Linear Precoding Gain for Large MIMO Configurations With QAM 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 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 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 technique to both perfect channel state information (CSI) at the transmitter as well as statistical CSI (Statistical CSI) scenarios, with large transmitting and receiving antenna sizes, and for constellation size up to $M = 64$. We show by the numerical results that the precoders developed offer significantly better performance than the configuration with no precoder as well as the maximum diversity precoder for QAM with constellation sizes $M = 16$, 32, and 64 and for MIMO channel size up to 100 × 100.**

***Index Terms*— Linear precoding, MIMO, mutual information, quadrature amplitude modulation.**

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, the problem of linear precoding design has been widely studied. Early work on MIMO precoding can be found in [1]–[3]. In addition, a variety of precoding concepts and under multiple scenarios can be found in [3]–[10]. Within more recent MIMO precoding research with realistic, finite alphabet inputs, the problem of designing an optimal linear precoder toward maximizing the mutual information between the input and output was considered in [11] and [12] where the optimal power allocation strategies were presented (e.g., Mercury Waterfilling (MWF)), together with general equations for the optimal precoder design. Furthermore, in [13] a channel pairing approach to MIMO precoding was presented. In addition, [14] also considered precoders for mutual information maximization and showed that the left singular vectors of the optimal precoder can be set equal to the right singular vectors of the channel. Finally, in [15], a mutual information maximizing precoder for a parallel layer MIMO detection system was set forth reducing the performance gap between maximum likelihood and parallel layer detection.

Recently, globally optimal linear precoding techniques were presented [16], [17] for scenarios employing perfect channel state information available at the transmitter (CSIT)[1] with finite alphabet inputs, capable of achieving mutual information rates much higher than the previously presented MWF [11] techniques by introducing input symbol correlation through a unitary input transformation matrix in conjunction with channel weight adjustment (power allocation). In addition, [18] presented an iterative algorithm for precoder optimization for sum rate maximization of Multiple Access Channels (MAC) with Kronecker MIMO channels. Furthermore, more recent work showed that when only Statistical CSI[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 [21], and applying a modified expression for the corresponding ergodic mutual information evaluation over all channel realizations. This development allows for a precoder optimization under Statistical CSI in a much easier way [21]. However, existing research in the area has not provided any results of optimal linear precoders in the case of QAM with constellation size $M \geq 16$, with the exception of [22]. In past research work, a major impediment toward developing optimal precoders for QAM has been a lack

Manuscript received March 6, 2016; revised July 6, 2016; accepted August 1, 2016. Date of publication August 9, 2016; date of current version October 14, 2016. The associate editor coordinating the review of this paper and approving it for publication was V. Raghavan.

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 the Center for Pervasive Communications and Computing, Department of Electrical Engineering and Computer Science, University of California at Irvine, Irvine, CA 92697 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/TCOMM.2016.2598836

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

[2]Statistical CSI pertains to the case in which the transmitter has knowledge of only the MIMO channel correlation matrices [19], [20] and the thermal noise variance.

of an accurate and efficient technique toward input-output mutual information evaluation, its gradients, and evaluation of the input-output minimum mean squared error (MMSE) covariance matrix, as required by the precoder optimization algorithm and other algorithms involved, e.g., the equivalent channel determination in the Statistical CSI case [21].

In this paper, we propose optimal linear precoding techniques for MIMO, suitable for QAM with constellation size $M \geq 16$. An additional advantage of these techniques is their ability to accommodate MIMO configurations with very large antenna sizes, e.g., 100×100. The only related work in this area is [22] which has antenna sizes up to 32×32. We show in the sequel that the proposed method is faster than the one in [22]. Carrying out this calculation 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 Gauss-Hermite quadrature rule [23] 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 [24] in order to determine the globally optimal linear precoder for MIMO systems, in a similar fashion to [16], 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 precoder (MDP) of [25] and the no-precoder case, at low signal-to-noise ratio (SNR) for a standard 2×2 MIMO channel, although the absolute utilization gain achieved is lower than 1 bps/Hz. We then proceed to show that significantly higher gains are available for different channels, e.g., a utilization gain of 1.30 bps/Hz at $\mathsf{SNR} = 10$ dB, when $M = 16$. We then employ larger 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 [26], and show very high gains available with reduced system complexity. Finally, we also employ Statistical CSI scenarios in conjunction with PGP and show very significant gains for large antenna sizes, e.g., 100×100 and $M = 16,\ 32,\ 64$. The main advantages of our work compared with other interesting proposals for large MIMO sizes, e.g., [22], lie over four main directions: a) It offers a precoder solution very close to the globally optimal one for each subgroup, instead of a locally optimal one, b) It is faster, c) It allows for larger constellation size, e.g., $M = 32,\ 64$, and d) It allows larger MIMO configurations, e.g., 100×100. 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 an almost globally optimal linear precoder for 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 instantaneous N_t transmit antenna, N_r receive antenna MIMO model 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, $\mathbf{G}$ 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, $\mathbf{n}$ 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 $\sigma^2 = \frac{1}{\mathsf{SNR}}$, SNR being the (coded) symbol signal-to-noise ratio. In this paper, a number of different channels are considered, e.g., channels comprising independent complex Gaussian components or spatially correlated Kronecker-type channels [19] (including those similar to the 3GPP spatial correlation model (SCM) [27]), or more generally canonical [28] or Weichselberger channels [20]. 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 called herein the "virtual" channel is given by [16]

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

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 a capacity approaching code, e.g., the LDPC code is employed in this MIMO system, the overall utilization in bps/Hz is determined by the mutual information between the transmitting symbols $\mathbf{x}$ and the receiving ones, $\mathbf{y}$ [29], [30]. It is shown [16] that the mutual information between $\mathbf{x}$ and $\mathbf{y}$, for channel realization $\mathbf{H}$, $I(\mathbf{x}; \mathbf{y})$, is only a function of $\mathbf{W} = \mathbf{V}_G \mathbf{\Sigma}_H^2 \mathbf{\Sigma}_G^2 \mathbf{V}_G^h$. The optimal CSIT precoder $\mathbf{G}$ is found by solving:

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

called the "original problem," and

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

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 $M \geq 16$ or large MIMO configurations. 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 [16], 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 CSIT. In the Statistical CSI case, the corresponding optimization problem becomes

$$\begin{aligned} &\underset{\boldsymbol{G}}{\text{maximize}}\ \mathbb{E}_{\mathbf{H}}\{I(\mathbf{x};\mathbf{y})\} \\ &\text{subject to } \mathrm{tr}(\mathbf{G}\mathbf{G}^h) = N_t, \end{aligned} \tag{6}$$

where the expectation is performed over all the channels $\mathbf{H}$. The ground-breaking work of [21] has shown that the problem in (6) for large antenna sizes can be solved by an approximate way of calculating the ergodic mutual information $\mathbb{E}_{\mathbf{H}}\{I(\mathbf{x};\mathbf{y})\}$ for a fixed precoding matrix $\mathbf{G}$ through well-determined parameters of a deterministic channel. More specifically, when $N_t,\ N_r \to \infty$ while the ratio N_t/N_r is kept constant, [21] has showed that an asymptotic methodology to approximate the ergodic mutual information exists, it also presented a methodology to determine the asymptotically optimal precoder for the Statistical CSI case for Weichselberger channels. The evaluation of the mutual information and the optimal asymptotic precoder, under these conditions, entail evaluations of equivalent deterministic channel mutual information multiple times. In [22] the same approach was used for Kronecker channels requiring the evaluation of an equivalent deterministic channel. These parameters include the mutual information of a corresponding deterministic channel defined in [22, eq. (19)], i.e., similar to CSIT scenario mutual information evaluation. Thus, methods that offer simplification of CSIT mutual information evaluation, $I(\mathbf{x};\mathbf{y})$, are also important in the Statistical CSI case toward determining the globally optimal linear precoder for 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.,

$$F \doteq \int_{-\infty}^{+\infty} \exp(-x^2) f(x) dx, \tag{7}$$

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

$$F \approx \sum_{l=1}^{L} c(l) f(\upsilon_l) = \mathbf{c}^t \mathbf{f}, \tag{8}$$

with $\mathbf{c} = [c(1)\cdots c(L)]^t$, $\{\upsilon_l\}_{l=1}^{L}$, and $\mathbf{f} = [f(\upsilon_1)\cdots f(\upsilon_L)]^t$, being the vector of the weights, the nodes, and function node values, respectively (see Appendix A), a 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 and imaginary parts as follows

$$\mathbf{n}_e = [n_{r1}\ n_{i1}\cdots n_{rN_r}\ n_{iN_r}]^t, \tag{9}$$

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

$$\mathbf{v}(\{k_{r\upsilon},\ k_{i\upsilon}\}_{\upsilon=1}^{N_r}) = [\upsilon_{kr1}\ \upsilon_{ki1}, \cdots, \upsilon_{krN_r}\ \upsilon_{kiN_r}]^t, \tag{10}$$

