Linear Precoding for MIMO with LDPC Coding and Reduced Receiver Complexity

T. Ketseoglou
ECE
California State University, Pomona, California

E. Ayanoglu
CPCC, EECS
University of California, Irvine, California

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, and thus results in independent transmitting/receiving streams between groups. We show with simulated results that PGP offers almost optimal performance, albeit with significant reduction both in the precoder optimization, and LDPC EXIT chart based decoding complexities.**

I. INTRODUCTION AND PROBLEM STATEMENT

In the area of MIMO research, BICMB [1], [2] has shown great potential for application, due its excellent diversity gains and its low relative decoding simplicity. 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) chats [3], [4]. Recently, linear precoding techniques were presented [5] capable of achieving mutual information rates much higher than the previously presented Mercury Water Filling (MWF) [6] techniques, by introducing input symbol correlation through a unitary input transformation together with channel weight adjustment. However, the gains presented in [5] come 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 [5] requires a significant computational complexity increase at the receiver, due to the dependence 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 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 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. 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 independence among different groups and it can be successfully applied to higher QAM constellations with $M \geq 16$.

II. LINEAR PRECODER OPTIMIZATION WITH REDUCED COMPLEXITY

The N_t transmit antenna, N_r receive antenna MIMO model is described by the following equation

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

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, $\mathbf{G}$ is the precoder matrix of size $N_t \times N_t$, $\mathbf{x}$ is the $N_t \times 1$ data vector with independent components 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

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

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

An equivalent model is given by [5]

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

where $\mathbf{\Sigma}_H$ and $\mathbf{\Sigma}_G$ are diagonal matrices containing the singular values of $\mathbf{H}$, $\mathbf{G}$, respectively and $\mathbf{V}_G$ is the matrix of the right singular vectors of $\mathbf{G}$. When LDPC 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, $\mathbf{y}$ [3], [4]. It is shown [5] that the mutual information between $\mathbf{x}$ and $\mathbf{y}$, $I(\mathbf{x};\mathbf{y})$, is only a function of $\mathbf{W} = \mathbf{V}_G\boldsymbol{\Sigma}_H^2\boldsymbol{\Sigma}_G^2\mathbf{V}_G^h$. The optimal precoder, $\mathbf{G}$ is found by solving:

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

The solution to (4) results in exponential complexity at both transmitter, receiver.

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. 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

$$H(\mathbf{y}) - H_I(\mathbf{y}) = H(\mathbf{y}) - \sum_{i=1}^{N_r} H_i(\mathbf{y}_i), \tag{5}$$

where $H(\cdot)$ represents the entropy of the argument in the parenthesis, $\mathbf{y}_i$ is the i receiving branch, and we defined $H_I(\mathbf{y}) \doteq \sum_{i=1}^{N_r} H_i(\mathbf{y}_i)$. By invoking the Pareto optimal surface criterion of optimality [7], the following equivalent optimization problems need to be solved in order to find this type of Pareto optimal precoder $\mathbf{G}$, or $\mathbf{W}$, respectively:

$$\begin{aligned} &\underset{\mathbf{G}}{\text{maximize}} && H(\mathbf{y}) - \lambda \sum_{i=1}^{N_r} H_i(\mathbf{y}_i) \\ &\text{subject to} && tr(\mathbf{G}\mathbf{G}^h) = N_t \\ &\text{and} && 0 \le \lambda < 1, \end{aligned} \tag{6}$$

called the "original problem," and

$$\begin{aligned} &\underset{\mathbf{V}_G, \boldsymbol{\Sigma}_G}{\text{maximize}} && H(\mathbf{y}) - \lambda \sum_{i=1}^{N_r} H_i(\mathbf{y}_i) \\ &\text{subject to} && tr(\boldsymbol{\Sigma}_G^2) = N_t \\ &\text{and} && 0 \le \lambda < 1, \end{aligned} \tag{7}$$

called the "equivalent problem", where the reception model of (3) is employed. 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 $\mathcal{S}$. The following two Minimum Mean Square Error (MMSE) matrices are instrumental in the results presented herein

$$\boldsymbol{\Phi}_{\mathbf{x}\mathbf{x}^h}(\mathbf{y}_\mathcal{S}) \doteq \mathbb{E}\left((\mathbf{x} - \mathbb{E}(\mathbf{x}|\mathbf{y}_\mathcal{S}))(\mathbf{x} - \mathbb{E}(\mathbf{x}|\mathbf{y}_\mathcal{S}))^h|\mathbf{y}_\mathcal{S}\right), \tag{8}$$

and

$$\boldsymbol{\Phi}_{\mathbf{x}\mathbf{x}^h,\mathcal{S}} \doteq \mathbb{E}(\boldsymbol{\Phi}_{\mathbf{x}\mathbf{x}^h}(\mathbf{y}_\mathcal{S})). \tag{9}$$

