{2N_t}$
UE	$8N_t^3$

TABLE II
MAP DETECTOR COMPUTATIONAL COMPLEXITY ORDER OF GLOBALLY OPTIMAL LINEAR PRECODER

Attribute	*Computational Complexity Order,* $\mathcal{O}$
MAP Detector	$M^{2N_t}\log_2(M)$

orders of the computational complexities for these attributes. We see that the transmitter computational complexity order of a global optimal precoder is M^{2N_t}.

2) Evaluation of Receiver Complexity: The major ramification introduced at the receiver due to transmitter precoding is the one regarding the MAP detector [6], [7] which evaluates the channel bit log-likelihood-ratios (LLR). Table II shows the corresponding complexity which is $O(M^{2N_t}\log_2(M))$.

3) Comparison of PGP to Global Optimization Complexity: Concerning the comparison of PGP complexity versus the globally optimal precoder one for the precoder and the MAP detector, when receiver branches are partitioned in independent groups (e.g., high values of RIF), each group represents exponentially smaller complexity both at the precoder determination and at the MAP detector, due to intergroup independence and smaller dimensions per group. To see how PGP reduces the MAP detector complexity, consider PGP in a MIMO system with N_g IGP groups. Let us focus on calculating the probability $\Pr(\mathbf{x}_i|\mathbf{y}) = \Pr(\mathbf{x}_i|\mathbf{y}_{\mathcal{S}(i)})$, for input symbol $\mathbf{x}_i$ in the system. Since PGP relies on an IGP to apply, let us use a corresponding IGP, $\mathcal{P} = \{\mathcal{S}_1, \cdots, \mathcal{S}_{N_g}\}$, we have N_g groups of nodes at both input and output domain. Each group in the partition, $\mathcal{S}_i$ with $1 \le i \le N_g$, comprises N_{gi} unique, non-intersecting input-output vector pairs, $(\mathbf{x}_{\mathcal{S}_i}, \mathbf{y}_{\mathcal{S}_i})$, as explained earlier, with $\sum_{i=1}^{N_g} N_{vi} = N_t$. Let $n_{in}(i,j), n_o(i,j)$ be the mapping from the partition groups to input and output indices, respectively. In other words, $n_{in}(i,j)$ represents the input index (taking values in $\{1, \cdots, N_t\}$) of the jth ($1 \le j \le N_{vi}$) input of the ith group ($1 \le i \le N_g$), and similarly for $n_o(i,j)$. Also, let $\mathcal{S}(l)(1 \le l \le N_t)$, be the mapping from the lth input $\mathbf{x}_l$ to its corresponding group. Finally, let $\mathbf{x}_{\mathcal{S}(i)} = \{\mathbf{x}_{n_{in}(i,1)}, \cdots, \mathbf{x}_i, \cdots, \mathbf{x}_{n_{in}(i,|\mathcal{S}(i)|)}\}$ be the set of input nodes present in the ith group in the IGP. Then, due to Lemma 2.2 and the fact that the MIMO channel in (3) is AWGN, we can write

$$\begin{aligned}\Pr(\mathbf{x}_i|\mathbf{y}) &= \Pr\left(\mathbf{x}_i|\mathbf{y}_{\mathcal{S}(i)}\right)\\ &\propto \frac{1}{M^{N_v \mathcal{S}(i)}} \cdot \sum_{x_{n_{in}(i,l)}, n_{in}(i,l)\neq i} p\left(\mathbf{y}_{\mathcal{S}(i)}|\mathbf{x}_{\mathcal{S}(i)}\right)\\ &= \frac{1}{M^{N_{vi}}} \frac{1}{\pi^{N_v}\sigma_n^{2N_v}} \cdot \\ &\quad \sum_{x_{n_{in}(i,l)}, n_{in}(i,l)\neq i} \exp\left(-\frac{\|\mathbf{y}_{\mathcal{S}(i)} - \tilde{\mathbf{H}}_{\mathcal{S}(i)}\mathbf{x}_{\mathcal{S}(i)}\|^2}{\sigma_n^2}\right), \end{aligned} \tag{20}$$

where $\propto$ means proportional to, i.e., equal except for a multiplication constant independent of $\mathbf{x}_i$, and $\tilde{\mathbf{H}}_{\mathcal{S}(i)}$ is the part of $\tilde{\mathbf{H}} \doteq \Sigma_H \Sigma_G \mathbf{V}_G^h$ that corresponds to $\mathbf{x}_{\mathcal{S}(i)}$. This equation clearly shows that due to the existing IGP, the MAP detector only needs to invoke inputs from the input group corresponding to the particular input under consideration. This also shows that the summations required in each evaluation are of the order $M^{N_{v|\mathcal{S}(i)|}}$, due to PGP, as there are only $N_{v|\mathcal{S}(i)|} - 1$ interfering symbols to $\mathbf{x}_i$, due to inter-group independence as different input groups connect to different antennas due to IGP. Further, as we need $M^{N_{v|\mathcal{S}(i)|}}$ values of this, i.e., over all possible group input combinations, one per different input combination, the total computational complexity order becomes $M^{2N_{v|\mathcal{S}(i)|}}$ per symbol, or $M^{2N_{v|\mathcal{S}(i)|}}\log_2(M)$ per bit in group $\mathcal{S}(i)$.

For N_g groups, and assuming for simplicity equal number of N_t/N_g beams per group, the corresponding complexity associated with PGP is found from Table I, II, after substituting N_t/N_g for N_t, and multiplying by N_g since PGP needs N_g smaller global optimal precoders. This means that PGP reduces the corresponding computational complexities at both the transmitter and receiver by $M^{2N_t(1-1/N_g)}$ times, thus offering a significant computational complexity reduction at both the transmitter and receiver. For example, by using a 4 × 4 MIMO system with QPSK modulation, PGP needs 128 times smaller computational complexity at both the transmitter and receiver. As the number of antennas and the modulation constellation size grow, this PGP complexity reduction becomes more significant.

III. NUMERICAL RESULTS

The results presented herein employ PGP as described above, as well as globally optimal precoders for comparison. Since PGP performs a number of N_g globally optimal precoder determinations, albeit of smaller size, it suffices to describe the globally optimal recoder implementation. The globally optimal precoder implementation methodology is performed by employing two backtracking line searches, one for $\mathbf{W}$, and another one for Σ_G^2, at each iteration, in a fashion similar to [12], but introducing some improvements. Similarly to [12], we follow a block coordinate gradient ascent maximization method to find the solution to the optimization problem described in (9), employing the virtual model of (3). It is proven in [12] that $I(\mathbf{x};\mathbf{y})$ is a concave function over $\mathbf{W}$ and Σ_G^2. It thus becomes efficient to employ two different gradient ascent methods, one for $\mathbf{W}$, and another one for Σ_G^2. We employ Θ, Σ to denote the $\mathbf{V}_G^h$ and Σ_G^2, respectively, evaluated during the optimization algorithm's execution.

The value of the step of each iteration over $\mathbf{W}$ and Σ_G^2 is determined through backtracking line searches, one for each of the variables $\mathbf{W}$ and Σ_G^2. We describe the backtracking line search for $\mathbf{W}$ first, then the one over Σ_G^2. At each iteration of the globally optimal algorithm over $\mathbf{W}$, the gradient of $I(\mathbf{x};\mathbf{y})$ over $\mathbf{W}$, $\nabla_{\mathbf{W}}I$ is required. We then select two parameters, α_1 and β_1, both smaller than one and positive, and perform the backtracking line search as follows: At each new trial, a parameter $t_1 > 0$ that represents the step size is updated by multiplying it with β_1. The initial value for t_1 is equal to 1. Then the algorithm checks if

$$I(\mathbf{W} + t\nabla_{\mathbf{W}}I) > I(\mathbf{W}) + \alpha_1 t_1 \|\nabla_{\mathbf{W}}I\|_F^2, \tag{21}$$