with $k_{r\upsilon},\ k_{i\upsilon}$ $(1 \le \upsilon \le N_r)$ 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

$$I(\mathbf{x};\mathbf{y}) \approx N_t \log_2(M) - \frac{N_r}{\log(2)} - \frac{1}{M^{N_t}} \sum_{k=1}^{M^{N_t}} \hat{f}_k, \tag{11}$$

where

$$\begin{aligned} \hat{f}_k = \left(\frac{1}{\pi}\right)^{N_r} &\sum_{k_{r1}=1}^{L}\sum_{k_{i1}=1}^{L}\cdots\sum_{k_{rN_r}=1}^{L}\sum_{k_{iN_r}=1}^{L} c(k_{r1})c(k_{i1}) \\ &\cdots c(k_{rN_r})c(k_{iN_r}) g_k(\sigma n_{k_{r1}}, \sigma n_{k_{i1}}, \cdots, \sigma n_{k_{rN_r}}, \sigma n_{k_{iN_r}}), \end{aligned} \tag{12}$$

with

$$g_k(\sigma \upsilon_{k_{r1}}, \sigma \upsilon_{k_{i1}}, \cdots, \sigma \upsilon_{k_{rN_r}}, \sigma \upsilon_{k_{iN_r}}) \tag{13}$$

being the value of the function

$$\log_2\left(\sum_m \exp\left(-\frac{1}{\sigma^2}||\mathbf{n} - \mathbf{H}\mathbf{G}(\mathbf{x}_k - \mathbf{x}_m)||^2\right)\right) \tag{14}$$

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

The proof of this lemma is presented in Appendix A, together with a derivation methodology available for this expression in the $N_t = N_t$ case. In Appendix A, we also show that $\hat{f}_k$ is an approximation to the physical quantity Input-Dependent Output Entropy (IDOE) which relates to the negative MIMO output entropy when the MIMO input is $\mathbf{x} = \mathbf{x}_k$. Let us stress that, the presented novel application of the Gauss-Hermite quadrature in the MIMO model allows for efficient evaluation of $I(\mathbf{x};\mathbf{y})$ for any channel matrix $\mathbf{H}$, and precoder $\mathbf{G}$, as required in the precoder optimization process, as explained below.

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 [16], 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 [16] 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 execution of the optimization algorithm. The value of the step of each iteration over $\mathbf{W}$ and $\mathbf{\Sigma}_G^2$ is determined through backtracking line searches, one for each of the variables $\mathbf{W}$ and $\mathbf{\Sigma}_G^2$. We describe the

backtracking line search for $\mathbf{W}$ first, then the one over $\mathbf{\Sigma}_G^2$. 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 employ a novel method in determining $\nabla_{\mathbf{W}} I$ at each iteration, as described in detail in the next subsection. 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

$$I(\mathbf{W} + t\nabla_{\mathbf{W}} I) > I(\mathbf{W}) + \alpha_1 t ||\nabla_{\mathbf{W}} I||_F^2, \tag{15}$$

where $||\nabla_{\mathbf{W}} I||_F$ is the Frobenius norm of $\nabla_{\mathbf{W}} I$ [31] and we used the notation $I(\mathbf{W})$ to mean the value of $I(\mathbf{x};\mathbf{y})$ at a fixed $\mathbf{W}$. If the condition is satisfied, the algorithm proceeds with calculating a new $\mathbf{\Theta}_{NEW}$ from $\mathbf{W}_{NEW} = \mathbf{W} + t_1 \nabla_{\mathbf{W}} I$, as follows:

$$\mathbf{W}_{NEW} = \mathbf{\Theta}_{NEW}^h \mathbf{\Sigma}_H^2 \mathbf{\Sigma}_G^2 \mathbf{\Theta}_{NEW} \tag{16}$$

by employing the eigenvalue decomposition (EVD) of $\mathbf{W}_{NEW}$, it updates $\mathbf{V}_G^h = \mathbf{\Theta}_{NEW}$, and then it proceeds to the backtracking line search over $\mathbf{\Sigma}_G^2$. 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 $\mathbf{\Sigma}_G^2$ 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

$$I(\mathbf{W} + t\nabla_{\mathbf{\Sigma}_G^2} I) > I(\mathbf{W}) + \alpha_2 t ||\nabla_{\mathbf{\Sigma}_G^2} I||_F^2, \tag{17}$$

where $||\nabla_{\mathbf{\Sigma}_G^2} I||_F$ is the Frobenius norm of $\nabla_{\mathbf{\Sigma}_G^2} I$ [31]. If the condition is satisfied, the algorithm proceeds with updating to the new $\mathbf{\Sigma}_{NEW}$, but after setting any negative terms in the main diagonal of $\mathbf{\Sigma}_{NEW}$ 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. We thus see that there will in general be multiple evaluations of $I(\mathbf{x};\mathbf{y})$, 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 α_1, α_2, β_1, β_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 the possibility of finding the globally optimal precoder for large QAM constellations with $M \leq 16$ or large MIMO configurations becomes reality, as our results demonstrate. The algorithm's pseudocode for a number of iterations t is presented under the heading Algorithm 1.

Algorithm 1 Global Precoder Optimization Algorithm With t Iterations

```
procedure PRECODER(Σ_H)
  while i ≤ t do
    Determine W_NEW = Θ^h Σ^2 Θ through backtracking line search (16)
    Set V_G^h = Θ
    Determine W_NEW = V_G Σ_G^2 Σ_H^2 V_G^h
    Determine Σ_{NEW,G}^2 through backtracking line search (17)
    Set negative entries on the diagonal of Σ_{NEW,G}^2 to zero
    Normalize Σ_{NEW,G}^2 to a trace equal to N_t
    Set Σ_G = Σ_{NEW,G}
    Determine W_NEW = V_G Σ_G^2 Σ_H^2 V_G^h
    Set W = W_NEW
    Evaluate I(W)
  end while
  return I(W)
end procedure
```

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., [21] 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 (11) 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

$$\tilde{\mathbf{y}} = \mathbf{V}_G \mathbf{y} = \mathbf{V}_G \mathbf{\Sigma}_H \mathbf{\Sigma}_G \mathbf{V}_G^h \mathbf{x} + \mathbf{V}_G \mathbf{n}, \tag{18}$$

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 (18), we see that we get the desired result, i.e., the value of $I(\mathbf{x};\mathbf{y})$ 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 theorem, an alternative proof of part of [16, Th. 1] can be developed, namely the fact that $I(\mathbf{x};\mathbf{y})$ is only a function of $\mathbf{W}$, as $\mathbf{My}$ is a sufficient statistic for

[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.

$\mathbf{y}$ and $\mathbf{M}$ is a function of $\mathbf{W}$. Assume without loss of generality that $N_t = N_r$. The gradient of $I(\mathbf{x};\mathbf{y})$ with respect to $\mathbf{M}$ can be found (see Appendix B for the derivation) from the Gauss-Hermite expression presented in (11) as follows

$$\nabla_{\mathbf{M}} I = -\frac{1}{\log(2)}\frac{1}{M^{N_t}}\left(\frac{1}{\pi}\right)^{N_r}\sum_{k_{r1}=1}^{L}\sum_{k_{i1}=1}^{L}\cdots\sum_{k_{rN_r}=1}^{L}\sum_{k_{iN_r}=1}^{L} \times c(k_{r1})\cdots c(k_{iN_r})\mathbf{R} \times (\sigma\upsilon_{k_{r1}}, \sigma\upsilon_{k_{i1}}, \cdots, \sigma\upsilon_{k_{rN_r}}, \sigma\upsilon_{k_{iN_r}}), \quad (19)$$

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

$$\sum_k \frac{1}{\sum_m \exp(-\frac{1}{\sigma^2}||\mathbf{n}-\mathbf{M}(\mathbf{x}_k-\mathbf{x}_m)||^2)} \times \sum_m \exp(-\frac{1}{\sigma^2}||\mathbf{n}-\mathbf{M}(\mathbf{x}_k-\mathbf{x}_m)||^2) \times ((\mathbf{n}-\mathbf{M}(\mathbf{x}_k-\mathbf{x}_m))(\mathbf{x}_k-\mathbf{x}_m)^h + ((\mathbf{n}-\mathbf{M}(\mathbf{x}_k-\mathbf{x}_m))(\mathbf{x}_k-\mathbf{x}_m)^h)^h) \quad (20)$$

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 [31]), 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.*

$$\nabla_{\mathbf{W}} I \approx \text{reshape}((\text{vec}(\nabla_{\mathbf{M}} I)^T\left((\mathbf{M}^*)\otimes\mathbf{I}+\mathbf{I}\otimes\mathbf{M}\right)^{-1}), N_t, N_t), \quad (21)$$

where reshape$(\mathbf{A}, k, n)$ *is the standard reshape of a 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.* Standard reshape emanates from the vector vec$(\mathbf{A})$ of matrix $\mathbf{A}$ which encompasses all columns of $\mathbf{A}$ starting from the leftmost one to the rightmost.

