Linear Precoding for Large MIMO Configurations and QAM Constellations

***Abstract*—In this paper, the problem of designing a linear precoder for Multiple-Input Multiple-Output (MIMO) systems in conjunction with Quadrature Amplitude Modulation (QAM) is addressed. First, a novel and efficient methodology to evaluate the input-output mutual information for a general Multiple-Input Multiple-Output (MIMO) system as well as its corresponding gradients is presented, based on the Gauss-Hermite quadrature rule. Then, the method is exploited in a block coordinate gradient ascent optimization process to determine the globally optimal linear precoder with respect to the MIMO input-output mutual information for QAM systems with relatively moderate MIMO channel sizes. The proposed methodology is next applied in conjunction with the complexity-reducing per-group processing (PGP) technique, which is semi-optimal, to both perfect channel state information at the transmitter (CSIT) as well as statistical channel state information (SCSI) scenarios, with high transmitting and receiving antenna size, and for constellation size up to $M = 64$. We show by numerical results that the precoders developed offer significantly better performance than the configuration with no precoder, and the maximum diversity precoder for QAM with constellation sizes $M = 16,\ 32$, and 64 and for MIMO channel size 100×100.**

I. INTRODUCTION

The concept of Multiple-Input Multiple-Output (MIMO) systems 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 linear precoding through exploitation of a unitary precoding matrix with success, 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-output mutual information [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 perfect channel state information available at the transmitter (CSIT)[1] scenarios with 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). Furthermore, more recent work has shown that when only statistical channel state information (SCSI)[2] is available at the transmitter, in asymptotic conditions when the number of transmitting and receiving antennas grows large, but with a constant transmitting to receiving antenna number ratio, one can design the optimal precoder by looking at an equivalent constant channel and its corresponding adjustments as per the pertinent theory [16], and applying a modified expression for the corresponding ergodic mutual information evaluation over all channel realizations. This development allows for a precoder optimization under SCSI in a much easier way [16]. However, existing research in the area does not provide any results of optimal linear precoders in the case of QAM with constellation size $M \geq 16$, with the exception of [17]. In past research work, a major impediment toward developing optimal precoders for QAM has been a lack of an accurate and efficient technique toward input-output mutual information evaluation, its gradients, and evaluation of the input-output minimum mean square error (MMSE) error covariance matrix, as required by the precoder optimization algorithm and other algorithms involved, e.g., the equivalent channel determination in the SCSI case [16].

In this paper, we propose optimal linear precoding techniques for MIMO with LDPC coding, suitable for QAM with constellation size $M \geq 16$. Carrying out this caculation has been very difficult to do until now due to the complexity involved in tackling this problem. Our approach entails a novel application of the Hermite-Gauss quadrature rule [18] which offers a very accurate and efficient way to evaluate the capacity of a MIMO system with QAM. We then apply this technique within the context of a block gradient ascent method [19] in order to determine the globally optimal linear precoder for MIMO systems, in a similar fashion to [12], for systems with CSIT and small antenna size. We show that for $M = 16,\ 32$, and 64 QAM, the optimal linear precoder offers 50% better mutual information than the maximal diversity

[1]Under CSIT the transmitter has perfect knowledge of the MIMO channel realization at each transmission.

[2]SCSI pertains to the case in which the transmitter has knowledge of only the MIMO channel correlation matrices [14], [15] and the thermal noise variance.

precoder (MDP) of [4] and the no precoder case, at low signal-to-noise ratio (SNR) for a standard 2×2 MIMO channel, however the absolute utilization gain achieved is lower than $1\ b/s/Hz$. We then proceed to show that significantly higher gains are available for different channels, e.g., a utilization gain of $1.30\ b/s/Hz$ at $SNR = 10\ dB$, when $M = 16$. We then employ higher antenna configurations, e.g., up to 40×40 with CSIT and $M = 16,\ 32$, and 64 together with the complexity reducing technique of per-group processing (PGP) which was originally presented in [20], and show very high gains available with reduced system complexity. Finally, we also employ SCSI scenarios in conjunction with PGP and show very significant gains for high antennas sizes and $M = 16,\ 32,\ 64$. Our main advantages compared with [17] lie over three main directions: a) It offers a globally optimal precoder solution for each subgroup, instead of a locally optimal one, b) It is faster, and c) It allows for higher constellation size, e.g., $M = 32,\ 64$ results with ease.

The paper is organized as follows: Section II presents the system model and problem statement. Then, in Section III, we present a novel Gauss-Hermite approximation to the evaluation of the input-output mutual information of a MIMO system that allows for fast, but otherwise very accurate evaluation of the input-output mutual information of a MIMO system, and thus represents a major facilitator toward determining the globally optimal linear precoder for LDPC MIMO. In Section IV, we present numerical results for the globally optimal precoder that implements the Gauss-Hermite approximation in the block coordinate gradient ascent method. Finally, our conclusions are presented in Section V.