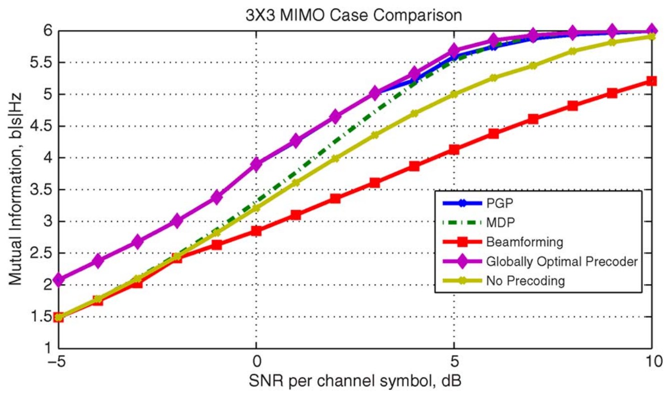


Fig. 3. Results for PGP, globally optimal precoding, no precoding, plain beamforming, and MDP for a 3×3 MIMO system and QPSK modulation.

where $\|\nabla_{\mathbf{W}} I\|_F$ is the Frobenius norm of $\nabla_{\mathbf{W}} I$ [19]. If the condition is satisfied, the algorithm proceeds with calculating a new Θ_{NEW} from $\mathbf{W}_{NEW} = \mathbf{W} + t_1 \|\nabla_{\mathbf{W}} I\|_F^2$, as follows:

$$\mathbf{W}_{NEW} = \Theta_{NEW}^h \Sigma_H^2 \Sigma_G^2 \Theta_{NEW}, \tag{22}$$

by employing the eigenvalue decomposition (EVD) of $\mathbf{W}_{NEW}$, updates $\mathbf{V}_G^h = \Theta_{NEW}$, and then it proceeds to the backtracking line search over Σ_G^2. If the condition is not satisfied, the search updates t_1 to its new and smaller value, $\beta_1 t_1$ and repeats the check on the condition, until the condition is satisfied or a maximum number of attempts in the first loop, n_1 has been reached. Then, the backtracking line search on Σ_G^2 takes place in a fashion similar to the search described for $\mathbf{W}$, but with some ramifications. First, based on the second backtracking line search loop parameters α_2 and β_2, the backtracking line search is as follows: At each new trial, a parameter $t_2 > 0$ that represents the step size is updated by multiplying it with β_2. The initial value for t_2 is equal to 1. Then the algorithm checks if

$$I\left(\mathbf{W} + t\nabla_{\Sigma_G^2} I\right) > I(\mathbf{W}) + \alpha_2 t_2 \left\|\nabla_{\Sigma_G^2} I\right\|_F^2, \tag{23}$$

where $\|\nabla_{\Sigma_G^2} I\|_F$ is the Frobenius norm of $\nabla_{\Sigma_G^2} I$ [19]. If the condition is satisfied, the algorithm proceeds with updating to the new Σ_{NEW}, but after setting any negative terms in the main diagonal of Σ_{NEW} to zero and renormalizing the remaining main diagonal entries. If the condition is not satisfied, the search updates t_2 to its new and smaller value, $\beta_2 t_2$ and repeats the check on the condition, until the condition is satisfied or a maximum number of attempts in the second loop, n_2 has been reached. For most cases presented, it is worth mentioning that only a few iterations (e.g., typically $<$20) are required to converge to the PGP solution results as presented herein, even for higher size MIMO configurations, e.g., 5×5 MIMO systems. All the results consider Quadrature Phase Shift Keying (QPSK) modulation on narrowband independently fading channels. For a 3×3 MIMO system with

$$\mathbf{H} = \begin{bmatrix} 1 & 0.5j & 0.3 \\ -0.5j & 1.5 & -0.1j \\ 0.3 & 0.1j & 0.5 \end{bmatrix},$$

Fig. 3 presents a plot of the mutual information achieved by PGP, globally optimal precoding, no precoding, plain beamforming ($\mathbf{V}_G = \Sigma_G = \mathbf{I}$ in the model presented in (3), which is known to be optimal in low *SNR* values [14]), and the MDP design of [4], as a function of the symbol *SNR* in dB. We observe very close agreement between the performance of the PGP and the globally optimal precoder. We also see that PGP significantly outperforms both the plain beamforming and no precoding case in the "medium" to "high" *SNR* range, by more than 1.5 b/s/Hz and 0.7 b/s/Hz, respectively, as expected (e.g., [12]), however with significantly reduced receiver complexity. In addition, PGP still outperforms MDP by about 0.5 b/s/Hz in the medium *SNR* range. However, this gain of PGP comes simultaneously with a much lower receiver complexity, due to the output branch independence present with PGP. The reason plain beamforming may appear to be performing worse than the no precoding case is as follows. Plain beamforming is known to be the optimal precoding technique for low SNR, but generally it lacks in performance in higher SNR. Since many authors consider it for comparison purposes, e.g., [14], we decided to include it in our results, as well. The rationale behind this low performance is that although the plain beamformer has full independence, it fails to achieve high information transfer. Based on the model of (3), not all precoders can achieve better information transfer than the original channel ($\mathbf{H}$) one.

For a better understanding of the Pareto optimization framework presented in Section II and its impact on system design, Table III presents results on the RIF for the globally optimal precoder and the PGP for different *SNR* values, all for the same 3×3 MIMO system presented above. We observe that as expected, the PGP approach, due to its built-in independence between groups, achieves significantly higher independence than the globally optimal precoder over all *SNR* values. The difference in the corresponding RIF values is about two times over all *SNR* values. Thus, based on this evidence, we see that PGP performs very close to the globally optimal precoder as far as the mutual information is concerned based on Fig. 3, while at the same time it offers much higher independence as shown in Table III which dramatically simplifies both the precoder design and the receiver complexity. This means that PGP achieves a better solution to the Pareto problem in (11) over a wide range of λ. To find the range of λ values $[\lambda_*, 1]$ for which PGP outperforms the globally optimal precoder, we need to find λ_*. Since $\lambda_* = \lambda_0^*/(1+\lambda_0^*)$, we need to find first the corresponding minimum λ_0 in (9). In general, let us denote by $H^{(M1)}(\mathbf{y}), H^{(M_2)}(\mathbf{y})$ and $RIF(M_1), RIF(M_2)$ the output entropies and RIF values achieved by two different precoding methods M_1 and M_2 respectively. Then, $\lambda_0^* = (H^{(M_2)}(\mathbf{y}) - H^{(M_1)}(\mathbf{y}))/(RIF(M_1) - RIF(M_2))$, provided that this value is non-negative. For our case we find that $\lambda_0^* = 0.01964$ at $SNR =$ 4 dB, giving $\lambda_* = 0.01926$ which means that the PGP precoder outperforms the globally optimal one based on the Pareto scalarization objective of (11) for a very wide range of λ, i.e., for λ in [0.01926, 1]. Practically this means that PGP offers a better solution to the MIMO precoding problem as it achieves almost the same performance with the globally optimal precoder albeit at lower complexity. In general, the closer λ_* is to zero, the smaller the gap between the globally optimal achievable mutual information and the PGP one, as it can be easily seen from the presented equations. Thus, based on this evidence, it is seen that the Pareto problem presented in

TABLE III
RIF COMPARISON FOR GLOBALLY OPTIMAL PRECODER AND PGP VERSUS *SNR*

SNR *per symbol*, *dB*	*Globally Optimal Precoder RIF*	*PGP RIF*
-2	-8.44	-4.46
0	-9.14	-4.33
2	-9.50	-4.27
4	-9.70	-4.10
6	-10.0	-4.01
8	-10.73	-4.00

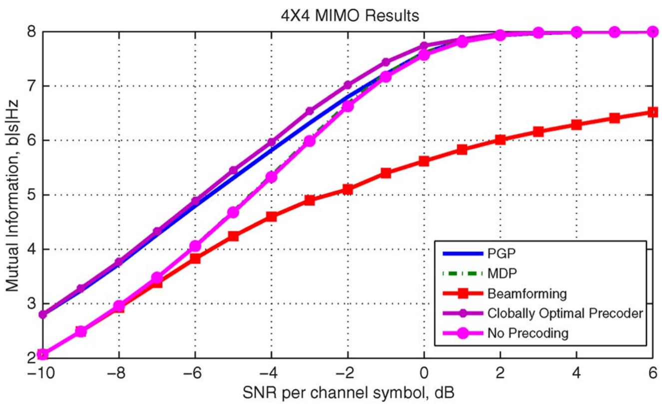


Fig. 4. Results for PGP, globally optimal precoding, no precoding, plain beamforming, and MDP for a 4×4 MIMO system and QPSK modulation.