Then, as $I(\mathbf{x};\mathbf{y})$ is a concave function of $\mathbf{W}$ [21], we can maximize over $\mathbf{W}$ in a straightforward way using closed form expressions. This is based on the fact that the approximated $I(\mathbf{x};\mathbf{y})$ through the Gauss-Hermite approximation is very accurate, as shown in the next section. The utilization of this lemma goes beyond the exploitation of the gradient of the mutual information in the precoder optimization algorithm. As it is well known from [16], $\nabla_{\mathbf{W}} I = \mathbf{\Phi}$, where $\mathbf{\Phi} = \mathbb{E}\left\{\mathbb{E}\{(\mathbf{x}-\mathbb{E}(\mathbf{x}|\mathbf{y}))(\mathbf{x}-\mathbb{E}(\mathbf{x}|\mathbf{y}))^h|\mathbf{y}\}\right\}$, is the minimum mean squared error (MMSE) covariance matrix of the channel. Thus, by using the current lemma, an accurate estimate of the MMSE covariance matrix of the MIMO channel can be achieved. This is very useful especially when dealing with, e.g., Statistical CSI cases where, the MMSE covariance matrix becomes instrumental in deriving the asymptotically optimal precoder [21]. Thus, based on the proposed Gauss-Hermite approximation, accurate, but otherwise simplified derivation of the MMSE covariance matrix of the channel becomes possible. Finally, since from [32] 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 procedure presented above.

TABLE I

COMPUTATIONAL COMPLEXITY PARAMETERS OF GAUSS-HERMIT QUADRATURE $I(\mathbf{x},\mathbf{y})$ EVALUATION

	Complexity	*Memory Requirement*	$I(\mathbf{x};\mathbf{y})$ *Single Run CPU sec*	$\nabla_{\mathbf{W}} I$ *Run CPU sec*
$M=16$	$16^{2N_t}(2L-1)^4L^4N_r(2N_t+N_r-1)$	L^{2N_r}	25	54
$M=32$	$32^{2N_t}(2L-1)^4L^4N_r(2N_t+N_r-1)$	L^{2N_r}	54	126
$M=64$	$64^{2N_t}(2L-1)^4L^4N_r(2N_t+N_r-1)$	L^{2N_r}	50.20	110.60

C. Complexity of the Globally Optimal Precoder Determination Method

The complexity of the globally optimal precoder determination method depends mainly on the evaluations of $I(\mathbf{x};\mathbf{y})$ as required by the optimization algorithm described above. By employing the Gauss-Hermite approximation described herein it becomes possible to significantly accelerate the evaluations of $I(\mathbf{x};\mathbf{y})$ at each iteration of the optimization algorithm. However, the complexity is still high: Each evaluation of $I(\mathbf{x};\mathbf{y})$ through the Gauss-Hermite approximation requires, as per the development in Appendix C, employing L weights and L nodes per each real and imaginary component of the noise vector $\mathbf{n}$, resulting in $2N_tL$ total weight and node dimensions. Then, evaluation of f_k in (30) requires $2N_r$ nested "DO" loops, each of length L, resulting in L^{2N_r} memory parameter values overall. For a good approximation in Gauss-Hermite quadrature, a value of $L \geq 3$ is required (please see the next section for relevant results) and thus the overall memory requirement becomes 3^{2N_r}. As far as the computational complexity in the number of operations (including both (complex) summations and multiplications) involved in calculating $I(\mathbf{x};\mathbf{y})$ through the Gauss-Hermite approximation, the corresponding complexity is shown in Appendix B to be $M^{2N_t}(2L-1)^{(2N_r)}L^4N_r(2N_t+N_r-1)$. Thus, for example, going from $M=16$ to $M=32$ QAM will result in about 4 times higher complexity with N_t and N_r held constant. On the other hand, increasing N_r has an even more profound effect on the complexity, due to its more complicated presence in this complexity equation. For example, increasing N_r from 2 to 3 while keeping all other parameters constant, will increase the complexity by a factor of $(2L-1)^2$ or for $L=3$ by 25 times. Our systematic numerical evaluations corroborate these numbers very closely. As we observe by comparing the figures in Table 1 to the numbers presented in [22], we see that the Gauss-Hermite approach is about 7 times faster per iteration of the precoder for $M=16$ and PGP with groups of size 2×2, i.e., the presented approach is faster. This is one of the major improvements due to the proposed methodology.

V. NUMERICAL RESULTS

The results presented in this subsection employ QAM with 16, 32, or 64 constellation sizes. We employ MIMO systems with $N_t = N_r = 2$ when global precoding optimization is

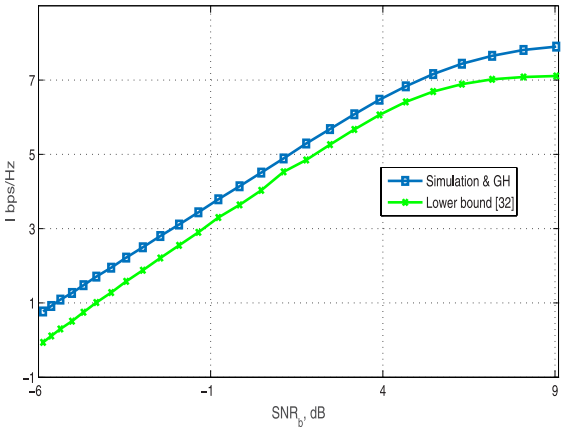


Fig. 1. Results for $I(\mathbf{x};\mathbf{y})$ without precoding for the $\mathbf{H}_1$ channel and QAM $M = 16$ modulation.

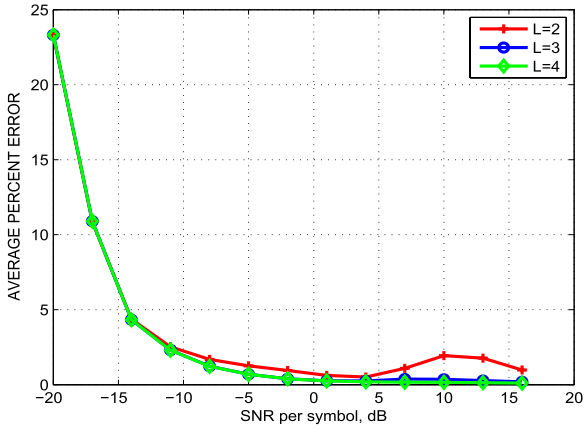


Fig. 2. Results for the average percent absolute normalized error in the Gauss Hermite $I(\mathbf{x};\mathbf{y})$ using 1000 generated random $\mathbf{H}$ channels and QAM $M = 16$ modulation.

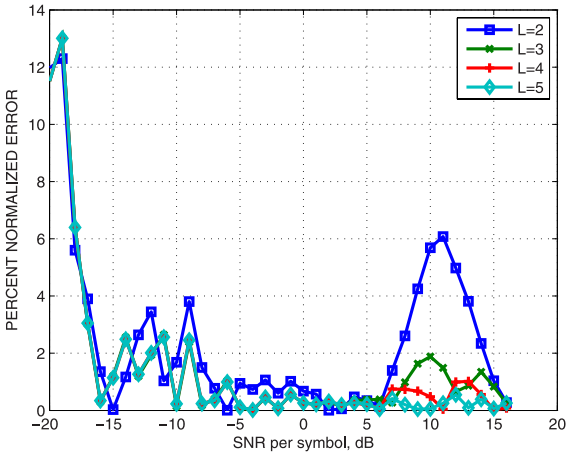


Fig. 3. Results for the average percent absolute normalized error in the Gauss Hermite $I(\mathbf{x};\mathbf{y})$ for the $\mathbf{H}_1$ channel and QAM $M = 16$ modulation.