II. SYSTEM MODEL AND PROBLEM STATEMENT

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, $\mathbf{G}$ 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 of size M, 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 $\sigma^2 = \frac{1}{SNR}$, SNR being the (coded) symbol signal-to-noise ratio. In this paper, a number of different channels will be considered, e.g., channels comprising independent complex Gaussian components or spatially correlated Kronecker-type channels [14] (including similar to the 3GPP spatial correlation model (SCM) ones [21] used in [17]) or [15]. The precoding matrix $\mathbf{G}$ needs to satisfy the following power constraint

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

where $\mathrm{tr}(\mathbf{A})$, $\mathbf{A}^h$ denote the trace and the Hermitian transpose of matrix $\mathbf{A}$, respectively. An equivalent model called herein the "virtual" channel is given by [12]

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

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, $\mathbf{y}$ [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 $\mathbf{W} = \mathbf{V}_G \mathbf{\Sigma}_H^2 \mathbf{\Sigma}_G^2 \mathbf{V}_G^h$. The optimal precoder, $\mathbf{G}$ is found by solving:

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

called the "original problem," and

$$\begin{aligned} \underset{\mathbf{V}_G, \mathbf{\Sigma}_G}{\text{maximize}} \quad & I(\mathbf{x};\mathbf{y}) \\ \text{subject to} \quad & \mathrm{tr}(\mathbf{\Sigma}_G^2) = N_t, \end{aligned} \tag{5}$$

called the "equivalent problem," where the reception model of (3) is employed. The solution to (4) or (5) results in exponential complexity at both transmitter and receiver, and it becomes especially difficult for QAM with constellation size $M \geq 16$. A major difficulty in the QAM case stems from the fact that there are multiple evaluations of $I(\mathbf{x};\mathbf{y})$ in the block coordinate ascent method employed for determining the globally optimal precoder. More specifically, for each block coordinate gradient ascent iteration, there are two line backtracking searches required [12], which demand one $I(\mathbf{x};\mathbf{y})$ plus its gradient evaluations per search trial, and one additional evaluation at the end of a successful search per backtracking line search. Thus, the need of a fast, but otherwise very accurate method of calculating $I(\mathbf{x};\mathbf{y})$ and its gradients prevails as instrumental toward determining the globally optimal linear precoder for LDPC MIMO.

III. ACCURATE APPROXIMATION TO $I(\mathbf{x};\mathbf{y})$ FOR MIMO SYSTEMS BASED ON GAUSS-HERMITE QUADRATURE

In Appendix A we prove that by applying the Gauss-Hermite quadrature theory for approximating the integral of a Gaussian function multiplied with an arbitrary real function $f(x)$, i.e.

$$F \doteq \int_{-\infty}^{+\infty} \exp(-x^2) f(x) dx, \tag{6}$$

which is approximated in the Gauss-Hermite approximation with L weights and nodes as

$$F \approx \sum_{l=1}^{L} c(l) f(v_l) = \mathbf{c}^t \mathbf{f}, \tag{7}$$

with $\mathbf{c} = [c(1) \cdots c(L)]^t$, $\{v_l\}_{l=1}^L$, and $\mathbf{f} = [f(v_1) \cdots f(v_L)]^t$, being the vector of the weights, the nodes, and function node values, respectively (see Appendix A), the following very accurate approximation is derived for $I(\mathbf{x};\mathbf{y})$ in a MIMO system, as presented in the following lemma. Let us first introduce some notations that make the overall understanding easier. Let $\mathbf{n}_e$ denote the equivalent to $\mathbf{n}$, real vector of length

$2N_r$ derived from $\mathbf{n}$ by separating its real, imaginary parts as follows

$$\mathbf{n}_e = [n_{r1}\ n_{i1} \cdots n_{rN_r}\ n_{iN_r}]^t, \tag{8}$$

with $n_{rv},\ n_{iv}$ being the values of the real, imaginary part of the v $(1 \leq v \leq N_r)$ element of $\mathbf{n}$, respectively. Let us also define the real vector $\mathbf{v}(\{k_{rv},\ k_{iv}\}_{v=1}^{N_r})$ of length $2N_r$ defined as follows

$$\mathbf{v}(\{k_{rv},\ k_{iv}\}_{v=1}^{N_r}) = [v_{kr1}\ v_{ki1}, \cdots, v_{krN_r}\ v_{kiN_r}]^t, \tag{9}$$

with $k_{rv},\ k_{iv}$ $(1 \leq v \leq N_r)$ being permutations of indexes in the set $\{1, 2, \cdots, L\}$. Then the following lemma is true concerning the Gauss-Hermite approximation for $I(\mathbf{x};\mathbf{y})$.

Lemma 1. *For the MIMO channel model presented in (1), the Gauss-Hermite approximation for $I(\mathbf{x};\mathbf{y})$ with L nodes per receiving antenna is given as*

$$I(\mathbf{x};\mathbf{y}) \approx N_t \log_2(M) - \frac{N_r}{\log(2)} - \frac{1}{M^{N_t}} \sum_{k=1}^{M^{N_t}} \hat{f}_k, \tag{10}$$

where

$$\hat{f}_k = \left(\frac{1}{\pi}\right)^{N_r} \sum_{k_{r1}=1}^{L} \sum_{k_{i1}=1}^{L} \cdots \sum_{k_{rN_r}=1}^{L} \sum_{k_{iN_r}=1}^{L} c(k_{r1})c(k_{i1}) \cdots c(k_{rN_r})c(k_{iN_r}) g_k(\sigma n_{k_{r1}}, \sigma n_{k_{i1}}, \cdots, \sigma n_{k_{rN_r}}, \sigma n_{k_{iN_r}}), \tag{11}$$

with

$$g_k(\sigma v_{k_{r1}}, \sigma v_{k_{i1}}, \cdots, \sigma v_{k_{rN_r}}, \sigma v_{k_{iN_r}}) \tag{12}$$

being the value of the function

$$\log_2\left(\sum_m \exp(-\frac{1}{\sigma^2}||\mathbf{n} - \mathbf{HG}(\mathbf{x}_k - \mathbf{x}_m)||^2)\right) \tag{13}$$

evaluated at $\mathbf{n}_e = \sigma \mathbf{v}(\{k_{rv},\ k_{iv}\}_{v=1}^{N_r})$.