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

Lemma 2.1: The functions $H_i(\mathbf{y}_i)$, $i = 1, 2, \cdots, N_t$ are concave functions of $\mathbf{W}$, with gradient $\nabla_\mathbf{W} H_i(\mathbf{y}_i) = \frac{1}{\sigma_n^2}\boldsymbol{\Phi}_{\mathbf{x}\mathbf{x}^h,\mathcal{S}=\{i\}}$. Here i in the subscript stands for selecting the ith receiving branch only.

Lemma 2.2: The function $H_I(\mathbf{y})$ is a concave function of $\mathbf{W}$ with gradient, $\nabla_\mathbf{W} H_I(\mathbf{y}) = \frac{1}{\sigma_n^2}\sum_{i=1}^{N_r}\boldsymbol{\Phi}_{\mathbf{x}\mathbf{x}^h,\mathcal{S}=\{i\}}$.

Theorem 2.3: The optimal precoder of (6) satisfies

$$\mathbf{H}^h\mathbf{H}\mathbf{G}\boldsymbol{\Phi}_{\mathbf{x}\mathbf{x}^h} - \lambda\sum_{i=1}^{N_r}\mathbf{H}_i^h\mathbf{H}_i\mathbf{G}\boldsymbol{\Phi}_{\mathbf{x}\mathbf{x}^h,\mathcal{S}=\{i\}} = \nu\mathbf{G}, \tag{10}$$

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

Theorem 2.4: The case of ordinary BICMB is the solution of the equivalent problem (7) with $\lambda = 1$, i.e., with $\mathbf{V}_G = \mathbf{I}$ being the optimal $\mathbf{V}_G$ and $\boldsymbol{\Sigma}_G$ an arbitrary non-negative element diagonal matrix satisfying $tr(\boldsymbol{\Sigma}_G^2) = N_t$.

Theorem 2.5: The objective in the equivalent problem in (7) is not a concave function of $\mathbf{W}$ when $\lambda = 1$ and $\mathbf{V}_G \neq \mathbf{I}$.

The proofs of these lemmas and theorems are presented in the Appendix. 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 [8] and invoking Hessian matrices and the Schur complement [7].

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 2.3. Further, as BICMB does not fully utilize the benefits of precoding as required by LDPC codes, its utility within the current context is limited. Armed with this result, we next investigate an alternative setup that leads to a different problem with a better solution for our purposes.

B. A Different Optimization Problem

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

$$\mathbf{y}_{\mathcal{S}_i} \cap \mathbf{y}_{\mathcal{S}_j} = \emptyset \quad (i \neq j) \quad \text{and} \quad \cup_{i=1}^{N_g}\mathbf{y}_{\mathcal{S}_i} = \mathbf{y},$$

the following inequality is true for fixed $\mathbf{V}_G$, $\boldsymbol{\Sigma}_G$:

$$H(\mathbf{y}) \le \sum_{i=1}^{N_g} H_i(\mathbf{y}_{\mathcal{S}_i}) \le \sum_{i=1}^{N_r} H_i(\mathbf{y}_i). \tag{11}$$

This is a very fundamental result in Information Theory [9] 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:

$$\begin{aligned} &\underset{\mathbf{V}_G,\ \boldsymbol{\Sigma}_G}{\text{maximize}} && H(\mathbf{y}) - \sum_{i=1}^{N_g} H_i(\mathbf{y}_{\mathcal{S}_i}) \\ &\text{subject to} && \mathbf{y}_{\mathcal{S}_i} \cap \mathbf{y}_{\mathcal{S}_j} = \emptyset \quad (i \neq j) \quad \text{and} \quad \cup_{i=1}^{N_g}\mathbf{y}_{\mathcal{S}_i} = \mathbf{y}. \end{aligned} \tag{12}$$