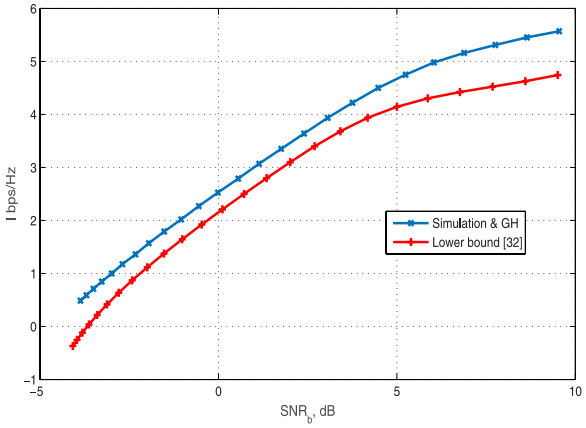


Fig. 4. Results for $I(\mathbf{x};\mathbf{y})$ without precoding for the $\mathbf{H}_2$ channel and QAM $M = 32$ modulation.

performed. We have used an $L = 3$ Gauss-Hermite approximation which results in 3^{2N_r} total nodes due to MIMO. The implementation of the globally optimizing methodology is performed by employing two backtracking line searches, one for $\mathbf{W}$ and another one for $\mathbf{\Sigma}_G^2$ at each iteration, in a fashion similar to [21]. For the results presented, it is worth mentioning that only a few iterations (e.g., typically < 8) are required to converge to the optimal solution results as presented in this paper. We apply the complexity reducing method of PGP [26] which offers semi-optimal results under exponentially lower transmitter and receiver complexity [26]. 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 Statistical CSI cases. We divide this section into five subsections. In the first subsection, we examine the accuracy of the proposed Gauss-Hermite approximation and provide a comparison with the lower bound technique presented in [32]. In the second subsection we show results for the globally optimal precoder for $M = 16,\ 32$ based on the approximation in conjunction with independent Gaussian channels and CSIT. In the third subsection we present results for Statistical CSI channels similar to the ones in 3GPP SCM [27], with antenna size up to 100×100 with PGP and modulation size $M = 16,\ 32$. In the fourth subsection, we present results for CSIT jointly with PGP and higher size of antennas and modulation. Finally, in the last subsection, we present results for a MIMO system with 100 base station antennas and 4 user antennas with $M = 16,\ 64$.

A. Accuracy of the Gauss-Hermite Approximation Technique

In Fig. 1 we present results for I through simulation, the Gauss-Hermite approximation (GH) with $L = 3$, and the lower bound developed in [32] versus the signal-to-noise ratio per bit (SNR_b) in dB, for QAM with $M = 16$, and for the commonly used channel [16], [25],

$$\mathbf{H}_1 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

In the same figure, we also show a lower bound for $I(\mathbf{x};\mathbf{y})$ which appeared in [32]. We see excellent accuracy for the approximation, i.e., no observable difference between the Gauss-Hermite approximation and the simulations, over all SNR_b values. In Fig. 2 we show results on the Gauss-Hermite approximation percent average absolute error normalized by the actual $I(\mathbf{x};\mathbf{y})$ for $L = 2,\ 3,\ 4$, for 1000 randomly generated $\mathbf{H}$ $N_t = N_r = 2$ channels, and $M = 16$, with respect to symbol SNR in dB. We see that $L = 2$ has a maximum normalized error of 6.25%. The average normalized error drops to less than 1% with $L = 3$ for $\mathsf{SNR} > -5$ dB.

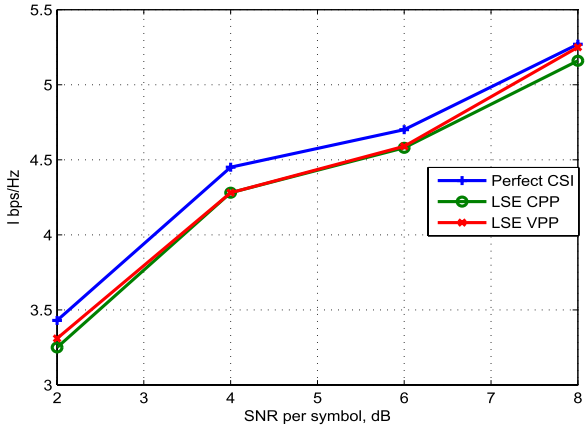


Fig. 5. Results for achievable $I(\mathbf{x};\mathbf{y})$ with channel estimation errors and $M = 16$ modulation.

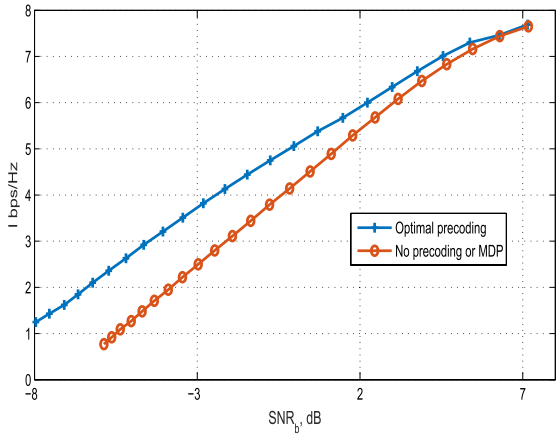


Fig. 6. $I(\mathbf{x};\mathbf{y})$ results for optimal precoding, MDP, and no-precoding cases for a 2×2 MIMO system and QAM $M = 16$ modulation.

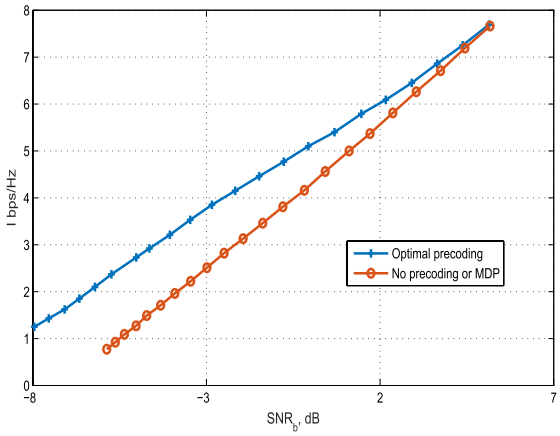


Fig. 7. $I(\mathbf{x};\mathbf{y})$ results for optimal precoding, MDP, and no-precoding cases for a 2×2 $\mathbf{H}_1$ MIMO system and QAM $M = 32$ modulation.

Furthermore, the results for $L = 4$ give little improvement over the ones with $L = 3$. Notice that the same type of behavior was found for $M = 32,\ 64$ over a wide variety of channels. We thus selected $L = 3$ in our work as the best tradeoff value between low approximation error and low system complexity. Furtermore, it is important to stress that the variance of the approximation error was found to be less than $4 \cdot 10^{-3}$ for $L \geq 2$ and $\mathsf{SNR} > -15$ dB. We thus see the Gauss-Hermite approximation offers excellent accuracy when $L = 3$ and for a wide SNR range. In Fig. 3 we present percent normalized absolute error results for $\mathbf{H}_1$ and $L = 1, 2, \cdots, 5$ in order to assess the impact of the value of L to the results, especially for higher L. We clearly see that by going from $L = 2$ to $L = 3$, a significant reduction on the error is achieved. Further increase of L offers an additional reduction to the error, however it requires significant additional complexity and processing delay. We thus see that $L = 3$ comes as the natural candidate due to its capability to attain percent error lower than 2% for a wide range of SNR. In Fig. 4 we present the corresponding results to Fig. 1 for $M = 32$ QAM in conjunction with the randomly generated channel

$$\mathbf{H}_2 = \begin{bmatrix} 1.98 + j0.12 & 0.0124 - j0.0016 \\ -0.2487 - j0.0314 & 0.0992 - j0.1 \end{bmatrix},$$

with the same type of behavior as before, i.e., the Gauss-Hermite approximation offers excellent accuracy and that the lower bound is lagging behind in performance, albeit by less than $N_r(1/\log(2) - 1) \approx 0.88$, which is the shift introduced in [32] in order to approximate I very closely for QPSK modulation. In Fig. 5 we present results with non-perfect channel estimation, i.e., channel estimation with errors for $N_t = N_r = 2$ [33]. To that end, we employ two channel estimators and use their imperfect estimation to model errors. We use two possible Least Squares Estimation (LSE) methods in conjunction with 2 orthogonal pilots, each of length 2. In the first method, a simple LSE with pilot power equal to the 7 dB independently of the message signal power, called Constant Pilot Power (CPP), while in the second one we use pilots with two times the message signal power, called Variable Pilot Power (VPP). We generate 50 random channels, employ LSE with both CPP and VCC, and then average the corresponding $I(\mathbf{x};\mathbf{y})$ on these channels. We observe that both proposed LSE methods attain similar results which show little impact of channel estimation errors to the overall system performance, e.g., at $\mathsf{SNR} = 4$ dB there is degradation of 3.8% in the precoder performance. In addition, since the two methods proposed are simple, in reality, with more sophisticated channel estimation, we expect a smaller impact of channel estimation errors to the precoder design and performance.

B. Results for Globally Optimal Precoding for Gaussian Channels and CSIT

In Fig. 6 we show results for the globally optimal precoder based on the Gauss-Hermite approximation with $L = 3$, and the Maximum Diversity Precoder (MDP) presented in [25], for $M = 16$, for $\mathbf{H}_1$. We observe that the optimal precoder offers significant utilization gains in the low SNR region, while its gain diminishes in the higher SNR region. For example, at $\mathsf{SNR}_b = -4$ dB, there is about 0.6 bps/Hz gain attainable with the globally optimal precoder over its MDP and no-precoding counterparts, which represents a 30% gain. Also, contrary to the BPSK/QPSK modulation case presented in [16] and [32], the MDP precoder offers no significant gain over the no-precoding case, a very important difference.