The proof of this lemma is presented in Appendix A.

IV. GLOBAL OPTIMIZATION OVER $\mathbf{G}$ TOWARD MAXIMUM $I(\mathbf{x};\mathbf{y})$ FOR QAM

A. Description of the Globally Optimal Precoder Method

Similarly to [12], we follow a block coordinate gradient ascent maximization method to find the solution to the optimization problem described in (4), 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 $\mathbf{\Sigma}_G^2$. It thus becomes efficient to employ two different gradient ascent methods, one for $\mathbf{W}$, and another one for $\mathbf{\Sigma}_G^2$. We employ $\mathbf{\Theta}$ and $\mathbf{\Sigma}$ to denote $\mathbf{V}_G^h$ and $\mathbf{\Sigma}_G^2$, respectively, evaluated during the optimization algorithm's execution.

There will in general be multiple evaluations of $I(\cdot)$, until the searches satisfy the conditions set or the maximum number of attempts allowed in a search has been reached. This explains the importance behind the requirement for an algorithm capable of efficient calculation of $I(\mathbf{x};\mathbf{y})$. In addition, as the parameters $\alpha_1,\ \alpha_2,\ \beta_1,\ \beta_2$ need to be optimized for faster and more efficient execution of the globally optimal precoder optimization, this requirement becomes even more essential. Finally, the role of n_1, n_2 is also very important as when the number of attempts within each loop grows, the corresponding differential value of the parameter decreases and after a few attempts, the corresponding value of the step size is almost zero. By employing the proposed approach herein the possibility of finding the globally optimal precoder for QAM with $M \geq 16$ becomes reality, as our results demonstrate.

B. Determination of $\nabla_{\mathbf{W}} I,\ \nabla_{\mathbf{\Sigma}_G^2} I$

We first set $\mathbf{M} = \mathbf{W}^{\frac{1}{2}}$. Then, it is easy to see that I is a function of $\mathbf{M}$ (see, e.g., [22] where the notion of sufficient statistic is employed to show that $I(\mathbf{x};\mathbf{y})$ depends on $\mathbf{W}$). The derivation of $\nabla_{\mathbf{W}} I$ is presented in Appendix B. The proof is based on the following theorem[3].

Theorem 1. *Substituting $\mathbf{M} = \mathbf{V}_G \mathbf{\Sigma}_H \mathbf{\Sigma}_G \mathbf{V}_G^h = \mathbf{W}^{\frac{1}{2}}$ for $\mathbf{HG}$ in (10) results in the same value of $I(\mathbf{x};\mathbf{y})$. In other words, since $\mathbf{M}$ is a function of $\mathbf{H}$, $\mathbf{G}$, $\mathbf{My}$ is a sufficient statistic for $\mathbf{y}$.*

Proof. The proof of the theorem is simple. First, recall that the "virtual" channel model in (3) is equivalent to the following model, which results by multiplying (3) by the unitary matrix $\mathbf{V}_G$ on the left, resulting in

$$\tilde{\mathbf{y}} = \mathbf{V}_G \mathbf{y} = \mathbf{V}_G \mathbf{\Sigma}_H \mathbf{\Sigma}_G \mathbf{V}_G^h \mathbf{x} + \mathbf{V}_G \mathbf{n}, \tag{14}$$

where the modified noise term $\mathbf{V}_G \mathbf{n}$ has the same statistics with $\mathbf{n}$, because $\mathbf{V}_G$ is unitary. By applying the Gauss-Hermite approximation to (14), we see that we get the desired result, i.e., the value of I remains the same, since both channel manifestations represent equivalent channels, i.e., the original one and its equivalent, thus their mutual information is the same. This completes the proof of the theorem. □

Note that using this simple theorem, we can prove easily part of Theorem 1 in [12], namely the fact that $I(\mathbf{x};\mathbf{y})$ is only a function of $\mathbf{W}$, as $\mathbf{My}$ a sufficient statistic for $\mathbf{y}$ and $\mathbf{M}$ is a function of $\mathbf{W}$.

Assume without loss of generality that $N_t = N_r$. The gradient of I with respect to $\mathbf{M}$ can be found (see Appendix B for the derivation) from the Gauss-Hermite expression presented in (10) as follows

$$\nabla_{\mathbf{M}} I = -\frac{1}{\log(2)} \frac{1}{M^{N_t}} \left(\frac{1}{\pi}\right)^{N_r} \sum_{k_{r1}=1}^{L} \sum_{k_{i1}=1}^{L} \cdots \sum_{k_{rN_r}=1}^{L} \sum_{k_{iN_r}=1}^{L} c(k_{r1})c(k_{i1}) \cdots c(k_{rN_r})c(k_{iN_r}) \times \mathbf{R}(\sigma v_{k_{r1}}, \sigma v_{k_{i1}}, \cdots, \sigma v_{k_{rN_r}}, \sigma v_{k_{iN_r}}), \tag{15}$$