(11), although intractable, provides a very useful tool in quantifying MIMO precoder tradeoffs between mutual information and complexity.

For the randomly generated 4×4 MIMO system with channel (see equation at the bottom of the page), and with $N_g = 2$, 2×2 subgroups, for a MIMO system with $N_t = N_r = 4$ and QPSK modulation, we get the results shown in Fig. 4. Clearly, PGP still outperforms MDP by about 0.7 b/s/Hz in the medium *SNR* range, while PGP outperforms the plain beamforming case by more than 1.5 b/s/Hz and MDP in the medium to high *SNR* range. At the same time PGP offers performance almost equal to the globally optimal precoder one in the low *SNR* range, while it is still very close to the globally optimal precoder at higher *SNR* values.

TABLE IV
PGP PERFORMANCE IN THREE DIFFERENT GROUP SELECTION SCENARIOS FOR 4×4 MIMO CASE

Group Combination Scheme	I $b/s/Hz$
$\mathcal{G}_1 = \{s_1, s_4\}$, $\mathcal{G}_2 = \{s_2, s_3\}$	6.78
$\mathcal{G}_1 = \{s_1, s_2\}$, $\mathcal{G}_2 = \{s_3, s_4\}$	5.40
$\mathcal{G}_1 = \{s_1, s_3\}$, $\mathcal{G}_2 = \{s_2, s_4\}$	5.02

Consider now the performance of the previous system at $SNR = -2$ dB for all different PGP subgroup formations. Denote by $\{s_1, s_2, s_3, s_4\}$ the set of the singular values of the channel in descending order. Combination 1 employs groups $\mathcal{G}_1 = \{s_1, s_4\}$, $\mathcal{G}_2 = \{s_2, s_3\}$, while Combination 2 employs $\mathcal{G}_1 = \{s_1, s_2\}$, $\mathcal{G}_2 = \{s_3, s_4\}$, and Combination 3 employs $\mathcal{G}_1 = \{s_1, s_3\}$, $\mathcal{G}_2 = \{s_2, s_4\}$. Table IV shows the performance of each selection method. We observe that by forming groups in the PGP by combining the most distant (in value) singular values, best performance is achieved by the PGP.

In Fig. 5 we present the corresponding mutual information results for a randomly generated 5×5 MIMO system with (see equation at the bottom of the page), where we use $N_g = 3$ groups in PGP, two of size 2×2 and one of size 1×1, i.e., a single input, single output one. The selection of the groups is based on the method presented above in which, the most distant eigenvalues are selected in each group. We observe the same type of behavior, i.e., PGP outperforms MDP slightly by a maximum of 0.85 b/s/Hz at a symbol $SNR = -12$ dB. We also observe that this PGP performance improvement over MDP is

$$\mathbf{H} = \begin{bmatrix} 1.0919 + 1.0036i & -0.7507 - 0.5688i & -0.1361 + 0.1525i & -1.0058 + 0.5107i \\ 0.0608 + 0.2062i & 1.6620 + 0.4926i & 0.6283 - 0.8244i & 0.3452 + 1.8282i \\ -1.0547 + 0.1399i & -0.4353 + 0.5905i & -0.5408 - 0.8117i & -0.1254 - 0.4716i \\ -0.5249 + 1.1227i & 0.5290 - 0.1723i & -0.9916 + 0.0742i & -0.1386 + 0.1325i \end{bmatrix}$$

$$\mathbf{H} = \begin{bmatrix} 0.3802 + 0.7316i & -0.9247 + 0.6282i & -0.9545 + 1.0171i & -0.1449 - 0.0723i & 0.4748 - 0.0212i \\ 1.2968 + 0.5140i & -0.3066 - 0.8111i & 2.1460 + 0.2299i & -0.0878 - 0.1707i & -0.8538 - 0.1166i \\ -1.5972 - 0.2146i & 0.2423 - 0.7558i & 0.5129 - 0.5338i & 1.0534 + 0.2257i & 0.5072 + 0.4439i \\ 0.6096 + 0.2078i & 2.5303 - 0.5724i & -0.0446 + 0.9689i & 0.9963 + 0.2212i & 1.1528 + 0.7731i \\ 0.2254 - 0.5567i & 1.9583 - 2.0819i & 0.5054 - 1.2102i & 1.0021 - 0.6116i & 0.3457 + 0.7844i \end{bmatrix}$$

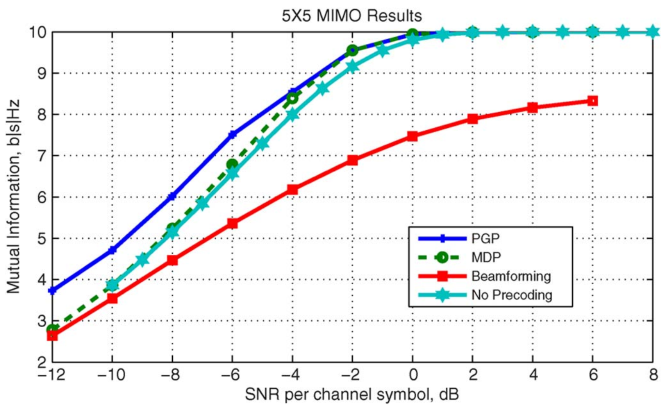


Fig. 5. Results for PGP, no precoding, plain beamforming, and MDP for a 5×5 MIMO system and QPSK modulation.

TABLE V
PGP Performance in Random Fading for PGP, No Precoding, and Globally Optimal Precoder for 3×3 MIMO Case

SNR, dB	*Gl. Opt. I*	*PGP I*	*No Precoding I*
−4	3.46	3.46	3.02
−2	4.37	4.37	3.84
0	5.09	5.09	4.63
2	5.52	5.49	5.23
4	5.92	5.84	5.72

still over the "low" *SNR* region and as *SNR* increases the two methods become equivalent as far as the offered mutual information is concerned. No results for globally optimal precoding are presented here, due to the complexity involved.

Finally, in Table V we present preliminary results for the mutual information in b/s/Hz for random fading channels with independently varying complex Gaussian (Rayleigh) fading assumed constant within each codeword transmission. For each channel realization a PGP and globally optimal precoder is determined, then the corresponding mutual information for each case, including the no precoding one, is evaluated. This process is repeated for multiple channels and the average mutual information is then found. In Table V, results for no precoding, PGP, and globally optimal precoder are presented for different *SNR* values. We observe very close performance agreement between PGP and the globally optimal precoder for all *SNR*, as expected. In addition, a gain on the order of 15% is achieved for PGP and the globally optimal precoder over the no precoding case at low *SNR* values.

IV. Conclusion

In this paper, the problem of designing a linear precoder for MIMO systems employing LDPC codes is addressed under the constraint of minimizing the dependence between the system's receiving branches, thus reducing the relevant transmitter and receiver complexities. This approach sees the overall precoding problem in an LDPC-coded system from a brand new angle allowing for a more practical deployment of higher dimension MIMO systems and higher QAM constellation sizes with very good performance over a wide *SNR* range.

We show that a Pareto surface optimization leads to an intractable problem, in general. We then target a generalization of BICMB (which is the only tractable solution to the $\lambda = 1$ Pareto problem) and show that this offers a solution to a very meaningful precoding optimization problem that allows for inter-group independence between different transmitting-receiving antenna pairs in the virtual channel domain. We call the new precoding solution PGP and we show, based on numerical results, that PGP offers indeed excellent performance, while its computational complexity is significantly reduced compared to the original solution. For the results presented, PGP is shown to attain mutual information very close to the globally optimal precoder and higher than MDP by 0.7–0.8 b/s/Hz in medium *SNR* ranges. Thus, based on presented evidence, PGP is a very good candidate for almost optimal precoding performance in LDPC coded systems with relatively low system complexity at both the transmitter and receiver. Our future work will look at generalizing the presented approach to higher size modulation constellations, e.g., QAM with $M \geq 16$.