In Fig. 7 we present results for $I(\mathbf{x};\mathbf{y})$, for the same channel, MDP, and no precoding, $N_t = N_r = 2$, and $M = 32$. There is no noticeable improvement offered by MDP over the no-precoding case, similarly to the $M = 16$ QAM case

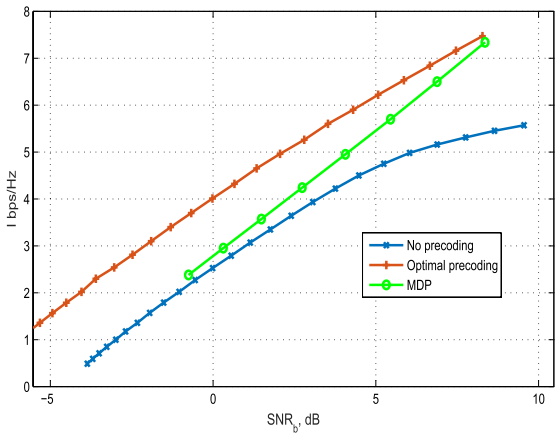


Fig. 8. $I(\mathbf{x};\mathbf{y})$ results for optimal precoding, MDP, and no-precoding cases for a 2×2 $\mathbf{H}_2$ MIMO system and QAM $M=32$ modulation.

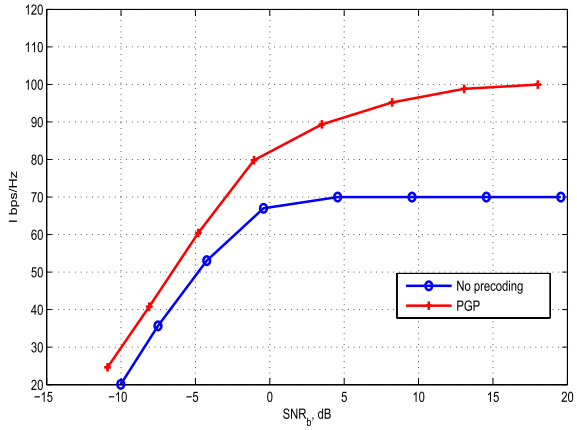


Fig. 9. $I(\mathbf{x};\mathbf{y})$ results for PGP and no-precoding cases for a 20×20 $\mathbf{H}$ Statistical CSI MIMO system and QAM $M=32$ modulation.

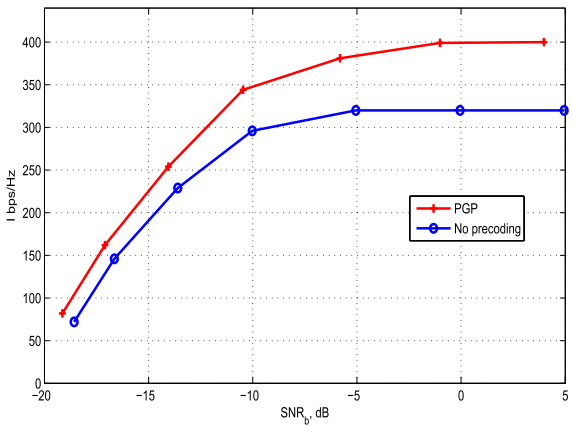


Fig. 10. $I(\mathbf{x};\mathbf{y})$ results for PGP and no-precoding cases for a 100×100 $\mathbf{H}$ Statistical CSI MIMO system and QAM $M=16$ modulation.

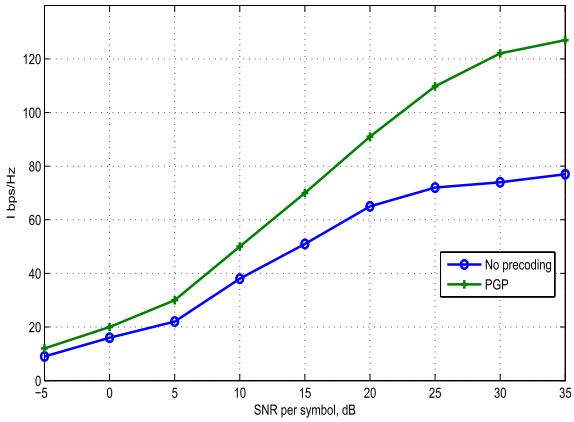


Fig. 11. $I(\mathbf{x};\mathbf{y})$ results for PGP and no-precoding cases for the 32×32 $\mathbf{H}$ Statistical CSI MIMO channel employed in [22] in conjunction with QAM $M=16$ modulation.

presented above. Clearly, the same fundamental conclusions as in the $M=16$ case hold true. The offered gain of the globally optimal precoder in low SNR is still around 50% over its no- precoding and MDP counterparts. We next present same type of results for $\mathbf{H}_2$. In Fig. 8 we observe that for this type of channel, the gains achieved by the globally optimal precoder are significantly higher and that in the high SNR regime the no-precoding case cannot achieve the maximum mutual information given by $N_t\log_2(M)=10$, i.e., it becomes saturated. We will see that for certain channels, which we call saturated channels, this type of behavior is also observed for the large MIMO channel configurations in both CSIT and Statistical CSI channel cases.

C. Results for Statistical CSI in Conjunction With PGP

We apply the asymptotic approximation to ergodic mutual information precoder optimization, e.g., by determining the parameters $\Xi_{eq},\Omega_{eq},\gamma_{eq},\psi_{eq},\mathbf{R}_{eq}$ in [22] for Kronecker channels. We first consider an $N_t=N_r=20$ SCM urban channel with $M=32$. We create the large size asymptotic approximation no-precoding results for a Statistical CSI channel based on [21]. For precoding to be efficient, but otherwise realistic, we employ PGP [26] with 10 groups of size 2×2 each. Due to the nature of this channel, we observe in Fig. 9 that the no-precoding case saturates at $I(\mathbf{x};\mathbf{y})=70$ bps/Hz. We see very significant gains offered by PGP in the high SNR_b regime. The PGP system attains the full capacity available in high SNR_b. In Fig. 10 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 large MIMO configurations are 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 bps/Hz while the no-precoding scheme saturates at 320 bps/Hz. The PGP system employed uses 50 groups of size 2×2 each. In Fig. 11 we present results for PGP versus a no-precoding SCM channel with $N_t=N_r=32$ and $M=16$, using the same SCM

$$\mathbf{H}=\begin{bmatrix} 1.5362+0.3151i & 0.5714+0.9123i & 0.1394-0.3407i & -0.0085+0.0081i \\ 1.5571+1.0171i & -0.3071+0.3765i & -0.3073+0.5680i & -0.0035+0.0041i \\ 0.4550-0.2484i & 0.7266-1.2195i & 0.0780+0.1645i & -0.0131+0.0008i \\ 0.2278+3.1243i & -0.6890-0.3397i & 0.0175-0.2322i & -0.0045-0.0064i \end{bmatrix},$$

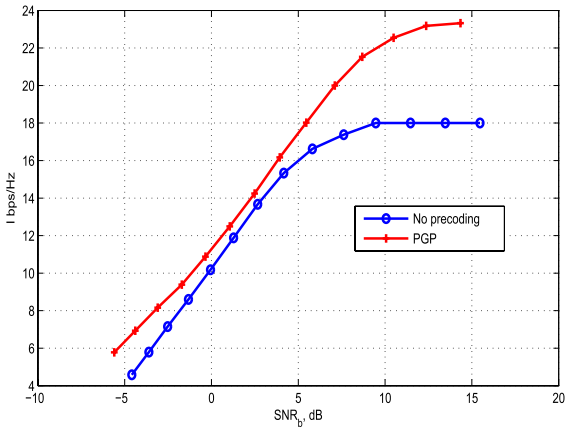


Fig. 12. $I(\mathbf{x};\mathbf{y})$ results for PGP and no-precoding cases for a 4×4 $\mathbf{H}$ CSIT MIMO system and QAM $M=64$ modulation.

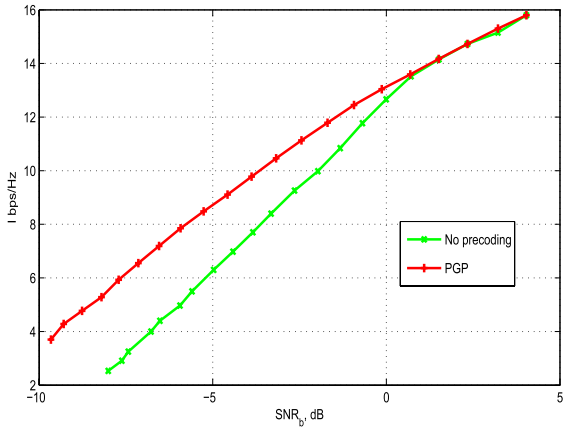


Fig. 13. $I(\mathbf{x};\mathbf{y})$ results for PGP and no-precoding cases for a randomly generated 10×4 $\mathbf{H}$ CSIT MIMO system and QAM $M=16$ modulation.

channel as in [22]. We observe that our proposed approach offers slightly better performance than the one employed in [22], although it is faster, as it is explained in Table 1.