[3]The theorem applies without loss of generality to the $N_t = N_r$ case. If $N_t \neq N_r$, then $\mathbf{\Sigma}_H$ needs to be either shrunk, or extended in size, by elimination or addition of zeros, respectively.

where $\mathbf{R}(\sigma v_{k_{r1}}, \sigma v_{k_{i1}}, \cdots, \sigma v_{k_{rN_r}}, \sigma v_{k_{iN_r}})$ is the value of the $N_t \times N_t$ matrix

$$\sum_k \frac{1}{\sum_m \exp(-\frac{1}{\sigma^2}||\mathbf{n} - \mathbf{M}(\mathbf{x}_k - \mathbf{x}_m)||^2)} \times \sum_m \exp(-\frac{1}{\sigma^2}||\mathbf{n} - \mathbf{M}(\mathbf{x}_k - \mathbf{x}_m)||^2) \times ((\mathbf{n} - \mathbf{M}(\mathbf{x}_k - \mathbf{x}_m))(\mathbf{x}_k - \mathbf{x}_m)^h + ((\mathbf{n} - \mathbf{M}(\mathbf{x}_k - \mathbf{x}_m))(\mathbf{x}_k - \mathbf{x}_m)^h)^h) \quad (16)$$

evaluated at $\mathbf{n}_e = \sigma \mathbf{v}(\{k_{rv}, \ k_{iv}\}_{v=1}^{N_r})$.

The required $\nabla_{\mathbf{W}} I$ for the execution of the optimization process can be found from Appendix B as per the next lemma, using an easily proven equation. Using the fact that for a Hermitian matrix such as $\mathbf{M}$, we need to add the Hermitian of the differential above in order to evaluate the actual gradient (see [23]), we get the desired result as follows (see Appendix B).

Lemma 2. *For the MIMO channel model presented in (1), the Gauss-Hermite approximation allows to approximate $\nabla_{\mathbf{W}} I$ as follows.*

$$\nabla_{\mathbf{W}} I \approx reshape((vec(\nabla_{\mathbf{M}} I)^T ((\mathbf{M}^*) \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{M})^{-1}), N_t, N_t), \quad (17)$$

where $reshape(\mathbf{A}, k, n)$ is the reshape of matrix $\mathbf{A}$ (with total number of elements kn) to a matrix with k rows, n columns and where $\otimes$ denotes Kronecker product of matrices.

Then, as I is a concave function of $\mathbf{W}$ [22], we can maximize over $\mathbf{W}$ in a straightforward way using closed form expressions. This is based on the fact that the approximated I through the Gauss-Hermite approximation is very accurate, as shown in the next section.

Finally, since from [24] we have that $\nabla_{\mathbf{\Sigma}_G^2} I = \text{diag}(\mathbf{V}_G^h \nabla_{\mathbf{W}} I \mathbf{V}_G \mathbf{\Sigma}_H^2)$, we can easily evaluate it through the above presented procedure.

V. NUMERICAL RESULTS

The results presented herein employ QAM with 16, 32 or 64 constellation sizes. We employ MIMO systems with $N_t = N_r = 2$ when global precoding optimization is performed. We have used an $L = 3$ Gauss-Hermite approximation which results in 3^{2N_r} total nodes due to MIMO. We apply the complexity reducing method of PGP [20] which offers semi-optimal results under exponentially lower transmitter and receiver complexity [20]. PGP divides the transmitting and receiving antennas into independent groups, thus achieving a much simpler detector structure while the precoder search is also dramatically reduced as well. Finally, we address both the CSIT and SCSI cases, as they both present significant application scenarios for 5G.

A. Results for SCSI in conjunction with PGP

In Fig. 7 we present results for PGP versus a no precoding SCM channel with $N_t = N_r = 100$ and $M = 16$. To the best of our knowledge, results for such high size antenna configurations were not available in the literature. Similar to the previous results, we observe high information rate gains in the high SNR_b regime as the PGP system achieves the full capacity of 400 $b/s/Hz$ while the no precoding scheme saturates at 320 $b/s/Hz$. The PGP system employed uses 50 groups of size 2×2 each.

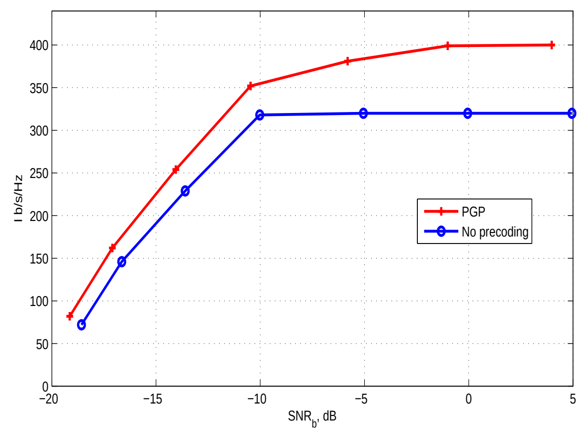

Fig. 7. Results for PGP and no precoding cases for a 100×100 $\mathbf{H}$ SCSI MIMO system and QAM $M = 16$ modulation.

