Semiunitary Precoding for Spatially Correlated MIMO Channels

Vasanthan Raghavan, *Member, IEEE*, Akbar M. Sayeed, *Senior Member, IEEE*, and Venugopal V. Veeravalli, *Fellow, IEEE*

***Abstract*—The focus of this paper is on spatial precoding in correlated multiantenna channels where the number of data-streams is adapted independent of the number of transmit antennas. Towards the goal of a low-complexity implementation, a statistical semiunitary precoder is studied where the precoder matrix evolves fairly slowly with respect to the channel evolution. While prior work on statistical precoding has focussed on information-theoretic limits, most of these computations result in complicated functional dependencies of the mutual information with the channel statistics that do not explicitly reveal the impact of statistics on performance. In contrast, estimates that are directly in terms of the channel statistics are obtained here for the relative mutual information loss of a semiunitary precoder with respect to a perfect channel information benchmark. Based on these estimates, matching metrics are developed that capture the degree of matching of a channel to the precoder structure *continuously* and allow ordering two matrix channels in terms of their mutual information performance. While these metrics are based on bounds, numerical studies are used to show that the proposed metrics capture the performance tradeoffs accurately. The main conclusion of this work is a simple-to-state fundamental principle in the context of signaling design for single-user MIMO systems: the best channel for the statistical precoder is the channel that is matched to it.**

***Index Terms*—Adaptive coding, correlated channels, low-complexity signaling, MIMO systems, multimode signaling, semiunitary precoding, spatial precoding.**

I. Introduction

MULTIPLE antenna communications has received significant attention over the last decade as a mechanism to increase the rate of information transfer, or the reliability of signal reception, or a combination of the two. The focus of this work is on point-to-point spatial precoding systems where the number of independent data-streams is constrained to be a subset, M, of the transmit dimension N_t. Initial works on precoding study optimal signaling strategies when perfect channel state information (CSI) is available at the transmitter and the receiver. These studies show that a *channel diagonalizing* input that corresponds to exciting the dominant M-dimensional eigen-space of the channel, with a power allocation that can be computed via waterfilling, is robust under different design metrics [1]–[10].

Although perfect CSI provides a benchmark on the performance, it is difficult to obtain in practice. More importantly, the system performance is not robust under CSI uncertainty. Small perturbations in the channel entries could result in large perturbations in a singular vector of the channel if the discernibility of the corresponding singular value diminishes. Furthermore, even if perfect CSI is available, tight constraints on complexity as well as energy consumption [11]–[14, Ch. 5] at the RF level in the mobile ends may disallow the implementation of optimal solutions in practice. This is because Third Generation wireless systems and beyond are expected to be multicarrier in nature and the burden of computing the optimal input is magnified by the number of subcarriers and the rate of evolution of the channel realizations. Besides this, the structure of the input could change, often dramatically, at the rate of evolution of the channel realizations, which also makes it difficult to implement. These reasons suggest that a slower rate of adaptation of the input signals, that is of low complexity and is more robust to CSI uncertainty, is preferred in practice.

In realistic wireless systems, where the channels are spatiotemporally correlated, the slow rate of statistical evolution implies that it is reasonable to assume perfect statistical knowledge of the channel at the transmitter. Since the spatial statistics experienced by the individual subcarriers are identical [15], [16], the burden of computing the optimal input with only the statistical information at the transmitter is equivalent to that of a narrowband system. Even in this setting, optimal precoding has been studied for different spatial correlation models [16]–[27]. These works show that the eigen-directions of the optimal input covariance matrix correspond to a set of the M-dominant eigenvectors of the transmit covariance matrix and are hence, easily adaptable to change in statistics. However, computing the power allocation across the M modes requires Monte Carlo averaging or gradient descent-type iterative approaches [22]–[25]. While the computational complexity of the power allocation algorithm may be affordable at the base station end, whether it is possible or not at the mobile end is questionable.

Many of the above works have also leveraged tools from asymptotic random matrix theory and made significant progress in characterizing the information-theoretic limits in correlated

Manuscript received May 28, 2008; revised January 08, 2010; accepted July 22, 2010. Date of current version February 18, 2011. This work was supported in part by NSF Grant #CCF-0049089 through the University of Illinois and in part by Grant #CCF-0431088 through the University of Wisconsin. The material in this paper was presented in part at the 42nd Annual Allerton Conference on Communications, Control, and Computing, Allterton, IL, September 2006, and at the IEEE International Symposium on Information Theory, Toronto, ON, Canada, 2008.

V. Raghavan was with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA. He is now with the Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, VIC 3052, Australia (e-mail: vasanthan_raghavan@ieee.org).

A. M. Sayeed is with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA (e-mail: akbar@engr.wisc.edu).

V. V. Veeravalli is with the Coordinated Science Laboratory and the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: vvv@illinois.edu).

Communicated by A. Goldsmith, Associate Editor for Communications.