D. Results for CSIT in Conjunction With PGP

For a 4×4 channel $\mathbf{H}$, as shown at the bottom of the previous page, used with $M=64$ we get the no-precoding and the PGP results using 2 groups of size 2×2 each depicted in Fig. 12. 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, these type of results for optimal precoding in conjunction with $M=64$ are not available in the literature. Finally, in Fig. 13 we present results for an asymmetric randomly generated MIMO channel with $N_t=4$, $N_r=10$, and $M=16$. PGP employs two 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.

E. Results for Massive MIMO

Massive MIMO [34]–[36] has attracted much interest recently, due to its potential to offer high data rates. We present results for the uplink and downlink of a Massive

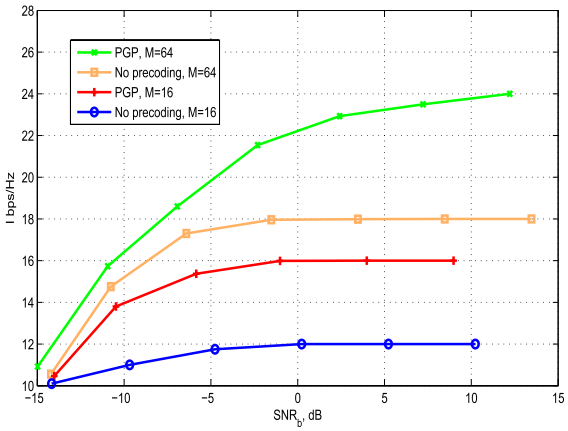


Fig. 14. $I(\mathbf{x};\mathbf{y})$ results for PGP and no-precoding cases for a randomly generated 100×4 uplink $\mathbf{H}$ CSIT MIMO system and QAM $M=16$ modulation.

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 in a single user configuration. Fig. 14 shows results for the 100×4 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 bps/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 $M=64$ uplink offers further gains by, e.g., achieving the maximum possible rate of 24 bps/Hz. For the downlink, in Fig. 15 we show results where the no-precoding case operates under 100 antenna inputs all correlated through the right singular vectors of the channel, thus creating a very demanding environment at the user, due to the exponentially increasing maximum a posteriori (MAP) detector complexity [26]. 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 bps/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., $\mathsf{SNR}_b=4$, the $M=64$ PGP scheme achieves 45% higher throughput that the $M=16$ one, a significant improvement. We would 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,

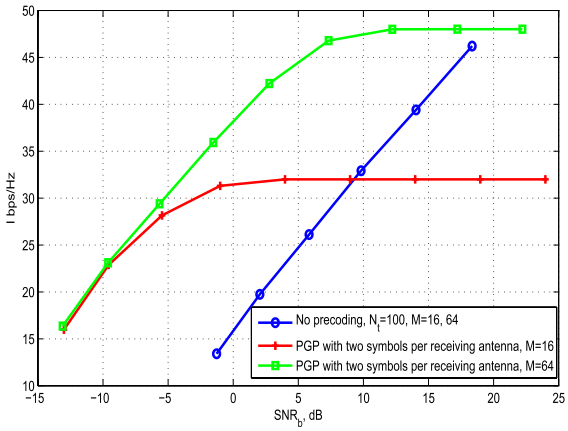


Fig. 15. $I(\mathbf{x};\mathbf{y})$ results for PGP and no-precoding cases for a randomly generated 4×100 $\mathbf{H}$ downlink CSIT MIMO system and QAM $M = 16, 64$ modulation.

in lower SNR, the PGP scheme achieves both much higher throughput with simultaneously exponentially lower MAP detector complexity at the user site detector. Note that the performance provided by PGP on the downlink depends on the number of symbols processed jointly. This number is currently limited by the computational complexity available. This limitation does not occur on the uplink since in that case $N_t < N_r$.

VI. Conclusions

In this paper, the problem of designing a linear precoder for MIMO systems toward mutual information maximization is addressed for QAM with $M \geq 16$ and in conjunction with large 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. We then applied a global optimization framework to derive an almost globally optimal precoder for the case of QAM with $M = 16$ and 32 and small antenna size configurations. We showed that under CSIT in this case, significant gains are available for the lower SNR range over no precoding, or MDP. We showed that for the standard 2×2 channel, although the globally optimal precoder offers significant gains over MDP and the no-precoder configurations in the low SNR region, it fails to offer gains as high as 1 bps/Hz. However, we demonstrated that by employing another 2×2 channel gains as high as 1.4 bps/Hz are possible. For systems of large MIMO configurations, we applied the complexity-simplifying PGP concept [26] to derive semi-optimal precoding results. Under Statistical CSI, we showed that for urban 3GPP SCM channels, an interesting saturation effect in the no-precoding case takes place, while the semi-optimal PGP precoder offers dramatically better results in this case and it does not experience any saturation as the SNR increases, e.g., it achieves the maximum information rate, $I(\mathbf{x};\mathbf{y}) = N_t \log_2(M)$ at high SNR. Furthermore, we applied the same Gauss-Hermite approximation approach to CSIT with a large number of antennas with the same success. We showed that for specific type of channels similar to urban 3GPP SCM [27], the PGP approach offers very high gains over the no-precoding case in the high SNR_b regime. Finally, we considered a Massive MIMO scenario in conjunction with CSIT and showed that by carefully designing the downlink and uplink precoders, the methodology shows very high gains, especially on the downlink, although it employs an exponentially 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 [16] and [21] 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. In addition, the presented Gauss-Hermite approximation offers important simplification in the evaluation of the MMSE covariance matrix of the MIMO channel which is required in, among other areas, the Statistical CSI equivalent channel determination [21].

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 [32]

$$\begin{aligned} H(\mathbf{x}|\mathbf{y}) &= \frac{N_r}{\log(2)} + \frac{1}{M^{N_t}} \sum_k \\ &\quad \times \mathbb{E}_{\mathbf{n}}\left(\log_2\left(\sum_m \exp\left(-\frac{1}{\sigma^2}||\mathbf{n} - \mathbf{HG}(\mathbf{x}_k - \mathbf{x}_m)||^2\right)\right)\right) \\ &= \frac{N_r}{\log(2)} + \frac{1}{M^{N_t}} \sum_k \int_{-\infty}^{+\infty} \mathcal{N}_c(\mathbf{n}|\mathbf{0},\sigma^2\mathbf{I}) \\ &\quad \times \log_2\left(\sum_m \exp\left(-\frac{1}{\sigma^2}||\mathbf{n} - \mathbf{HG}(\mathbf{x}_k - \mathbf{x}_m)||^2\right)\right) d\mathbf{n}, \end{aligned} \tag{22}$$

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

$$\begin{aligned} f_k &\doteq \int_{-\infty}^{+\infty} \mathcal{N}_c(\mathbf{n}|\mathbf{0},\sigma^2\mathbf{I}) \\ &\quad \times \log_2\left(\sum_{\mathbf{x}_m} \exp\left(-\frac{1}{\sigma^2}||\mathbf{n} - \mathbf{HG}(\mathbf{x}_k - \mathbf{x}_m)||^2\right)\right) d\mathbf{n}. \end{aligned} \tag{23}$$

There is an intimate connection between f_k and the parameter $H_k(\mathbf{y}) \doteq \mathbb{E}_{\mathbf{y}|\mathbf{x}_k}\{-\log_2(p(\mathbf{y}))|\mathbf{x} = \mathbf{x}_k\}$ called the Input-Dependent Output Entropy (IDOE) herein. Note that IDOE represents the entropy at the receiver output when the input is $\mathbf{x}_k$. This parameter is different that the conditional entropy. By using standard entropic identities, we can easily see that

$$f_k = -H_k(\mathbf{y}) + 2N_r \log_2(\sigma) + N_r \log_2(\pi) + N_t \log_2(M). \tag{24}$$

Thus, since our Gauss-Hermite approximation focuses on finding approximations to each of the f_k terms, one per input symbol, it equivalently offers estimates of the IDOE terms $H_k(\mathbf{y})$. This gives a physical meaning to the estimated terms $\hat{f}_k$ presented below.