B. Results for CSIT in conjunction with PGP

For a $4X4$ channel

$$\mathbf{H} = \begin{bmatrix} -2.2536 & 1.4748 & -0.2612 & 0.0179 \\ 0.2824 - 0.2468i & -0.6570 - 0.5204i & -0.6028 - 0.2290i & 0.0201 + 0.0355i \\ 0.7920 - 1.7400i & 0.4156 - 0.3760i & -0.1992 + 0.3091i & -0.0685 + 0.0110i \\ 1.3780 + 2.2864i & 0.5010 + 0.7468i & -0.0585 + 0.1318i & -0.0081 + 0.0560i \end{bmatrix},$$

used with $M = 64$ we get the no precoding and the PGP results using 2 groups of size 2×2 each depicted in Fig. 8. This example represents the corresponding CSIT case example that is similar to the SCM channels used in the previous subsection. We observe very high gains of PGP over the no precoding case in the high SNR_b regime. To the best of our knowledge this type of results for optimal precoding in conjunction with

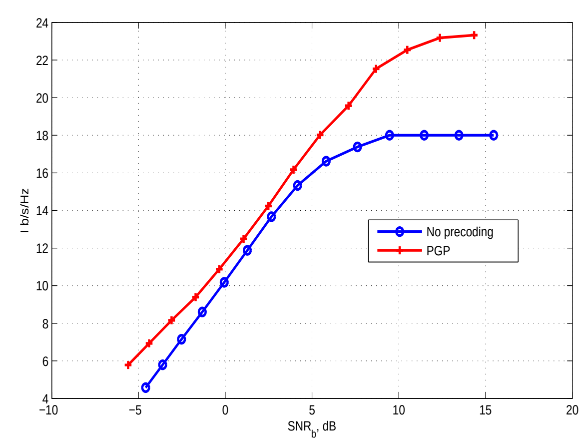

Fig. 8. Results for PGP and no precoding cases for a 4×4 $\mathbf{H}$ CSIT MIMO system and QAM $M = 64$ modulation.

$M = 64$ were not available in the literature. Finally, in Fig. 9 we present results for an asymmetric randomly generated MIMO channel with $N_t = 4,\ N_r = 10$, and $M = 16$. PGP employs 2 groups of size $N_r = 5,\ N_t = 2$ each. In the current scenario, we observe that significant gains are shown in the low SNR_b regime, e.g., around 3 dB in SNR_b lower than $-7\ dB$.

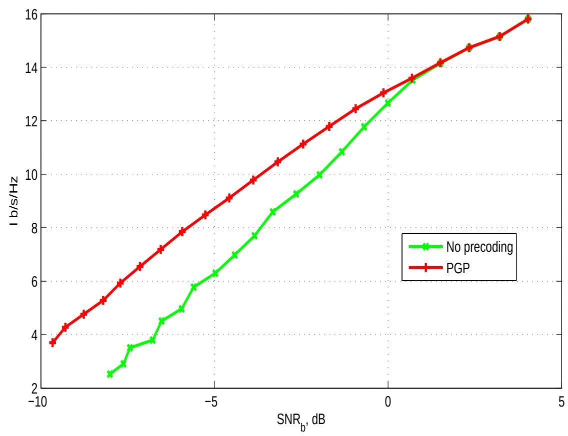

Fig. 9. Results for PGP and no precoding cases for a randomly generated 10×4 $\mathbf{H}$ CSIT MIMO system and QAM $M = 16$ modulation.

C. Results for Massive MIMO

Massive MIMO [25]–[27] has attracted much interest recently, due to its potential to offer high data rates with low signal power. We present results for the uplink, and downlink of a Massive MIMO system based on 100 base station, 4 user antennas, respectively, with $M = 16,\ 64$, and for a Kronecker-based 3GPP SCM urban channel in a CSIT scenario. Fig. 10 shows results for the 4×100 uplink of the system. We employ PGP to dramatically reduce the system complexity at the transmitter and receiver sites. Under no precoding, the channel saturates and fails to meet the maximum possible mutual information of 16 $b/s/Hz$, while with PGP the system clearly achieves the maximum mutual information rate, thus achieving high gains on the uplink in the high SNR regime. We stress the much higher throughput possible with $M = 64$ over the $M = 16$ case. For example, the no precoding $M = 16$ uplink significantly outperforms the PGP $M = 16$ uplink. Second, the PGP $M = 6$ uplink offers further gains by, e.g., achieving the maximum possible rate of 24 $b/s/Hz$. For the downlink, in Fig. 11 we show results where the no precoding case operates under 100 antenna inputs all correlated through the right eigenvectors of the channel, thus creating a very demanding environment at the user, due to the exponentially increasing maximum a posteriori (MAP) detector complexity [20]. On the other hand, employing PGP with only two input symbols per receiving antenna, i.e., with dramatically reduced decoding complexity, the PGP system achieves much higher throughput in the lower SNR regime, with SNR gain on the order of 10 dB, albeit achieving a maximum of $32,\ 48$ $b/s/Hz$ as there are a total of 8 $M = 16,\ 64$ QAM data symbols employed, respectively. We observe the superior performance of $M = 64$ over its $M = 16$ counterpart due to its increased constellation size. For example, at medium SNR_b, e.g., $SNR_b = 4$, the $M = 64$ PGP scheme achieves 45% higher throughput that the $M = 16$ one, a significant improvement. We also like to emphasize that the no precoding scheme requires a very high exponential MAP detector complexity, on the order of M^{100}, while for the low-SNR-superior PGP, this complexity is on the order of M^2 only. Thus, even in the higher SNR region where the no precoding scheme can achieve a higher throughput, the complexity required at the user site becomes prohibitive. This demonstrates the superiority of PGP on the Massive MIMO downlink. On the other hand, in lower SNR, the PGP scheme achieves both much higher throughput with simultaneously exponentially lower MAP detector complexity at the user site detector.