Appendix A
Proof of Lemma 2.1

For the *if* part, which is easier, let $\mathbf{x}_{\mathcal{S}}$ be the restriction of the input vector $\mathbf{x}$ to the set of indices described by $\mathcal{S}$ (which is a subset of $\{1,2,\cdots,N_t\}$, using the notation of (3)). We need to prove that if an IGP exists in a MIMO system with output groups $\{\mathbf{y}_{\mathcal{S}_{y_i}}\}_{i=1}^{N_g}$, then the output $\mathbf{y}$ pdf, $p(\mathbf{y})$, can be written as $\prod_{m=1}^{N_g} p(\mathbf{y}_{\mathcal{S}_m})$, i.e., the output groups are independent. Recalling that an IGP with N_g groups, $\mathcal{P}$, is defined as follows: $\mathcal{P} = \{\mathcal{S}_1,\cdots,\mathcal{S}_{N_g}\}$, where each group in the partition, $\mathcal{S}_i$ with $1 \leq i \leq N_{vi}$, comprises N_{vi} unique, non-intersecting input-output vector pairs (with $\sum_{i=1}^{N_g} N_{vi} = N_t$), $(\mathbf{x}_{\mathcal{S}_i}, \mathbf{y}_{\mathcal{S}_i})$, which satisfy $\cup_{i=1}^{N_g} \mathcal{N}(\mathbf{x}_{\mathcal{S}_i}) = \cup_{i=1}^{N_g} \mathcal{N}(\mathbf{y}_{\mathcal{S}_i}) = \{1,2,\cdots,N_t\}$ and $\mathcal{N}(\mathbf{x}_{\mathcal{S}_i}) \cap_{j\neq i} \mathcal{N}(\mathbf{x}_{\mathcal{S}_j}) = \mathcal{N}(\mathbf{y}_{\mathcal{S}_i}) \cap_{j\neq i} \mathcal{N}(\mathbf{y}_{\mathcal{S}_j}) = \emptyset$, where $\mathcal{N}(\cdot)$ represents the index set (nodes) present in the argument set (within the parenthesis). In addition, IGP requires that each input node, n, in a partition set $\mathcal{S}_i$ is connected only to the corresponding outputs in $\mathcal{S}_i$, in other words, the elements of $\mathbf{V}_G$ corresponding to all other outputs are set to zero, i.e., $\mathbf{V}_G(n, j) = 0$, for output nodes j with $j \notin \mathcal{N}(\mathbf{y}_{\mathcal{S}_i})$, and where $\mathbf{V}_G(n, j)$ represents the ith row, jth column entry of matrix $\mathbf{V}_G$ of (3). All the IGP schemes considered herein use square group structure, i.e., with same number of input and output elements in each group. In other words, $|\mathcal{N}(\mathbf{x}_{\mathcal{S}_i})| = |\mathcal{N}(\mathbf{y}_{\mathcal{S}_i})| = N_{vi}$, with $\sum_{i=1}^{N_g} N_{vi} = N_t$. We have

$$\begin{aligned} p(\mathbf{y}) &= p\left(\mathbf{y}_1,\cdots,\mathbf{y}_{N_g}\right) \\ &= \sum_{\mathbf{x}_{\mathcal{S}_1}} \cdots \sum_{\mathbf{x}_{\mathcal{S}_{N_g}}} p\left(\mathbf{y}_{\mathcal{S}_1},\cdots,\mathbf{y}_{\mathcal{S}_{N_g}} | \mathbf{x}_{\mathcal{S}_1},\cdots,\mathbf{x}_{\mathcal{S}_{N_g}}\right) \cdot \\ &\quad \Pr\left(\mathbf{x}_{\mathcal{S}_1},\cdots,\mathbf{x}_{\mathcal{S}_{N_g}}\right) \\ &= \sum_{\mathbf{x}_{\mathcal{S}_1}} \cdots \sum_{\mathbf{x}_{\mathcal{S}_{N_g}}} \prod_{m=1}^{N_g} p\left(\mathbf{y}_{\mathcal{S}_m} | \mathbf{x}_{\mathcal{S}_m}\right) \Pr\left(\mathbf{x}_{\mathcal{S}_m}\right) \\ &= \prod_{m=1}^{N_g} p\left(\mathbf{y}_{\mathcal{S}_m}\right), \end{aligned} \tag{24}$$

where we have used the conditional independence of outputs from inputs which do not enter their antennas, since they belong to different input groups.

For the *only if* part, we will use mathematical induction on N_g. We also need to invoke the fact that the noise in the MIMO channel in (3) is AWGN and thus the conditional pdf of the output $\mathbf{y}$ given the input vector $\mathbf{x}$ is circular complex Gaussian. For $N_g = 2$ the statement becomes as follows, after setting $\tilde{\mathbf{H}} \doteq \Sigma_H \Sigma_G \mathbf{V}_G^h$, employing the model of (3): If

$$p(\mathbf{y}) = \prod_{l=1}^{2} p\left(\mathbf{y}_{\mathcal{S}_l}\right) \tag{25}$$

where $\mathbf{y}_{\mathcal{S}_1}, \mathbf{y}_{\mathcal{S}_2}$ represent a partition of the output vector $\mathbf{y}$, then an IGP with $N_g = 2$ needs to exist in the MIMO system. For the model in (3) we can write (here $N_v = \min\{N_t, N_r\}$, where min stands for the minimum of a set of elements)

$$\begin{aligned} p(\mathbf{y}) &= \sum_{\mathbf{x}} p(\mathbf{y}|\mathbf{x}) \Pr(\mathbf{x}) \\ &= \frac{1}{M^{N_v}} \frac{1}{\pi^{N_v} \sigma_n^{2N_v}} \sum_{\mathbf{x}} \exp\left(-\frac{\|\mathbf{y} - \tilde{\mathbf{H}}\mathbf{x}\|^2}{\sigma_n^2}\right). \end{aligned} \tag{26}$$

Let us separate the $\mathbf{x}$ vector into three, in general, parts: $\mathbf{x}_{\mathcal{S}_1}, \mathbf{x}_{\mathcal{S}_2}, \mathbf{x}_{\mathcal{S}_c}$, representing inputs entering only outputs in $\mathbf{x}_{\mathcal{S}_1}$, $\mathbf{x}_{\mathcal{S}_2}$, and inputs entering both, respectively. Then,

$$\begin{aligned} p(\mathbf{y}) = &\frac{1}{M^{N_v}} \frac{1}{\pi^{N_v} \sigma_n^{2N_v}} \sum_{\mathbf{x}_{\mathcal{S}_1}} \sum_{\mathbf{x}_{\mathcal{S}_2}} \sum_{\mathbf{x}_{\mathcal{S}_c}} \exp \\ &\left(-\frac{\left\|\mathbf{y}_{\mathcal{S}_1} - \tilde{\mathbf{H}}_{\mathcal{S}_1}\mathbf{x}\right\|^2}{\sigma_n^2}\right) \cdot \exp\left(-\frac{\left\|\mathbf{y}_{\mathcal{S}_2} - \tilde{\mathbf{H}}_{\mathcal{S}_2}\mathbf{x}\right\|^2}{\sigma_n^2}\right), \end{aligned} \tag{27}$$

where $\tilde{\mathbf{H}}_{\mathcal{S}_l}$, $l = 1, 2$ represents the rows of $\tilde{\mathbf{H}}$ that correspond to the OIGP l. We can now further break down $\tilde{\mathbf{H}}_{\mathcal{S}_l}\mathbf{x}$ as follows

$$\tilde{\mathbf{H}}_{\mathcal{S}_l}\mathbf{x} = \tilde{\mathbf{H}}_{\mathcal{S}_l}\mathbf{x}_{\mathcal{S}_l} + \tilde{\mathbf{H}}_{\mathcal{S}_c}\mathbf{x}_{\mathcal{S}_c}, \tag{28}$$

where $\tilde{\mathbf{H}}_{\mathcal{S}_{l1}}, \tilde{\mathbf{H}}_{\mathcal{S}_{lc}}$ are appropriate matrices derived from $\tilde{\mathbf{H}}_{\mathcal{S}_l}$. Substituting this into the equation and after expanding the norms, we get