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 $n_{r\upsilon}, n_{i\upsilon}$, with $\upsilon = 1, \cdots, N_r$, the υth receiving antenna real and imaginary noise component, respectively. Also define by $(\mathbf{HG}(\mathbf{x}_k - \mathbf{x}_m))_{r\upsilon}$ and $(\mathbf{HG}(\mathbf{x}_k - \mathbf{x}_m))_{i\upsilon}$, the υth receiving antenna real and imaginary component of $(\mathbf{HG}(\mathbf{x}_k - \mathbf{x}_m))$, respectively. We then have

$$\mathcal{N}_c(\mathbf{n}|\mathbf{0}, \sigma^2\mathbf{I}) = \frac{1}{\pi^{N_r}\sigma^{2N_r}} \exp(-\frac{\sum_l n_{r\upsilon}^2 + n_{i\upsilon}^2}{\sigma^2}), \quad (25)$$

$$d\mathbf{n} = \prod_{\upsilon=1}^{N_r} dn_{r\upsilon} dn_{i\upsilon}, \quad (26)$$

and

$$\sum_m \exp(-\frac{1}{\sigma^2}||\mathbf{n} - \mathbf{HG}(\mathbf{x}_k - \mathbf{x}_m)||^2) = \sum_m \exp(-\frac{1}{\sigma^2}(\sum_\upsilon (n_{r\upsilon} - (\mathbf{HG}(\mathbf{x}_k - \mathbf{x}_m))_{r\upsilon})^2 + \sum_\upsilon (n_{i\upsilon} - (\mathbf{HG}(\mathbf{x}_k - \mathbf{x}_m))_{i\upsilon})^2)). \quad (27)$$

The Gauss-Hermite quadrature is as follows:

$$\int_{-\infty}^{+\infty} \exp(-x^2) f(x) dx \approx \sum_{l=1}^{L} c(l) f(\upsilon_l), \quad (28)$$

for any real function $f(x)$, and with vector $\mathbf{c} = [c(1) \cdots c(L)]^T$ being the "weights," and υ_l are the "nodes" of the approximation. The approximation is based on the following weights and nodes [23]

$$c(l) = \frac{2^{L-1} L! \sqrt{2}\pi}{L^2 (H_{L-1}(\upsilon_l))^2} \quad (29)$$

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 υ_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 (23), and after changing variables, we get that

$$f_k \approx \hat{f}_k = \left(\frac{1}{\pi}\right)^{N_r} \sum_{k_{r1}=1}^{L} \sum_{k_{i1}=1}^{L} \cdots \sum_{k_{rN_r}=1}^{L} \sum_{k_{iN_r}=1}^{L} c(k_{r1}) c(k_{i1}) \cdots \times c(k_{iN_r}) g_k(\sigma n_{k_{r1}}, \sigma n_{k_{i1}}, \cdots, \sigma n_{k_{rN_r}}, \sigma n_{k_{iN_r}}), \quad (30)$$

where

$$g_k(\sigma n_{k_{r1}}, \sigma n_{k_{i1}}, \cdots, \sigma n_{k_{rN_r}}, \sigma n_{k_{iN_r}}) \quad (31)$$

is the value of the function (from (30))

$$\log_2 \left(\sum_m \exp(-\frac{1}{\sigma^2} ||\mathbf{n} - \mathbf{HG}(\mathbf{x}_k - \mathbf{x}_m)||^2) \right) \quad (32)$$

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

For the special case $N_r = 2$, (30) becomes

$$\hat{f}_k = \left(\frac{1}{\pi}\right)^2 \sum_{k_{r1}=1}^{L} \sum_{k_{i1}=1}^{L} \sum_{k_{r2}=1}^{L} \sum_{k_{i2}=1}^{L} c(k_{r1}) c(k_{i1}) \times c(k_{r2}) c(k_{i2}) g_k(\sigma n_{k_{r1}}, \sigma n_{k_{i1}}, \sigma n_{k_{r2}}, \sigma n_{k_{i2}}), \quad (33)$$

which using basic properties of bilinear forms can be rewritten as

$$\hat{f}_k = \left(\frac{1}{\pi}\right)^2 \sum_{k_{r1}=1}^{L} \sum_{k_{i1}=1}^{L} \sum_{k_{r2}=1}^{L} \sum_{k_{i2}=1}^{L} c(k_{r1}) c(k_{i1}) c(k_{r2}) \times c(k_{i2}) g_k(\sigma n k_{r1}, \sigma n_{k_{i1}}, \sigma n_{k_{r2}}, \sigma n_{k_{i2}}) = \left(\frac{1}{\pi}\right)^2 \mathbf{c}^t \mathbf{F}_{k,1} \mathbf{c}, \quad (34)$$

where $\mathbf{F}_{k,1}$ is an $L \times L$ matrix with k_{r1}, k_{i1} element equal to

$$\mathbf{F}_{k,1}[k_{r1}, k_{i1}] = \mathbf{c}^t \mathbf{V}_{k,1} \mathbf{c}, \quad (35)$$

with $\mathbf{V}_{k,1}$ being an $L \times L$ matrix with k_{r2}, k_{i2} element equal to

$$\mathbf{V}_{k,1}[k_{r2}, k_{i2}] = g_k(\sigma n_{k_{r1}}, \sigma n_{k_{i1}}, \sigma n_{k_{r2}}, \sigma n_{k_{i2}}), \quad (36)$$

so that the different f_k can be approximated efficiently, and then summing them over the different $\mathbf{x}_k$, as per (22), we get an approximation for $H(\mathbf{x}|\mathbf{y})$,

$$\begin{aligned} H(\mathbf{x}|\mathbf{y}) &\approx \frac{N_r}{\log(2)} + \frac{1}{M^{N_t}} \sum_k \hat{f}_k \\ &= \frac{N_r}{\log(2)} + \frac{1}{M^{N_t}} \sum_k \mathbf{c}^t \mathbf{V}_k \mathbf{c} \\ &= \frac{N_r}{\log(2)} + \frac{1}{2M^{N_t}} \mathbf{c}^t (\sum_k \mathbf{V}_k) \mathbf{c} \\ &= \frac{N_r}{\log(2)} + \frac{1}{2M^{N_t}} \mathbf{c}^t \mathbf{V} \mathbf{c}, \end{aligned} \quad (37)$$

where $\mathbf{V} \doteq \sum_k \mathbf{V}_k$.

The procedure presented above for the $N_t = N_r = 2$ case can be directly generalized to any $N_t = N_r$ by following a recursive procedure as follows. We repeat the above recursive procedure until either $\mathbf{F}_{k,l}$ or $\mathbf{V}_{k,l}$ have included all the indices k_{rj}, k_{ij}. The procedure will end at a $\mathbf{V}_{k,l}$ if N_r is even and an $\mathbf{F}_{k,l}$ if $N - r$ is odd.

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 $\mathbf{HG}$,

$$I(\mathbf{x}; \mathbf{y}) \approx N_t \log_2(M) - \frac{N_r}{\log(2)} - \frac{1}{M^{N_t}} \sum_k \hat{f}_k. \quad (38)$$

In order to derive the gradient of $I(\mathbf{x};\mathbf{y})$ with respect to $\mathbf{W}$, we first derive the gradient of $I(\mathbf{x};\mathbf{y})$ with respect to $\mathbf{M}^*$. Start with the differential of $I(\mathbf{x};\mathbf{y})$ with respect to $\mathbf{M}^*$ in (38) and approximate the f_k by $\hat{f}_k$,

$$d_{\mathbf{M}^*} I(\mathbf{x};\mathbf{y}) \approx -\frac{1}{M^{N_t}} d_{\mathbf{M}^*}\left(\sum_k \hat{f}_k\right) = -\frac{1}{M^{N_t}}\left(\frac{1}{\pi}\right)^{N_r} \times \sum_{k_{r1}=1}^{L}\sum_{k_{i1}=1}^{L}\cdots\sum_{k_{rN_r}=1}^{L}\sum_{k_{iN_r}=1}^{L} c(k_{r1})c(k_{i1}) \cdots c(k_{rN_r}) \times c(k_{iN_r}) \sum_k d_{\mathbf{M}^*} \times \left(g_k(\sigma n_{k_{r1}}, \sigma n_{k_{i1}}, \cdots, \sigma n_{k_{rN_r}}, \sigma n_{k_{iN_r}})\right). \quad (39)$$

Taking into account that $g_k(\sigma n_{k_{r1}}, \sigma n_{k_{i1}}, \cdots, \sigma n_{k_{rN_r}}, \sigma n_{k_{iN_r}})$ is the value of

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

evaluated at $\mathbf{n}_e = \sigma\mathbf{v}(\{k_{rv},\ k_{iv}\}_{v=1}^{N_r})$, we can develop (39) further, by using well-known results [31], as follows

$$d_{\mathbf{M}^*} I(\mathbf{x};\mathbf{y}) \approx -\frac{1}{\log(2)}\frac{1}{M^{N_t}}\left(\frac{1}{\pi}\right)^{N_r} \sum_{k_{r1}=1}^{L}\sum_{k_{i1}=1}^{L} \cdots \times \sum_{k_{rN_r}=1}^{L}\sum_{k_{iN_r}=1}^{L} c(k_{r1})c(k_{i1})\cdots c(k_{rN_r})c(k_{iN_r}) \times \mathrm{tr}\left(\{\mathbf{R}(\sigma n_{r1,k_{r1}}, \sigma n_{i1,k_{i1}}, \cdots, \sigma n_{rN_r,k_{rN_r}}, \sigma n_{iN_r,k_{iN_r}})\}^t \times d\mathbf{M}^*\right), \quad (40)$$