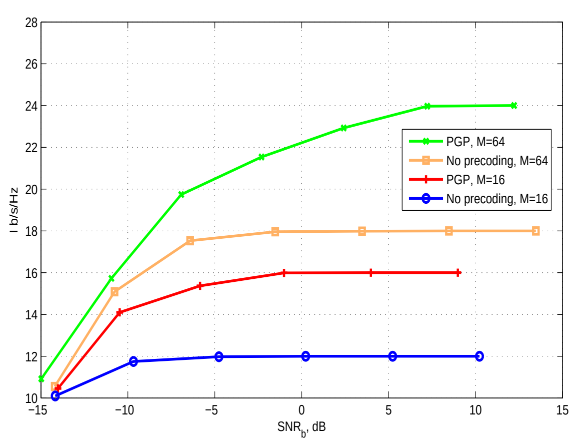

Fig. 10. Results for PGP and no precoding cases for a randomly generated 100×4 uplink $\mathbf{H}$ CSIT MIMO system and QAM $M = 16$ modulation.

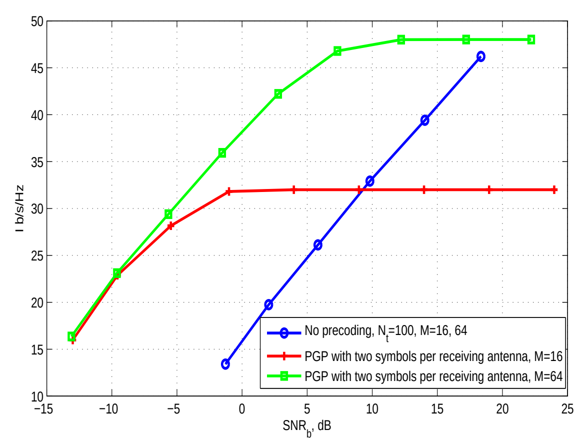

Fig. 11. Results for PGP and no precoding cases for a randomly generated 4×100 $\mathbf{H}$ downlink CSIT MIMO system and QAM $M = 16$ modulation.

VI. CONCLUSIONS

In this paper, the problem of designing a linear precoder for MIMO systems employing, e.g., LDPC codes toward mutual information maximization is addressed for QAM with $M \geq 16$

and in conjunction with high MIMO system size. A major obstacle toward this goal is a lack of efficient techniques for evaluating $I(\mathbf{x}, \mathbf{y})$ and its derivatives. We have presented a novel solution to this problem based on the Gauss-Hermite quadrature. $I(\mathbf{x}; \mathbf{y}) = N_t \log_2(M)$ at high SNR. Furthermore, we apply the same Gauss-Hermite approximation approach to CSIT high antenna size channels with the same success. We show that for specific type of channels similar to urban 3GPP SCM [21], the PGP approach offers very high gains over the no precoding case in the high SNR_b regime. Finally, we consider a Massive MIMO scenario in conjunction with CSIT and show that by carefully designing the downlink, uplink precoders, the methodology shows very high gains, especially on the downlink, although it employs an expenentially simpler MAP detector at the user site.

Based on the evidence presented, the novel application of the Gauss-Hermite quadrature rule in the MIMO scenario allows for generalizing the interesting results presented in [12], [16] to the QAM case with ease. Because of the simplification achieved by the combination of PGP and the Gauss-Hermite approximation, we were able to derive results with, e.g., $N_t = N_r = 100$ as well as with $M = 64$ efficiently. The novel combination of the Gauss-Hermite approximation with PGP allows for application of optimized procoding to 5G oriented massive MIMO systems with ease. In addition, the presented Gauss-Hermite approximation offers important simplification in the evaluation of the MMSE error covariance matrix of the MIMO channel which is required in, among other areas, the SCSI equivalent channel determination as per [16]. Finally, compared with other interesting proposals for large MIMO sizes, e.g., [17], which also employs PGP, the presented approach has the following advantages: a) It offers a globally optimal precoder solution for each subgroup, instead of a locally optimal one, b) It is faster, and c) It allows for higher constellation size, e.g., $M = 32,\ 64$ results with ease, thus it seems to be more efficient in these regards.

Appendix A
Gauss-Hermite Quadrature Approximation in MIMO Input Output Mutual Information

$I(\mathbf{x}; \mathbf{y}) = H(\mathbf{x}) - H(\mathbf{x}|\mathbf{y}) = N_t \log_2(M) - H(\mathbf{x}|\mathbf{y})$, where the conditional entropy, $H(\mathbf{x}|\mathbf{y})$ can be written as [12]

$$\begin{aligned} &H(\mathbf{x}|\mathbf{y}) = \frac{N_r}{\log(2)} + \frac{1}{M^{N_t}} \sum_k \\ &\mathbb{E}_{\mathbf{n}} \left(\log_2 \left(\sum_m \exp(-\frac{1}{\sigma^2} ||\mathbf{n} - \mathbf{HG}(\mathbf{x}_k - \mathbf{x}_m)||^2) \right) \right) = \\ &\frac{N_r}{\log(2)} + \frac{1}{M^{N_t}} \sum_k \int_{-\infty}^{+\infty} \mathcal{N}_c(\mathbf{n}|\mathbf{0}, \sigma^2 \mathbf{I}) \\ &\times \log_2 \left(\sum_m \exp(-\frac{1}{\sigma^2} ||\mathbf{n} - \mathbf{HG}(\mathbf{x}_k - \mathbf{x}_m)||^2) \right) d\mathbf{n}, \end{aligned} \tag{18}$$