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. [9]). This constraint is valid only if inputs are also partitioned. In other words,

$$\mathcal{S}_i = (\mathbf{x}_i, \mathbf{y}_i)$$

is a partition of $(\mathbf{x}, \mathbf{y})$ where $\cup_{i=1}^{N_g}(\mathcal{S}_i) = (\mathbf{x}, \mathbf{y})$, and $\mathcal{S}_i \cap \mathcal{S}_j = \emptyset$ for $i \neq j$. To take advantage of this, consider the following problem:

$$\begin{aligned} &\underset{\mathbf{V}_G, \mathbf{\Sigma}_G}{\text{maximize}} && H(\mathbf{y}) \\ &\text{subject to} && H(\mathbf{y}) = \sum_{i=1}^{N_g} H_i(\mathbf{y}_{\mathcal{S}_i}) \\ &\text{and} && tr(\mathbf{\Sigma}_G^2) = N_t. \end{aligned} \tag{13}$$

This way, the previous result can be easily utilized to exploit inter-group independence. This is the generalized PGP problem (G-PGP). A simpler version is obtained if we further specialize the power constraint in (13) to $tr(\mathbf{\Sigma}_{G_i}^2) = N_{t_i}$, 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) ($i = 1, 2, \cdots, N_g$):

$$\begin{aligned} &\underset{\mathbf{W}_{s_i}}{\text{maximize}} && I(\mathbf{x}_{s_i}, \mathbf{y}_{s_i}) \\ &\text{subject to} && \mathbf{W}_{s_i}^h = \mathbf{W}_{s_i} \\ &\text{and} && tr(\mathbf{\Sigma}_{G_i}^2) = N_{t_i}. \end{aligned} \tag{14}$$

Theorem 2.6: The PGP solution is in the feasible region of the original problem (4).

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

$$\mathbf{V} = diag[\mathbf{V}_1 \cdots \mathbf{V}_{N_g}] =$$

$$= \begin{bmatrix} \mathbf{V}_{P_1} & 0 & & & & \cdots & 0 \\ 0 & \mathbf{V}_{P_2} & 0 & \cdots & & & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & & & & \cdots & 0 & \mathbf{V}_{P_{N_g}} \end{bmatrix}$$

is a unitary matrix, and the diagonal matrix

$$\mathbf{\Sigma}_G = diag[\mathbf{\Sigma}_1 \cdots \mathbf{\Sigma}_{N_g}]$$

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

$\boldsymbol{Technique}$	$\boldsymbol{I}$ $(b/s/Hz)$
PGP	5.56
Optimal	5.73

TABLE I
PGP PERFORMANCE VS. OPTIMAL PERFORMANCE.

C. *Complexity of PGP at the Transmitter and the Receiver*

The original linear precoder optimization problem as described by (4) is solved in [5]. For an $N_t \times N_r$ MIMO system, this scheme requires evaluation of the derivatives of $N_t(N_t - 1)/2$ Givens matrices with respect to the two parameters ω_{pq}, ν_{pq}, a total of $2N_t(N_t - 1)$ differentiations at each iteration of the optimization required backtracking line search for $\mathbf{W}$. With PGP, the same calculation will entail $N_g \frac{N_t - 1}{N_t - N_g}$ differentiations instead, an important reduction in the transmitter processing complexity in this particular evaluation. Similar gains are observed in all aspects of the transmitter-based determination of the optimal precoder.

At the receiver, due to MIMO, the decoding complexity of the optimal precoder grows exponentially with N_t and is proportional to $N_r N_t M^{N_t} \log_2(M)$, due to the high complexity involved in the calculation of the extrinsic coded bit Log Likelihood Ratio (LLR) of the MIMO Detector required by the EXIT chart [3]. It can be easily shown, but it is beyond the scope of this paper, that because of the inherent inter-group independence of PGP, the corresponding complexity becomes proportional to $\frac{N_r N_t}{N_g} M^{N_t/N_g} \log_2(M)$. This represents a very significant reduction in complexity, allowing for the possibility of employing high numbers of transmitting antennas, even when a remote receiver (i.e. with low complexity) is employed.

III. NUMERICAL RESULTS

The results presented herein employ PGP as described above. The implementation of the PGP methodology is performed by employing two backtracking line searches, one for $\mathbf{W}$, and another one for $\mathbf{\Sigma}_G^2$, in a fashion similar to [5]. For most cases presented, it is worth mentioning that only a few iterations are required to converge to the PGP solution results as presented herein, even for large MIMO configurations, e.g., 6×6 MIMO systems. All the results consider Quadrature Phase Shift Keying (QPSK) modulation on narrowband independently Fading channels. For

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

at symbol signal-to-noise ratio, $SNR = 6.54\ dB$ with QPSK modulation (corresponding to average SNR of $8\ dB$, as per published results in [5]). Consider PGP where there is a 2×2 subgroup using the maximum and minimum channel singular values, and a 1×1 group using the third channel singular value. Table I shows a comparison of PGP with the optimal precoder.