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

$$\sum_k \frac{1}{\sum_l \exp(-\frac{1}{\sigma^2}||\mathbf{n}-\mathbf{M}(\mathbf{x}_k-\mathbf{x}_l)||^2)} \times \sum_m \left(\exp(-\frac{1}{\sigma^2}||\mathbf{n}-\mathbf{M}(\mathbf{x}_k-\mathbf{x}_m)||^2) \times \left((\mathbf{n}-\mathbf{M}(\mathbf{x}_k-\mathbf{x}_m))(\mathbf{x}_k-\mathbf{x}_m)^h\right)\right) \quad (41)$$

evaluated at $\mathbf{n}_e = \sigma\mathbf{v}(\{k_{rv},\ k_{iv}\}_{v=1}^{N_r})$. In this derivation we have used the fact that

$$d_{\mathbf{M}}\left(\exp(-\frac{||\mathbf{n}-\mathbf{M}(\mathbf{x}_k-\mathbf{x}_m)||^2}{\sigma^2})\right) = \frac{1}{\sigma^2}\times\exp(-\frac{||n-\mathbf{M}(\mathbf{x}_k-\mathbf{x}_m)||^2}{\sigma^2})(\mathbf{n}-\mathbf{M}(\mathbf{x}_k-\mathbf{x}_m)) \times (\mathbf{x}_k-\mathbf{x}_m)^h. \quad (42)$$

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 [31]), we get (19).

Now, since $\mathbf{W} = \mathbf{M}^2$, by taking vectors of the matrices on each side of this equation and applying some identities from [31], we get that

$$d\mathrm{vec}(\mathbf{W}) = \left((\mathbf{M}^*)\otimes\mathbf{I} + \mathbf{I}\otimes\mathbf{M}\right) d\mathrm{vec}(\mathbf{M}), \quad (43)$$

where the notation $\mathrm{vec}(\mathbf{A})$ denotes the vector found from matrix $\mathbf{A}$ by taking its columns one at a time, starting from the leftmost one. Since $\mathbf{M} = \mathbf{W}^{\frac{1}{2}}$, it is straightforward to derive (23) by employing (43) as follows. The relationship between the gradient and the derivative is through reshaping the derivative row vector [31], we can thus write

$$\nabla_{\mathbf{W}} I \approx \mathrm{reshape}((\mathrm{vec}(\nabla_{\mathbf{M}} I)^T \left((\mathbf{M}^*)\otimes\mathbf{I} + \mathbf{I}\otimes\mathbf{M}\right)^{-1}), N_t, N_t), \quad (44)$$

where $\mathrm{reshape}(\mathbf{A}, k, n)$ is the standard reshape of matrix $\mathbf{A}$ with total elements kn to a matrix with k rows, n columns.

Appendix C
Evaluation of Computational Complexity in Calculating $I(\mathbf{x};\mathbf{y})$ Through Gauss-Hermite Quadrature

Let's start with $N_r = 2$. Then, the complexity involved in the Gauss-Hermite quadrature approximation of $I(\mathbf{x};\mathbf{y})$ is determined by the one required in the calculation of $\hat{f}_k$ in (34). From (35), this is equal to $(2L-1)^2$ times the number of operations, including summations and products, required in evaluating each element of the matrix $\mathbf{V}_k$. Since $\mathbf{V}_k$ is a size $L\times L$ matrix with elements given in (36), it can be seen from (32) that the complexity of each element $\mathbf{V}_k[k_{r2}, k_{i2}]$ is $L^2N_r(2N_t+N_r-1)$. Since there are M^{N_t} summation terms over k (the size of the overall multiple input constellation), the total complexity becomes $M^{2N_t}(2L-1)^4L^4N_r(2N_t+1)$. For a general value of N_r, in a similar fashion, the corresponding complexity becomes $M^{2N_t}(2L-1)^{(2N_r)}L^4N_r(2N_t+N_r-1)$.

References

[1] K.-H. Lee and D. P. Petersen, "Optimal linear coding for vector channels," *IEEE Trans. Commun.*, vol. 24, no. 12, pp. 1283–1290, Dec. 1976.

[2] J. Salz, "Digital transmission over cross-coupled linear channels," *AT&T Tech. J.*, vol. 64, no. 6, pp. 1147–1159, Jul. 1985.

[3] J. Yang and S. Roy, "On joint transmitter and receiver optimization for multiple-input-multiple-output (MIMO) transmission systems," *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3221–3231, Dec. 1994.

[4] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers. I. Unification and optimal designs," *IEEE Trans. Signal Process.*, vol. 47, no. 7, pp. 1988–2006, Jul. 1999.

[5] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1051–1064, May 2002.

[6] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, Dec. 2001.

[7] J. Yang and S. Roy, "Joint transmitter-receiver optimization for multi-input multi-output systems with decision feedback," *IEEE Trans. Inf. Theory*, vol. 40, no. 5, pp. 1334–1347, Sep. 1994.

[8] H. Sampath and A. Paulraj, "Linear precoding for space-time coded systems with known fading correlations," *IEEE Commun. Lett.*, vol. 6, no. 6, pp. 239–241, Jun. 2002.

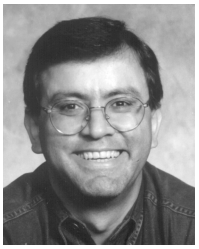
[9] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.

[10] V. Raghavan, A. M. Sayeed, and V. V. Veeravalli, "Semiunitary precoding for spatially correlated MIMO channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1284–1298, Mar. 2011.

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

[12] A. Lozano, A. M. Tulino, and S. Verdu, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3033–3051, Jul. 2006.
[13] S. K. Mohammed, E. Viterbo, Y. Hong, and A. Chockalingam, "Precoding by pairing subchannels to increase MIMO capacity with discrete input alphabets," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4156–4169, Jul. 2011.
[14] M. Payaro and D. P. Palomar, "On optimal precoding in linear vector Gaussian channels with arbitrary input distribution," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun./Jul. 2009, pp. 1085–1089.
[15] 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.*, Jun. 2011, pp. 346–350.
[16] C. Xiao, Y. R. 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.
[17] M. Lamarca, "Linear precoding for mutual information maximization in MIMO systems," in *Proc. Int. Symp. Wireless Commun. Syst.*, Sep. 2009, pp. 26–30.
[18] M. A. Girnyk, M. Vehkapera, and L. K. Rasmussen, "Large-system analysis of correlated MIMO multiple access channels with arbitrary signaling in the presence of interference," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 2060–2073, Apr. 2014.
[19] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.
[20] W. Weichselberger, M. Herdin, H. Ozcelik, and E. Bonek, "A stochastic MIMO channel model with joint correlation of both link ends," *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 90–100, Jan. 2006.
[21] 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 Trans. Wireless Commun.*, vol. 14, no. 2, pp. 983–997, Feb. 2015.
[22] Y. Wu, C.-K. Wen, D. W. K. Ng, R. Schober, and A. Lozano, "Low-complexity MIMO precoding with discrete signals and statistical CSI," in *Proc. ICC*, 2016, pp. 1–6.
[23] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. Washington, DC, USA: U.S. Government Printing Office, 1972.
[24] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[25] Y. Xin, Z. Wang, and G. B. Giannakis, "Space-time diversity systems based on linear constellation precoding," *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 294–309, Mar. 2003.
[26] T. Ketseoglou and E. Ayanoglu, "Linear precoding for MIMO with LDPC coding and reduced complexity," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2192–2204, Apr. 2015.
[27] J. Salo *et al. MATLAB Implementation of the 3GPP Spatial Channel Model*. [Online]. Available: http://www.tkk.fi/Units/Radio/scm/
[28] V. Raghavan, J. H. Kotecha, and A. M. Sayeed, "Why does the Kronecker model result in misleading capacity estimates?" *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4843–4864, Oct. 2010.
[29] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 670–678, Apr. 2004.
[30] Y. Jiang, A. Ashikhmin, 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.
[31] A. Hjørungnes, *Complex-Valued Matrix Derivatives: With Applications in Signal Processing and Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[32] W. Zeng, C. 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. 38–41, Feb. 2012.
[33] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.
[34] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
[35] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.
[36] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

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, including senior level positions with Siemens, Ericsson, Rockwell, and Omnipoint. From 1996 to 1998, he participated in TIA TR45.5 (now 3GPP2) 3G standardization, making significant contributions to the cdma2000 standard. He holds several essential patents in wireless communications. Since 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 with 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 with the University of California at 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 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 to 2002, he was a Systems Architect with Cisco Systems, Inc. Since 2002, he has been a Professor with the Department of Electrical Engineering and Computer Science, University of California at Irvine, CA, USA, where he served as the Director of the Center for Pervasive Communications and Computing and held the Conexant-Broadcom Endowed Chair from 2002 to 2010. He was a 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 to 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 IEEE TRANSACTIONS ON COMMUNICATIONS. 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.