where $\mathcal{N}_c(\mathbf{n}|\mathbf{0}, \sigma^2 \mathbf{I})$ represents the probability density function (pdf) of the circularly symmetric complex AWGN. Let us define

$$\begin{aligned} f_k \doteq & \int_{-\infty}^{+\infty} \mathcal{N}_c(\mathbf{n}|\mathbf{0}, \sigma^2 \mathbf{I}) \times \\ & \log_2 \left(\sum_{\mathbf{x}_m} \exp(-\frac{1}{\sigma^2} ||\mathbf{n} - \mathbf{HG}(\mathbf{x}_k - \mathbf{x}_m)||^2) \right) d\mathbf{n}. \end{aligned} \tag{19}$$

Since $\mathbf{n}$ has independent components over the different receiving antennas, and over the real and imaginary dimensions, the integral above can be partitioned into $2N_r$ real integrals in tandem, in the following manner: Define by n_{rv}, n_{iv}, with $v = 1, \cdots, N_r$, the vth receiving antenna real and imaginary noise component, respectively. Also define by $(\mathbf{HG}(\mathbf{x}_k - \mathbf{x}_m))_{rv}$ and $(\mathbf{HG}(\mathbf{x}_k - \mathbf{x}_m))_{iv}$, the vth receiving antenna real and imaginary component of $(\mathbf{HG}(\mathbf{x}_k - \mathbf{x}_m))$, respectively. We then have

$$\mathcal{N}_c(\mathbf{n}|\mathbf{0}, \sigma^2 \mathbf{I}) = \frac{1}{\pi^{N_r} \sigma^{2N_r}} \exp(-\frac{\sum_l n_{rv}^2 + n_{iv}^2}{\sigma^2}), \tag{20}$$

$$d\mathbf{n} = \prod_{v=1}^{N_r} dn_{rv} dn_{iv}, \tag{21}$$

and

$$\begin{aligned} &\sum_m \exp(-\frac{1}{\sigma^2} ||\mathbf{n} - \mathbf{HG}(\mathbf{x}_k - \mathbf{x}_m)||^2) \\ &= \sum_m \exp(-\frac{1}{\sigma^2} (\sum_v (n_{rv} - (\mathbf{HG}(\mathbf{x}_k - \mathbf{x}_m))_{rv})^2 \\ &+ \sum_v (n_{iv} - (\mathbf{HG}(\mathbf{x}_k - \mathbf{x}_m))_{iv})^2)). \end{aligned} \tag{22}$$

The Gauss-Hermite quadrature is as follows:

$$\int_{-\infty}^{+\infty} \exp(-x^2) f(x) dx \approx \sum_{l=1}^{L} c(l) f(v_l), \tag{23}$$

for any real function $f(x)$, and with vector $\mathbf{c} = [c(1) \cdots c(L)]^T$ being the "weights," and v_l are the "nodes" of the approximation. The approximation is based on the following weights and nodes [18]

$$c(l) = \frac{2^{L-1} L! \sqrt{2}\pi}{L^2 (H_{L-1}(v_l))^2} \tag{24}$$

where $H_{L-1}(x) = (-1)^{L-1} \exp(x^2) \frac{d^{L-1}}{dx^{L-1}} (\exp(-x^2))$ is the $(L-1)$th order Hermitian polynomial, and the value of the node v_l equals the root of $H_L(x)$ for $l = 1, 2, \cdots, L$.

Applying the Gauss-Hermite quadrature $2N_r$ times in tandem, to the integral in (19), and after changing variables, we get that

$$\begin{aligned} f_k \approx \hat{f}_k = \left(\frac{1}{\pi}\right)^{N_r} \sum_{k_{r1}=1}^{L} \sum_{k_{i1}=1}^{L} \cdots \sum_{k_{rN_r}=1}^{L} \sum_{k_{iN_r}=1}^{L} c(k_{r1}) c(k_{i1}) \\ \cdots \times c(k_{rN_r}) c(k_{iN_r}) g_k(\sigma n_{k_{r1}}, \sigma n_{k_{i1}}, \cdots, \sigma n_{k_{rN_r}}, \sigma n_{k_{iN_r}}), \end{aligned} \tag{25}$$

where

$$g_k(\sigma n_{k_{r1}}, \sigma n_{k_{i1}}, \cdots, \sigma n_{k_{rN_r}}, \sigma n_{k_{iN_r}}) \tag{26}$$

is the value of the function (from (25))

$$\log_2\left(\sum_m \exp(-\frac{1}{\sigma^2}||\mathbf{n}-\mathbf{HG}(\mathbf{x}_k-\mathbf{x}_m)||^2)\right) \quad (27)$$

evaluated at $\mathbf{n}_e = \sigma\mathbf{v}(\{k_{rv},\ k_{iv}\}_{v=1}^{N_r})$.