$$\begin{aligned} p(\mathbf{y}) = &\frac{1}{M^{N_v}} \frac{1}{\pi^{N_v} \sigma_n^{2N_v}} \sum_{\mathbf{x}_{\mathcal{S}_1}} \sum_{\mathbf{x}_{\mathcal{S}_2}} \sum_{\mathbf{x}_{\mathcal{S}_c}} \exp\left(\left(-\frac{1}{\sigma_n^2}\right)\left\{\sum_{l=1}^{2}\right.\right. \\ &\left[\|\mathbf{y}_{\mathcal{S}_l}\|^2 + \left\|\tilde{\mathbf{H}}_{\mathcal{S}_{l1}}\mathbf{x}_{\mathcal{S}_l}\right\|^2 + \left\|\tilde{\mathbf{H}}_{\mathcal{S}_{lc}}\mathbf{x}_{\mathcal{S}_c}\right\|^2\right. \\ &\quad -2\left(\mathrm{Re}(\mathbf{y}_{\mathcal{S}_l}^h \tilde{\mathbf{H}}_{\mathcal{S}_{l1}} \mathbf{x}_{\mathcal{S}_l})\right) \\ &\left.\left.\left. -2\left(\mathrm{Re}(\mathbf{y}_{\mathcal{S}_l}^h \tilde{\mathbf{H}}_{\mathcal{S}_{lc}} \mathbf{x}_{\mathcal{S}_c})\right) + 2\mathrm{Re}\left(\mathbf{x}_{\mathcal{S}_l}^h \tilde{\mathbf{H}}_{\mathcal{S}_{l1}}^h \tilde{\mathbf{H}}_{\mathcal{S}_{lc}} \mathbf{x}_{\mathcal{S}_c}\right)\right]\right\}\right). \end{aligned} \tag{29}$$

Summing first over $\mathbf{x}_{\mathcal{S}_c}$, we observe that due to the term $\exp\left(-\frac{\sum_{l=1}^{2} -2\mathrm{Re}(\mathbf{y}_{\mathcal{S}_l}^h \tilde{\mathbf{H}}_{\mathcal{S}_{lc}} \mathbf{x}_{\mathcal{S}_c})}{\sigma_n^2}\right)$, the summation over $\mathbf{x}_{\mathcal{S}_c}$ when $\mathcal{S}_c \neq \emptyset$ will result in a function of both $\mathbf{y}_{\mathcal{S}_1}, \mathbf{y}_{\mathcal{S}_2}$ which is not a product of two factors containing one of the two vectors each. In other words, the two output groups cannot be independent, except if $\mathcal{S}_c = \emptyset$, in which case we have an IGP. This proves our assertion.

Assume that the statement is true for $N_g = k$: If

$$p(\mathbf{y}) = \prod_{l=1}^{k} p\left(\mathbf{y}_{\mathcal{S}_l}\right) \tag{30}$$

where $\mathbf{y}_{\mathcal{S}_1}, \cdots, \mathbf{y}_{\mathcal{S}_k}$ represent a partition of the output vector $\mathbf{y}$, then an IGP with $N_g = k$ needs to exist in the MIMO system. Then, we need to prove that for $N_g = k+1$, the statement is still true: If

$$p(\mathbf{y}) = \prod_{l=1}^{k+1} p\left(\mathbf{y}_{\mathcal{S}_l}\right) \tag{31}$$

where $\mathbf{y}_{\mathcal{S}_1}, \cdots, \mathbf{y}_{\mathcal{S}_{k+1}}$ represent a partition of the output vector $\mathbf{y}$, then an IGP with $N_g = k+1$ needs to exist in the MIMO system. We prove this by invoking the already proved case of $N_g = 2$, together with the assumption for $N_g = k$. Writing

$$p(\mathbf{y}) = p\left(\mathbf{y}_{\mathcal{S}_1}\right)\left(\prod_{l=2}^{k+1} p\left(\mathbf{y}_{\mathcal{S}_l}\right)\right), \tag{32}$$

we see that we have an OIGP with $N_g = 2$ output groups defined by $\mathbf{y}_{\mathcal{S}_1}$ and the rest of the outputs, $\{\mathbf{y}_{\mathcal{S}_l}\}_{l=2}^{N_g}$. Thus, based on the proven $N_g = 2$ case, the system entails an IGP with $N_g = 2$. Now apply Lemma 1 one more time to the output groups $\{\mathbf{y}_{\mathcal{S}_l}\}_{l=2}^{N_g}$, of size k. Thus, because we assumed that for $N_g = k$ Lemma 1 is valid, we get that there is an IGP for $\{\mathbf{y}_{\mathcal{S}_l}\}_{l=2}^{N_g}$. Totally, including $\{\mathbf{y}_{\mathcal{S}_l}\}_{l=2}^{N_g}$, and taking the two partitions together, the overall MIMO system has an IGP of size $N_g = k+1$ groups, proving the case $N_g = k+1$.

Appendix B
Proof of Lemma 2.2

For an input-output MIMO IGP, $\mathcal{P} = \{\mathcal{S}_1, \cdots, \mathcal{S}_{N_g}\}$, we have N_g groups of nodes at both input and output domain. Each group in the partition, $\mathcal{S}_i$ with $1 \leq i \leq N_g$, comprises N_{gi} unique, non-intersecting input-output vector pairs, $(\mathbf{x}_{\mathcal{S}_i}, \mathbf{y}_{\mathcal{S}_i})$, as explained in the main text, with $\sum_{i=1}^{N_g} N_{vi} = N_t$. Let $n_{in}(i,j), n_o(i,j)$ be the mapping from the partition groups to input and output indices, respectively. In other words, $n_{in}(i,j)$ represents the input index (taking values in $\{1, \cdots, N_v\}$) of the jth ($1 \leq j \leq N_{vi}$) input of the ith group ($1 \leq i \leq N_g$), and similarly for $n_o(i,j)$. Also, let $\mathcal{S}(l)(1 \leq l \leq N_t)$, be the mapping from the lth input $\mathbf{x}_l$ to its group number. Finally, let $\mathbf{x}_{\mathcal{S}(i)} = \{\mathbf{x}_{n_{in}(i,1)}, \cdots, \mathbf{x}_i, \cdots, \mathbf{x}_{n_{in}(i,|\mathcal{S}(i)|)}\}$ be the set of input nodes present in the ith group of the IGP. We can then write, employing $p(\cdot)$ for the pdf of the variables in the parenthesis

$$\begin{aligned} \Pr(\mathbf{x}_i|\mathbf{y}) &= \frac{p(\mathbf{y}, \mathbf{x}_i)}{p(\mathbf{y})} \\ &= \frac{\sum_{l \neq i, l \in \mathcal{S}(i)} \sum_{m=1, m \neq i}^{N_g} p\left(\mathbf{y}|\mathbf{x}_{\mathcal{S}_1}, \cdots, \mathbf{x}_{\mathcal{S}(i)}, \cdots, \mathbf{x}_{\mathcal{S}_{N_g}}\right) \Pr(\mathbf{x})}{p(\mathbf{y})} \\ &= \frac{1}{\prod_{m=1}^{N_g} p\left(\mathbf{y}_{\mathcal{S}_m}\right)} \cdot \sum_{l \neq i, l \in \mathcal{S}(i)} \sum_{m=1, m \neq i}^{N_g} \\ &\quad \prod_{m=1, m \neq i}^{N_g} p\left(\mathbf{y}_{\mathcal{S}_m}|\mathbf{x}_{\mathcal{S}_m}\right) \Pr\left(\mathbf{x}_{\mathcal{S}_m}\right) p\left(\mathbf{y}_{\mathcal{S}_i}|\mathbf{x}_{\mathcal{S}_i}\right) \Pr\left(\mathbf{x}_{\mathcal{S}_i}\right) \end{aligned}$$

$$= \frac{1}{\prod_{m=1}^{N_g} p\left(\mathbf{y}_{\mathcal{S}_m}\right)} \sum_{l\neq i, l\in\mathcal{S}(i)} \sum_{m=1,m\neq i}^{N_g} \prod_{m=1,m\neq i}^{N_g} p\left(\mathbf{y}_{\mathcal{S}_m}|\mathbf{x}_{\mathcal{S}_m}\right)\Pr\left(\mathbf{x}_{\mathcal{S}_m}\right) p\left(\mathbf{y}_{\mathcal{S}_i}|\mathbf{x}_{\mathcal{S}_i}\right)\Pr\left(\mathbf{x}_{\mathcal{S}_i}\right)$$

$$= \frac{p\left(\mathbf{y}_{\mathcal{S}_i},\mathbf{x}_i\right)}{p\left(\mathbf{y}_{\mathcal{S}(i)}\right)} = \Pr\left(\mathbf{x}_i|\mathbf{y}_{\mathcal{S}(i)}\right), \tag{33}$$

where the last four equations come from the conditional independence of outputs given inputs of the other groups and the fact that the inputs are independent, the independence of the outputs of different groups, the total probability law, and the definition of conditional probability, respectively. Note that $\Pr(\mathbf{x})$ is the probability of the overall input vector (N_t components), which due to independent input assumption and equally likely inputs, can be written as $\Pr(\mathbf{x}) = \prod_{l=1}^{N_t}\Pr(\mathbf{x}_i) = \prod_{k=1}^{N_g}\Pr(\mathbf{x}_{\mathcal{S}_k}) = \frac{1}{M^{N_t}}$.