Combination no.	I $b/s/Hz$
1	6.05
2	4.34
3	5.65

TABLE II
PGP PERFORMANCE IN THREE DIFFERENT GROUP SELECTION SCENARIOS.

For

$$\mathbf{H} = \begin{bmatrix} -0.7558+0.7731i & 0.2299-0.8585i & -0.0723-0.5442i & -0.6116-0.7701i \\ -0.5724+0.7844i & -0.5338-0.7874i & -0.1707+0.2626i & -0.0212+0.0230i \\ -2.0819-0.6107i & 0.9689-0.0048i & 0.2257-0.1595i & -0.1166+0.3907i \\ 1.0171+0.0547i & -1.2102+1.0837i & 0.2212+0.7901i & 0.4439+0.7782i \end{bmatrix}$$

and with $N_g = 2$, 2×2 subgroups, for a MIMO system with $N_t = N_r = 4$, 3 line backtracking search iterations in steepest descent optimization, and QPSK modulation, we get the results shown in Fig. 1.

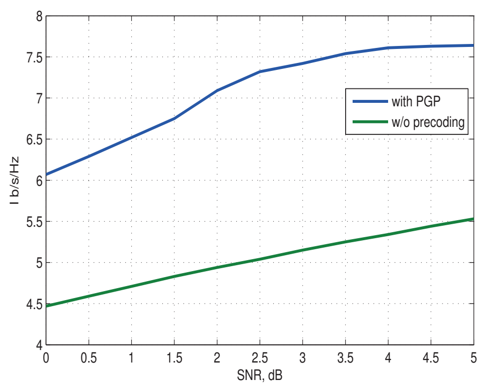

Fig. 1. Results for PGP vs. no precoding for a 4×4 MIMO system and QPSK modulation.

Consider now the performance of the previous system 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 II 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.

IV. CONCLUSIONS

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. This approach sees the overall precoding problem in an LDPC coded system from a brand new angle allowing for practical deployment of higher dimension MIMO systems with very good performance over a wide SNR range.

We show that a Pareto surface optimization leads to a very intractable problem. 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. We call the new precoding solution PGP and we show, based on simulation results, that PGP offers indeed excellent performance, while its computational complexity is significantly reduced compared to the original solution. 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.

APPENDIX
PROOF OF LEMMA 2.1

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

$$\mathbf{y}_i = \mathbf{H}_i \mathbf{G} \mathbf{x} + \mathbf{n}_i. \tag{15}$$

This represents an $N_t \times 1$ MIMO system. Thus, based on the general results presented in [5], 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

$$\nabla_{\mathbf{W}_i} H_i(\mathbf{y}_i) = \frac{1}{\sigma_n^2} \mathbf{\Phi}_{\mathbf{x}\mathbf{x}^h, \mathcal{S}=\{i\}}. \tag{16}$$

Thus, 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

$$dvec(\mathbf{W}) = \sum_i dvec(\mathbf{W}_i),$$

where $dvec(\mathbf{F})$ represents the vector differential of (complex) matrix function $\mathbf{F}$. 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 composite function differentiation results of [8]. 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 [5] 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 2.1.

APPENDIX
PROOF OF LEMMA 2.2

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 2.1, $H_I(\mathbf{y}) = \sum_{i=1}^{N_r} H_i(\mathbf{y}_i)$ 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 2.1 proves the assertion.

APPENDIX
PROOF OF THEOREM 2.3

The proof of this Theorem is as follows: First, write the Lagrangian of the 'original' problem in (6), as

$$H(\mathbf{y}) - \lambda \sum_{i=1}^{N_r} H_i(\mathbf{y}_i) - \nu(tr(\mathbf{G}\mathbf{G}^h) - N_t), \qquad (17)$$

where without loss we employed the constraint $tr(\mathbf{G}\mathbf{G}^h) \leq N_t$, 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 Lemmas of this paper, and results from [5] we can easily get the desired critical point equation described in Theorem 2.3.