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/TIT.2010.2103810

MIMO channels. However, most of them rely on the implicit characterization of the limiting eigenvalue distribution of random matrices (given by the Stieltjes transformational formula [28], [29]) and obtain fixed-point equations which can be solved at any fixed SNR to produce asymptotic capacity formulas; see [22]–[25], [29], [30], and references therein. While this approach is valid in the antenna asymptotics for any fixed SNR, insights on the impact of the channel statistics (the transmit and receive covariance matrices) on capacity is rendered difficult due to the complicated nature of the fixed-point equations.

With this background in mind, we restrict our theoretical attention to the mutual information performance of a class of statistical *semiunitary*[1] precoders where the eigen-directions of the input correspond to the dominant eigenvectors of the transmit covariance matrix and the power allocation is uniform. Our focus here is on two questions: 1) can the performance of a semiunitary precoder be captured as a function of the channel statistics *transparently*, in contrast to existing implicit characterizations?, 2) when is the semiunitary precoder near-optimal with respect to a perfect CSI benchmark and what is the "gap"[2] in performance in terms of the system and the channel parameters?

Towards answering these questions, we use tools from asymptotic random matrix theory to bound the relative average loss in mutual information between the perfect CSI and statistical semiunitary precoders. These bounds are *transparent* and in terms of the eigenvalues of the transmit and receive covariance matrices. Motivated by these bounds, we introduce the notion of *matching metrics* that abstractly capture the degree of channel-to-precoder matching. On one extreme is a perfectly matched channel where: 1) the M-dominant eigenvalues of the transmit covariance matrix are *well-conditioned*[3] whereas the remaining $(N_t - M)$ eigenvalues are *ill-conditioned* away from the dominant ones, and 2) the receive covariance matrix is also *well-conditioned*. On the other extreme is a perfectly mismatched channel where both the transmit and receive covariance matrices are ill-conditioned with the additional constraint that $\mathsf{rank}(\mathbf{H}) \geq M$ with probability 1.

Our work establishes the following simple-to-state fundamental principle, akin to existing source-channel matching paradigms, in the context of signaling design for single-user MIMO systems. While there exists no metric for ordering two matrices [31], multiantenna channel matrices can be ordered continuously with respect to their average mutual information performance with a semiunitary precoder of a fixed rank using the matching metrics. In particular, the two extreme cases of channels (as above) correspond to the setting where the mutual information of the semiunitary precoder is closest and farthest to the perfect CSI precoder, respectively. While the matching metrics have been defined based on bounds and these bounds have only been established under certain special assumptions (antenna asymptotics and high SNR), we provide numerical studies to show that the matching metrics capture the performance tradeoffs accurately for all SNRs and even small antenna numbers.

Despite the growing importance of statistical (semiunitary) precoding in wireless standardization efforts, a comprehensive study of the performance limits of statistical precoding is lacking in the literature and the channel-to-precoder matching principle established here provides some intuition on what type of precoder is best suited to a specific channel statistics.

A. Organization

After elucidating the system model in Section II, we benchmark the structure of the optimal precoder with perfect CSI and only statistical knowledge at the transmitter in Section III. We also motivate the need to study statistical semiunitary precoding in this section. In Section IV and the appendices, using tools from random matrix theory and eigenvector perturbation theory, we study the asymptotic (in antenna dimensions) performance of a statistical semiunitary precoder. We discuss the implications of our results and illustrate them numerically in Section V. Concluding remarks are provided in Section VI.

B. Notation

The M-dimensional identity matrix is denoted by $\mathbf{I}_M$. The i,j-th and i-th diagonal entries of a matrix $\mathbf{X}$ are denoted by $\mathbf{X}(i,j)$ and $\mathbf{X}(i)$, respectively. In more complicated settings (for example, when the matrix $\mathbf{X}$ is represented as a product or sum of many matrices), the above entries are denoted by $\mathbf{X}_{ij}$ and $\mathbf{X}_i$, respectively. The complex conjugate, conjugate transpose, and inverse operations are denoted by $(\cdot)^\star$, $(\cdot)^H$, and $(\cdot)^{-1}$ while the expectation, the trace and the determinant operators are given by $E[\cdot]$, $\mathsf{Tr}(\cdot)$ and $\det(\cdot)$, respectively. The standard big-Oh ($\mathcal{O}$) and little-oh (o) notations are used along with the decreasing ordering for eigenvalues of an $n \times n$ Hermitian matrix $\mathbf{X}$: $\lambda_1(\mathbf{X}) \geq \cdots \geq \lambda_n(\mathbf{X})$. The largest and the smallest eigenvalues are also denoted by $\lambda_{\max}(\mathbf{X})$ and $\lambda_{\min}(\mathbf{X})$, respectively. The notation x^+ stands for $\max(x, 0)$. All logarithms are to base e unless mentioned otherwise.