This proves that independence between groups in the partition results in a MAP detector in which, only the outputs of the group to which an input symbol belongs are relevant for the symbol's MAP detector probability evaluation. This phenomenon is the main reason behind the complexity reduction offered by, e.g., PGP.

APPENDIX C
PROOF OF LEMMA 2.3

Consider the output of antenna $i (1 \leq i \leq N_r)$ in the model of (1), written as

$$\mathbf{y}_i = \mathbf{H}_i\mathbf{G}\mathbf{x} + \mathbf{n}_i. \tag{34}$$

This represents an $N_t \times 1$ MIMO system. Thus, based on the general results presented in [12], we see that the gradient of $H(\mathbf{y}_i)$ with respect to $\mathbf{W}_i = \mathbf{G}^h\mathbf{H}_i^h\mathbf{H}_i\mathbf{G}$ equals

$$\nabla_{\mathbf{W}_i} H_i(\mathbf{y}_i) = \frac{1}{\sigma_n^2}\Phi_{\mathbf{xx}^h,\mathcal{S}=\{i\}}. \tag{35}$$

It still remains to show that $\nabla_{\mathbf{W}_i}H_i(\mathbf{y}_i) = \nabla_{\mathbf{W}}H_i(\mathbf{y}_i)$. Toward this end, we notice that

$$\mathbf{W} = \mathbf{G}^h\mathbf{H}^h\mathbf{H}\mathbf{G} = \mathbf{G}^h\left(\sum_{i=1}^{N_r}\mathbf{H}_i^h\mathbf{H}_i\right)\mathbf{G} = \sum_{i=1}^{N_r}\mathbf{W}_i, \tag{36}$$

from which we get

$$\text{dvec}(\mathbf{W}) = \sum_i \text{dvec}(\mathbf{W}_i) \tag{37}$$

where $\text{dvec}(\mathbf{F})$ represents the vector differential of the (complex) matrix function $\mathbf{F}$. This means that the derivative of $\mathbf{W}$ with respect to $\mathbf{W}_i$, $\mathcal{D}_{\mathbf{W}_i}\mathbf{W} = I_{N_t}$, and also that $\mathcal{D}_{\mathbf{W}_i}\mathbf{W}^* = \mathbf{0}_{N_t}$, with $\mathbf{W}^*$ representing the conjugate matrix of $\mathbf{W}$. Now, $H(\mathbf{y}_i)$ is a function of $\mathbf{W}_i$ only [12], but $\mathbf{W}_i$ is a function of $\mathbf{W}$. Thus, $H(\mathbf{y}_i)$ is a composite function of $\mathbf{W}$ and then based on the chain rule [19],

$$\mathcal{D}_{\mathbf{W}}H(\mathbf{y}_i) = \mathcal{D}_{\mathbf{W}_i}H(\mathbf{y}_i)\mathcal{D}_{\mathbf{W}}\mathbf{W}_i + \mathcal{D}_{\mathbf{W}_i^*}H(\mathbf{y}_i)\mathcal{D}_{\mathbf{W}}\mathbf{W}_i^*. \tag{38}$$

Notice that based on the conjugation property of differentiation, $\mathcal{D}_{\mathbf{W}}\mathbf{W}_i^* = \mathbf{0}_{N_t}$. From the last equation, we can get the desired result that $\nabla_{\mathbf{W}_i}H_i(\mathbf{y}_i) = \nabla_{\mathbf{W}}H_i(\mathbf{y}_i)$ by invoking the transformation for scalar functions from derivatives to gradients. Concerning the concavity of $H(\mathbf{y}_i)$ with respect to $\mathbf{W}$, we note that $H_i(\mathbf{y}_i)$ is concave with respect to $\mathbf{W}_i$ as it equals $I(\mathbf{x};\mathbf{y}_i)$ minus a constant, where $I(\mathbf{x};\mathbf{y}_i)$ represents the mutual information from $\mathbf{x}$ to $\mathbf{y}_i$, and it is known from [12] that $I(\mathbf{x};\mathbf{y}_i)$ is concave with respect to $\mathbf{W}_i$. As $H_i(\mathbf{y}_i)$ only depends on $\mathbf{W}$ through $\mathbf{W}_i$, it becomes evident that $H_i(\mathbf{y}_i)$ is also a concave function of $\mathbf{W}$. This completes the proof of Lemma 2.3.

APPENDIX D
PROOF OF LEMMA 2.4

To prove this, we note that the sum of concave functions of a matrix variable is also concave in that variable. Thus, based on Lemma 2.3, $H_I(\mathbf{y}) = \sum_{i=1}^{N_r}H_i(\mathbf{y}_i)$ is a concave function of $\mathbf{W}$ with gradient that equals the sum of the gradients of $H_i(\mathbf{y}_i)$ with respect to $\mathbf{W}$ which by invoking Lemma 2.3 proves the assertion.

APPENDIX E
PROOF OF THEOREM 2.5

The proof of this Theorem is as follows: First, write the Lagrangian of the "original" problem in (10), as

$$H(\mathbf{y}) - \lambda\sum_{i=1}^{N_r}H_i(\mathbf{y}_i) - \nu\left(\text{tr}(\mathbf{G}\mathbf{G}^h) - N_t\right), \tag{39}$$

where without loss we employed the constraint $\text{tr}(\mathbf{G}\mathbf{G}^h) \leq N_t$, and where ν is a positive parameter. The critical points are now found by setting the gradient of the Lagrangian with respect to $\mathbf{G}$ equal to zero, as part of the Karush-Kuhn-Tucker (KKT) conditions. Noticing that $\nabla_{\mathbf{G}}tr(\mathbf{G}\mathbf{G}^h) = \mathbf{G}$, and upon invoking the two previous Lemmas of this paper, (35), and the fact that $\nabla_{\mathbf{W}}H(\mathbf{y}) = \frac{1}{\sigma_n^2}\Phi_{\mathbf{xx}^h,\mathcal{S}}$ [12], we get the desired critical point equation described in Theorem 2.5.

APPENDIX F
PROOF OF THEOREM 2.6

For $\lambda = 1$, the "equivalent" optimization problem described by (11) becomes

$$\begin{array}{ll} \underset{\mathbf{W}}{\text{maximize}} & H(\mathbf{y}) - \sum_{i=1}^{N_r}H_i(\mathbf{y}_i) \\ \text{subject to} & \mathbf{W}^h = \mathbf{W} \\ \text{and} & \text{tr}\left(\Sigma_G^2\right) = N_t \\ \text{and} & 0 \leq \lambda < 1. \end{array} \tag{40}$$

Performing the maximization first over $\mathbf{V}_G$, then over Σ_G we see that

$$\max_{\Sigma_G}\left\{\max_{\mathbf{V}_G}\left\{H(\mathbf{y}) - \sum_{i=1}^{N_r}H_i(\mathbf{y}_i)\right\}\right\} = 0, \tag{41}$$

since $H(\mathbf{y}) \leq \sum_{i=1}^{N_r}H_i(\mathbf{y}_i)$, with equality if and only if $\Pr(\mathbf{y}) = \prod_{i=1}^{N_r}\Pr(\mathbf{y}_i)$ [20]. Thus, the maximum in (41) is achievable if and only if the receiving outputs $\mathbf{y}_i (i = 1, \cdots, N_r)$ are independent. For this to happen, $\mathbf{V}_G$ needs to be equal to the identity matrix, $\mathbf{I}$. Taking into account the power constraint in (11) completes the proof.