Appendix B
Derivation of $\nabla_{\mathbf{W}} I$ through the Gauss-Hermite approximation

Without loss of generality, let's assume that $N_t = N_r$. Using Theorem 1, we can write by using the Gauss-Hermite approximation with $\mathbf{M}$ instead of $\mathbf{HG}$,

$$I(\mathbf{x};\mathbf{y}) \approx N_t \log_2(M) - \frac{N_r}{\log(2)} - \frac{1}{M^{N_t}}\sum_k \hat{f}_k. \quad (28)$$

In order to derive the gradient of I with respect to $\mathbf{W}$, we first derive the gradient of I with respect to $\mathbf{M}$. Start with the differential of I with respect to $\mathbf{M}^*$ in (28) and approximate the f_k by $\hat{f}_k$, to get for the differential of $I(\mathbf{x};\mathbf{y})$ over $\mathbf{M}^*$. The full details can be found in [28], but are omitted from here, due to lack of space. We can then employ identities from [23] to get the desired result.

References

[1] E. Akay, E. Sengul, 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] B. Li and E. Ayanoglu, "Diversity Analysis of Bit-Interleaved Coded Multiple Beamforming with Orthogonal Frequency Division Multiplexing," *IEEE Transactions on Communications*, vol. 61, pp. 3794–3805, September 2013.

[4] Y. Xin, Z. Wang, and G. Giannakis, "Space-Time Diversity Systems Based on Linear Constellation Precoding," *IEEE Transactions on Wireless Communications*, vol. 2, pp. 294–309, March 2003.

[5] V. Tarokh, N. Seshadri, and A. Calderbank, "Space-Time Codes for High Data Rate Communications: Performance Criterion and Code Construction," *IEEE Transactions on Information Theory*, vol. 44, pp. 744–765, March 1998.

[6] 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.

[7] 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.

[8] 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.

[9] A. Lozano, A. Tulino, and S. Verdu, "Optimal Power Allocation for Parallel Gaussian Channels With Arbitrary Input Distributions," *IEEE Transactions on Information Theory*, vol. 52, pp. 3024–3051, July 2006.

[10] M. Payaro and D. Palomar, "On Optimal Precoding in Linear Vector Gaussian Channels with Arbitrary Input Distribution," in *Proceedings IEEE International Symposium on Information Theory*, 2009.

[11] E. Ohlmer, U. Wachsmann, and G. Fettweis, "Mutual Information Maximizing Linear Precoding for Parallel Layer MIMO Detection," in *Proceedings 12th International Workshop on Signal Processing Advances in Wireless Communications*, 2011.

[12] C. 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.

[13] M. Lamarca, "Linear Precoding for Mutual Information Maximization in MIMO Systems," in *Proceedings International Symposium of Wireless Communication Systems 2009*.

[14] D. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading Correlation and Its Effect on the Capacity of Multielement Antenna Systems," *IEEE Transactions on Communications*, vol. 48, pp. 502–513, March 2000.

[15] W. Weichselberger, M. Herdin, H. Ozcelik, and E. Bonek, "A Stochastic MIMO Channel Model with Joint Correlation of Both Links," *IEEE Transactions on Wireless Communications*, vol. 5, pp. 90–100, January 2006.

[16] Y. Wu, C.-K. Wen, C. Xiao, X. Gao, and R. Schober, "Linear precoding for the MIMO Multiple Access Channel With Finite Alphabet Inputs and Statistical CSI," *IEEE Transactions on Wireless Communications*, pp. 983–997, February 2015.

[17] Y. Wu, C.-K. Wen, D. Ng, R. Schober, and A. Lozano, "Low-Complexity MIMO Precoding with Discrete Signals and Statistical CSI," manuscript submitted for publication.

[18] M. Abramowitz and I. Stegun, *Handbookof Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Washington D.C.: U.S. Government Printing Office, 1972.

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

[20] T. Ketseoglou and E. Ayanoglu, "Linear precoding for MIMO with LDPC Coding and Reduced Complexity," *IEEE Transactions on Wireless Communications*, pp. 2192–2204, April 2015.

[21] J. Salo, G. D. Galdo, J. Salmi, P. Kyosti, M. Milojevic, D. Laselva, and C. Schneider, "MATLAB Implementaion of the 3GPP Spatial Channel Model," Available at http://www.tkk.fi/Units/Radio/scm/.

[22] Y. Wu, C. Wen, C. Xiao, X. Gao, and R. Schober, "Linear Precoding for the MIMO Multiple Access Channel with Finite Alphabet Inputs and Statistical CSI," *IEEE Transactions on Signal Processing*, vol. 60, pp. 3134–3148, February 2012.

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

[24] W. Zeng, C. Xiao, and J. Lu, "A Low Complexity Design of Linear Precoding for MIMO Channels with Finite Alphabet Inputs," *IEEE Wireless Communications Letters*, vol. 1, pp. 38–42, February 2012.

[25] T. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 3590–3600, November 2010.

[26] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot Contamination and Precoding in Multi-Cell TDD Systems," *IEEE Transactions on Wireless Communications*, vol. 10, pp. 2640–2651, August 2011.

[27] H. Ngo, E. Larsson, and T. Marzetta, "Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems," *IEEE Transactions on Communications*, vol. 61, pp. 1436–1449, April 2013.

[28] T. Ketseoglou and E. Ayanoglu, "Linear Precoding for MIMO Channels with QAM Constellations and Reduced Complexity," 2016, eprint: arXiv: 1601.03141v1 [cs.IT].