APPENDIX
PROOF OF THEOREM 2.4

For $\lambda = 1$, the "equivalent" optimization problem described by (7) becomes

$$\begin{aligned} &\underset{\mathbf{W}}{\text{maximize}} && H(\mathbf{y}) - \sum_{i=1}^{N_r} H_i(\mathbf{y}_i) \\ &\text{subject to} && \mathbf{W}^h = \mathbf{W} \\ &\text{and} && tr(\mathbf{\Sigma}_G^2) = N_t \\ &\text{and} && 0 \leq \lambda < 1. \end{aligned} \qquad (18)$$

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

$$\max_{\mathbf{\Sigma}_G}\{\max_{\mathbf{V}_G}\{H(\mathbf{y}) - \sum_{i=1}^{N_r} H_i(\mathbf{y}_i)\}\} = \\ = 0,$$

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 (7) completes the proof.

APPENDIX
PROOF OF THEOREM 2.5

The proof of the Theorem is based on contradiction of Schur-type complements conditions [7] and Kronecker product properties (e.g., [8]). We also exploit properties of MMSE matrices without proof. Let us denote by $\otimes$ the Kronecker matrix product. The upper left part of the composite Hessian matrix of $H(\mathbf{y}) - \sum_{i=1}^{N_r} H_i(\mathbf{y}_i)$ ($\lambda = 1$) with respect to $\mathbf{W}$, denoted as $\mathcal{H}_{\mathbf{W},\mathbf{W}^*}$, is given as follows, employing results from [5] and the previous Lemmas and Theorems of this paper:

$$\mathcal{H}_{\mathbf{W},\mathbf{W}^*}\left(H(\mathbf{y}) - \sum_{i=1}^{N_r} H_i(\mathbf{y}_i)\right) =$$

$$= -\frac{1}{\sigma_n^2}\mathbb{E}\left(\mathbf{\Phi}^*_{\mathbf{x}\mathbf{x}^h}(\mathbf{y}) \otimes \mathbf{\Phi}_{\mathbf{x}\mathbf{x}^h}(\mathbf{y}) - \sum_{i=1}^{N_r} \mathbf{\Phi}^*_{\mathbf{x}\mathbf{x}^h,\mathcal{S}=\{i\}}(\mathbf{y}_i) \otimes \mathbf{\Phi}_{\mathbf{x}\mathbf{x}^h,\mathcal{S}=\{i\}}(\mathbf{y}_i)\right).$$

Now notice that

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

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

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

is not positive semidefinite. Based on properties of MMSE matrices we know that

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

is negative semidefinite. Then, using properties of Kronecker products we can easily show that

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

is negative semidefinite. Thus, our assertion is proven.

APPENDIX
PROOF OF THEOREM 2.6

It suffices to show that

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

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

$$\mathbf{V}\mathbf{V}^h = diag[\mathbf{V}_1\mathbf{V}_1^h \cdots \mathbf{V}_{N_g}\mathbf{V}_{N_g}^h] = \mathbf{I}.$$

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

REFERENCES

[1] E. Akay and E. Ayanoglu, "Bit Interleaved Coded Multiple Beamforming," *IEEE Transactions on Communications*, vol. 55, pp. 1802–1810, September 2007.

[2] B. Li, H. J. Park, and E. Ayanoglu, "Constellation Precoded Multiple Beamforming," *IEEE Transactions on Communications*, vol. 59, pp. 1275–1286, May 2011.

[3] S. ten Brink, G. Kramer, and A. Ashihkmin, "Design of Low-Density Parity-Check Codes for Modulation and Detection," *IEEE Transactions on Communications*, vol. 52, pp. 670–678, April 2004.

[4] Y. Jian, A. Ashihkmin, and N. Sharma, "LDPC Codes for Flat Rayleigh Fading Channels with Channel Side Information," *IEEE Transactions on Communications*, vol. 56, pp. 1207–1213, August 2008.

[5] G. Xiao, Y. Zheng, and Z. Ding, "Globally Optimal Linear Precoders for Finite Alphabet Signals Over Complex Vector Gaussian Channels," *IEEE Transactions on Signal Processing*, vol. 59, pp. 3301–3314, July 2011.

[6] F. Perez-Cruz, M. Rodriguez, and S. Verdu, "MIMO Gaussian Channels with Arbitrary Inputs: Optimal Precoding and Power Allocation," *IEEE Transactions on Information Theory*, vol. 56, pp. 1070–1086, March 2010.

[7] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[8] A. Hj∅rugnes, *Complex-Valued Matrix Derivatives With Applications in Signal Processing and Communications*. Cambridge, UK: Cambridge University Press, 2011.

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