APPENDIX G
PROOF OF THEOREM 2.7

The proof of the Theorem is based on contradiction of Schur-type complement conditions [16] and Kronecker product properties (e.g., [19]). We also exploit properties of MMSE. Let us denote by $\otimes$ the Kronecker matrix product. The complex Hessian matrix of a scalar function of a complex matrix $\mathbf{Z}$, $\mathcal{CH}(f(\mathbf{Z}))$ is defined as [19]:

$$\mathcal{CH}\left(f(\mathbf{Z})\right) = \begin{bmatrix} \mathcal{H}_{\mathbf{Z},\mathbf{Z}^*} f & \mathcal{H}_{\mathbf{Z}^*,\mathbf{Z}^*} f \\ \mathcal{H}_{\mathbf{Z},\mathbf{Z}} f & \mathcal{H}_{\mathbf{Z}^*,\mathbf{Z}} f \end{bmatrix}, \tag{42}$$

where $\mathbf{A}^*$ is the conjugate of matrix $\mathbf{A}$. In our case,

$$f(\mathbf{W}) = H(\mathbf{y}) - \sum_{i=1}^{N_r} H_i(\mathbf{y}_i). \tag{43}$$

By applying the results for the complex Hessian matrix for $\mathbf{y}$, and $\mathbf{y}_i (i = 1, \cdots, N_r)$ from [12], complex Hessian matrix properties, and generalizing the results of Lemma 2, we can see that the different parts of the composite Hessian matrix (42) of $H(\mathbf{y}) - \sum_{i=1}^{N_r} H_i(\mathbf{y}_i)(\lambda = 1)$ with respect to $\mathbf{W}$, denoted as $\mathcal{H}_{\mathbf{W},\mathbf{W}^*} f$, is given as follows, employing results on the Hessian matrix of the mutual information with respect to $\mathbf{W}$ from [12], and the previous lemmas and theorems of this paper:

$$\begin{aligned} &\mathcal{H}_{\mathbf{W},\mathbf{W}^*}\left(f(\mathbf{W})\right) \\ &\quad = \left(\mathcal{H}_{\mathbf{W}^*,\mathbf{W}}\left(f(\mathbf{W})\right)\right)^* \\ &\quad = -\frac{1}{\sigma_n^2}\mathbb{E}\Bigg(\mathbf{\Phi}^*_{\mathbf{xx}^h}(\mathbf{y}) \otimes \mathbf{\Phi}_{\mathbf{xx}^h}(\mathbf{y}) \\ &\qquad - \sum_{i=1}^{N_r} \mathbf{\Phi}^*_{\mathbf{xx}^h,\mathcal{S}=\{i\}}(\mathbf{y}_i) \otimes \mathbf{\Phi}_{\mathbf{xx}^h,\mathcal{S}=\{i\}}(\mathbf{y}_i)\Bigg), \end{aligned} \tag{44}$$

and

$$\begin{aligned} &\mathcal{H}_{\mathbf{W},\mathbf{W}}\left(f(\mathbf{W})\right) \\ &\quad = \left(\mathcal{H}_{\mathbf{W}^*,\mathbf{W}^*}\left(f(\mathbf{W})\right)\right)^* \\ &\quad = -\frac{1}{\sigma_n^2}\mathbb{E}\Bigg(\mathbf{\Psi}_{\mathbf{xx}^t}(\mathbf{y}) \otimes \mathbf{\Psi}^*_{\mathbf{xx}^t}(\mathbf{y}) \\ &\qquad - \sum_{i=1}^{N_r} \mathbf{\Psi}_{\mathbf{xx}^t,\mathcal{S}=\{i\}}(\mathbf{y}_i) \otimes \mathbf{\Psi}^*_{\mathbf{xx}^t,\mathcal{S}=\{i\}}(\mathbf{y}_i)\Bigg), \end{aligned} \tag{45}$$

where

$$\mathbf{\Psi}_{\mathbf{xx}^t}(\mathbf{y}_\mathcal{S}) \doteq \mathbb{E}\left(\left(\mathbf{x} - \mathbb{E}(\mathbf{x}|\mathbf{y}_\mathcal{S})\right)\left(\mathbf{x} - \mathbb{E}(\mathbf{x}|\mathbf{y}_\mathcal{S})\right)^t |\mathbf{y}_\mathcal{S}\right), \tag{46}$$

$$\mathbf{\Psi}_{\mathbf{xx}^t,\mathcal{S}} \doteq \mathbb{E}\left(\mathbf{\Psi}_{\mathbf{xx}^t}(\mathbf{y}_\mathcal{S})\right), \tag{47}$$

similarly to (12), (13), and where $\mathbf{A}^t$ is the transpose of matrix $\mathbf{A}$. It is worth stressing that $\mathcal{CH}(f\mathbf{W}))$ in our case can be put in the form required by the complex Schur complement, i.e.

$$\mathcal{CH}\left(f(\mathbf{W})\right) = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^h & \mathbf{C} \end{bmatrix}, \tag{48}$$

since $(\mathbf{\Psi}_{\mathbf{xx}^t}(\mathbf{y}_\mathcal{S}) \otimes \mathbf{\Psi}^*_{\mathbf{xx}^t}(\mathbf{y}_\mathcal{S}))^t = \mathbf{\Psi}_{\mathbf{xx}^t}(\mathbf{y}_\mathcal{S}) \otimes \mathbf{\Psi}^*_{\mathbf{xx}^t}(\mathbf{y}_\mathcal{S})$, and that for a matrix in this form to be seminegative definite it is necessary that $\mathbf{A}$ be seminegative definite, based on Schur complement properties [16]. Now notice that

$$\mathbf{\Phi}^*_{\mathbf{xx}^h}(\mathbf{y}) \otimes \mathbf{\Phi}_{\mathbf{xx}^h}(\mathbf{y}) - \sum_{i=1}^{N_r} \mathbf{\Phi}^*_{\mathbf{xx}^h,\mathcal{S}=\{i\}}(\mathbf{y}_i) \otimes \mathbf{\Phi}_{\mathbf{xx}^h,\mathcal{S}=\{i\}}(\mathbf{y}_i) \tag{49}$$

is not a positive semidefinite matrix. To show this, it suffices to show that

$$\mathbf{\Phi}^*_{\mathbf{xx}^h}(\mathbf{y}) \otimes \mathbf{\Phi}_{\mathbf{xx}^h}(\mathbf{y}) - \mathbf{\Phi}^*_{\mathbf{xx}^h,\mathcal{S}=\{1\}}(\mathbf{y}_1) \otimes \mathbf{\Phi}_{\mathbf{xx}^h,\mathcal{S}=\{1\}}(\mathbf{y}_1) \tag{50}$$

is not positive semidefinite, since the residual sum is not positive definite.

Based on properties of MMSE matrices we know that

$$\mathbf{\Phi}_{\mathbf{xx}^h}(\mathbf{y}) - \mathbf{\Phi}_{\mathbf{xx}^h,\mathcal{S}=\{1\}}(\mathbf{y}_1)$$

is negative semidefinite, since $\mathbf{\Phi}_{\mathbf{xx}^h,\mathcal{S}=\{1\}}(\mathbf{y}_1)$ represents the Mean Square Error (MSE) matrix of a suboptimal estimator of $\mathbf{x}$ given $\mathbf{y}$. Then, using properties of Kronecker products we can easily show that

$$\mathbf{\Phi}^*_{\mathbf{xx}^h}(\mathbf{y}) \otimes \mathbf{\Phi}_{\mathbf{xx}^h}(\mathbf{y}) - \mathbf{\Phi}^*_{\mathbf{xx}^h,\mathcal{S}=\{1\}}(\mathbf{y}_1) \otimes \mathbf{\Phi}_{\mathbf{xx}^h,\mathcal{S}=\{1\}}(\mathbf{y}_1) \tag{51}$$

is negative semidefinite. This comes from the following property: If positive semidefinite complex matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ of same size satisfy $\mathbf{A} \prec \mathbf{B}$, and $\mathbf{C} \prec \mathbf{D}$ (here $\prec$ is generalized sense ordering between matrices, as per, e.g., [16], i.e., $\mathbf{A} \prec \mathbf{B}$ means that $\mathbf{B} - \mathbf{A}$ is semipositive definite), then $\mathbf{A} \otimes \mathbf{C} \prec \mathbf{B} \otimes \mathbf{D}$, which can be easily proven as follows: Consider the Kronecker product $(\mathbf{A} - \mathbf{B}) \otimes \mathbf{C} \prec \mathbf{0}$, as the Kronecker product of a semipositive definite matrix $\mathbf{C}$, with a seminegative one $(\mathbf{A} - \mathbf{B})$. Thus, $\mathbf{A} \otimes \mathbf{C} \prec \mathbf{B} \otimes \mathbf{C}$. Similarly, we can show that $\mathbf{B} \otimes \mathbf{C} \prec \mathbf{B} \otimes \mathbf{D}$. Substituting $\mathbf{A} = \mathbf{\Phi}^*_{\mathbf{xx}^h}(\mathbf{y})$, $\mathbf{B} = \mathbf{\Phi}_{\mathbf{xx}^h}(\mathbf{y})$, $\mathbf{C} = \mathbf{\Phi}^*_{\mathbf{xx}^h,\mathcal{S}=\{1\}}(\mathbf{y}_1)$, and $\mathbf{D} = \mathbf{\Phi}_{\mathbf{xx}^h,\mathcal{S}=\{1\}}(\mathbf{y}_1)$ shows that the matrix in (51) is indeed seminegative definite. Thus, $\mathcal{H}_{\mathbf{W},\mathbf{W}^*}(f(\mathbf{W}))$ is also seminegative definite as it is a seminegative definite matrix minus a semipositive definite one, as per (45). After taking expectation in (44), we have a semipositive definite matrix, which contradicts the Schur complement requirement. Thus, our assertion is proven.

APPENDIX H
PROOF OF THEOREM 2.8

It suffices to show that

$$\mathbf{V}\mathbf{V}^h = \mathbf{I},$$

for $\mathbf{V} = \text{diag}[\mathbf{V}_1, \cdots, \mathbf{V}_{N_g}]$. By multiplying $\mathbf{V}$ by $\mathbf{V}^h$ in blocks we see that

$$\mathbf{V}\mathbf{V}^h = \text{diag}\left[\mathbf{V}_1, \mathbf{V}_1^h, \cdots, \mathbf{V}_{N_g}\mathbf{V}_{N_g}^h\right] = \mathbf{I}. \tag{52}$$

Since the component matrices are unitary, i.e. $\mathbf{V}_i\mathbf{V}_i^h = \mathbf{I}$, thus proving our assertion. We do not consider the power constraint proof, since this is obvious due to the component power constraints required by PGP.

REFERENCES

[1] E. Akay, E. Sengul, and E. Ayanoglu, "Bit interleaved coded multiple beamforming," *IEEE Trans. Commun.*, vol. 55, no. 9, pp. 1802–1810, Sep. 2007.

[2] B. Li, H. J. Park, and E. Ayanoglu, "Constellation precoded multiple beamforming," *IEEE Trans. Commun.*, vol. 59, no. 5, pp. 1275–1286, May 2011.

[3] B. Li and E. Ayanoglu, "Diversity analysis of bit-interleaved coded multiple beamforming with orthogonal frequency division multiplexing," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3794–3805, Sep. 2013.

[4] Y. Xin, Z. Wang, and G. Giannakis, "Space-time diversity systems based on linear constellation precoding," *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 294–309, Mar. 2003.

[5] V. Tarokh, N. Seshadri, and A. Calderbank, "Space-time codes for high data rate communications: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

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

[7] Y. Jian, A. Ashihkmin, and N. Sharma, "LDPC codes for flat rayleigh fading channels with channel side information," *IEEE Trans. Commun.*, vol. 56, no. 8, pp. 1207–1213, Aug. 2008.

[8] F. Perez-Cruz, M. Rodriguez, and S. Verdu, "MIMO Gaussian channels with arbitrary inputs: Optimal precoding and power allocation," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 1070–1086, Mar. 2010.

[9] A. Lozano, A. Tulino, and S. Verdu, "Optimal Power Allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3024–3051, Jul. 2006.

[10] M. Payaro and D. Palomar, "On Optimal precoding in linear vector Gaussian channels with arbitrary input distribution," in *Proc. IEEE Int. Symp. Inf. Theory*, 2009, pp. 1085–1089.

[11] E. Ohlmer, U. Wachsmann, and G. Fettweis, "Mutual information maximizing linear precoding for parallel layer MIMO detection," in *Proc. 12th Int. Workshop Signal Process. Adv. Wireless Commun.*, 2011, pp. 346–350.

[12] G. Xiao, Y. Zheng, and Z. Ding, "Globally optimal linear precoders for finite alphabet signals over complex vector Gaussian channels," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3301–3314, Jul. 2011.

[13] M. Lamarca, "Linear precoding for mutual information maximization in MIMO systems," in *Proc. Int. Symp. Wireless Commun. Syst.*, 2009, pp. 26–30.

[14] W. Zeng, G. Xiao, and J. Lu, "A low complexity design of linear precoding for MIMO channels with finite alphabet inputs," *IEEE Wireless Commun. Lett.*, vol. 1, no. 1, pp. 38–42, Feb. 2012.

[15] W. Zeng, G. Xiao, W. Mingxi, and J. Lu, "Linear precoding for finite alphabet inputs over MIMO fading channels with statistical CSI," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 3134–3148, Feb. 2012.

[16] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

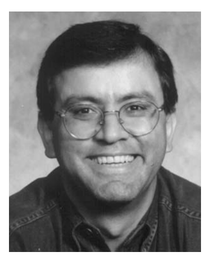
[17] D. Palomar and S. Verdu, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 141–154, Jan. 2006.

[18] L. Xie, J. Ding, and F. Ding, "Gradient based iterative solutions for general linear matrix equations," *Comput. Math. Commun.*, vol. 58, no. 7, pp. 1441–1448, Oct. 2009.

[19] A. HjØrugnes, *Complex-Valued Matrix Derivatives With Applications in Signal Processing and Communications*. Cambridge, UK: Cambridge Univ. Press, 2011.

[20] T. Cover and J. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

Thomas Ketseoglou (S'85–M'91–SM'96) received the B.S. degree from the University of Patras, Patras, Greece, in 1982, the M.S. degree from the University of Maryland, College Park, MD, USA, in 1986, and the Ph.D. degree from the University of Southern California, Los Angeles, CA, USA, in 1990, all in electrical engineering. He worked in the wireless communications industry, and held senior level positions with Siemens, Ericsson, Rockwell, and Omnipoint. From 1996 through 1998, he participated in TIA TR45.5 (now 3GPP2) 3G standardization, making significant contributions to the cdma2000 standard. He has been inventor and co-inventor on several essential patents in wireless communications. Since September 2003, he has been with the Electrical and Computer Engineering Department, California State Polytechnic University, Pomona, CA, USA, where he is a Professor. He spent his sabbatical leave in 2011 at the Digital Technology Center, University of Minnesota, Minneapolis, MN, USA, where he taught digital communications and performed research on network data and machine learning techniques. He is a part-time Lecturer at the University of California, Irvine. His teaching and research interests are in wireless communications, signal processing, and machine learning, with current emphasis on MIMO, optimization, localization, and link prediction.



Ender Ayanoglu (S'82–M'85–SM'90–F'98) received the B.S. degree from the Middle East Technical University, Ankara, Turkey, in 1980 and the M.S. and Ph.D. degrees from Stanford University, Stanford, CA, USA, in 1982 and 1986, respectively, all in electrical engineering. He was with the Communications Systems Research Laboratory, part of AT&T Bell Laboratories, Holmdel, NJ, USA, until 1996, and Bell Labs, Lucent Technologies, until 1999. From 1999 until 2002, he was a Systems Architect at Cisco Systems, Inc. Since 2002, he has been a Professor with the Department of Electrical Engineering and Computer Science, University of California, Irvine, CA, USA, where he served as the Director of the Center for Pervasive Communications and Computing and held the Conexant-Broadcom Endowed Chair during 2002–2010. He was the recipient of the IEEE Communications Society Stephen O. Rice Prize Paper Award in 1995 and the IEEE Communications Society Best Tutorial Paper Award in 1997. From 1993 until 2014, he was an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS and served as its Editor-in-Chief from 2004 to 2008. Currently, he is serving as a Senior Editor of the same journal. From 1990 to 2002, he served on the Executive Committee of the IEEE Communications Society Communication Theory Committee, and from 1999 to 2001, he was its Chair.