II. System Setup

We consider a communication system with N_t transmit and N_r receive antennas where M $(1 \leq M \leq N_t)$ independent data-streams are used in signaling. That is, the M-dimensional input vector $\mathbf{s}$ is precoded into an N_t-dimensional vector via the $N_t \times M$ precoding matrix $\mathbf{F}$ and transmitted over the channel. With a transmit power constraint of ρ, the discrete-time baseband signal model used is

$$\mathbf{y} = \sqrt{\frac{\rho}{M}}\mathbf{HFs} + \mathbf{n} \tag{1}$$

where $\mathbf{y}$ is the N_r-dimensional received vector, $\mathbf{H}$ is the $N_r \times N_t$-dimensional channel matrix, and $\mathbf{n}$ is the N_r-dimensional (zero mean, unit variance) additive white Gaussian noise. The most general decomposition of the precoder is

$$\mathbf{F} = \mathbf{V_F}\mathbf{\Lambda}_{\mathbf{F}}^{1/2}\mathbf{U}_{\mathbf{F}}^H \tag{2}$$

[1]An $N_t \times M$ matrix $\mathbf{X}$ with $M \leq N_t$ is said to be semiunitary if it satisfies $\mathbf{X}^H\mathbf{X} = \mathbf{I}_M$.

[2]This gap can possibly be bridged with a *limited feedback* scheme that provides partial channel information to the transmitter.

[3]If $\mathbf{\Lambda}_t(1) \geq \cdots \geq \mathbf{\Lambda}_t(M)$ denote the first M eigenvalues of the transmit covariance matrix and $\frac{\mathbf{\Lambda}_t(1)}{\mathbf{\Lambda}_t(M)}$ is (or is not) significantly larger than 1, we loosely say that these eigenvalues are ill-(or well-)conditioned.

where $\mathbf{V_F}$ is $N_t \times M$ semiunitary, $\mathbf{\Lambda_F}$ is an $M \times M$ nonnegative definite power shaping (allocation) matrix, and $\mathbf{U_F}$ is $M \times M$ unitary. Under the assumption that $\mathbf{s}$ has i.i.d. components with zero mean and unit variance, the transmit power constraint is met with $\mathsf{Tr}(\mathbf{\Lambda_F}) \leq M$.

A. Channel Model

In this work, we make the reasonable assumption that the receiver has perfect CSI. The main emphasis here is on the impact of transmitter knowledge of statistics of the channel process on performance. We assume a block fading, narrowband model for the time-frequency correlation of $\mathbf{H}$ and focus on the spatial correlation. It is well-known that Rayleigh fading (zero mean complex Gaussian) is an accurate model for $\mathbf{H}$ in a non line-of-sight setting and hence, the complete spatial statistics are described by the second-order moments of $\{\mathbf{H}(i,j)\}$.

The most general, mathematically tractable spatial correlation model is a *canonical decomposition*[4] of the channel along the transmit and receive covariance bases [24], [26], [32]. In this model, we assume that the auto- and cross-covariance matrices of all rows of $\mathbf{H}$ have the same unitary eigen-basis (denoted by $\mathbf{U}_t$), and the auto- and cross-covariance matrices of all the columns of $\mathbf{H}$ have the same unitary eigen-basis ($\mathbf{U}_r$). Thus, we can decompose $\mathbf{H}$ as

$$\mathbf{H} = \mathbf{U}_r \mathbf{H}_{\mathsf{ind}} \mathbf{U}_t^H \tag{3}$$

where $\mathbf{H}_{\mathsf{ind}}$ has independent, but not necessarily identically distributed entries. The transmit and receive covariance matrices are defined as

$$\mathbf{\Sigma}_t \triangleq E[\mathbf{H}^H\mathbf{H}] = \mathbf{U}_t E[\mathbf{H}_{\mathsf{ind}}^H \mathbf{H}_{\mathsf{ind}}]\mathbf{U}_t^H = \mathbf{U}_t \mathbf{\Lambda}_t \mathbf{U}_t^H \tag{4}$$
$$\mathbf{\Sigma}_r \triangleq E[\mathbf{H}\mathbf{H}^H] = \mathbf{U}_r E[\mathbf{H}_{\mathsf{ind}} \mathbf{H}_{\mathsf{ind}}^H]\mathbf{U}_r^H = \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{U}_r^H \tag{5}$$

where $\mathbf{\Lambda}_t = E[\mathbf{H}_{\mathsf{ind}}^H \mathbf{H}_{\mathsf{ind}}]$ and $\mathbf{\Lambda}_r = E[\mathbf{H}_{\mathsf{ind}} \mathbf{H}_{\mathsf{ind}}^H]$ are diagonal. Note that the eigenvalues of the transmit covariance matrix are

$$\left\{ \sum_{i=1}^{N_r} \sigma_{ik}^2, k = 1, \ldots, N_t \right\} \tag{6}$$

where σ_{ij}^2 denotes the variance of $\mathbf{H}_{\mathsf{ind}}(i,j)$. Given a correlated channel, we will assume that $M \leq \mathsf{rank}(\mathbf{\Lambda}_t) \leq N_t$. We will also assume that the columns of $\mathbf{H}_{\mathsf{ind}}$ are arranged in the decreasing order of transmit eigenvalues.

Under certain conditions, the model in (3) reduces to some well-known spatial correlation models such as the i.i.d. model, the separable correlation [33] and the virtual representation [15], [23] frameworks. For example, in the separable case, under the normalization that

$$\mathsf{Tr}(\mathbf{\Lambda}_t) = \mathsf{Tr}(\mathbf{\Lambda}_r) = \rho_c = N_t N_r, \tag{7}$$

we can write $\mathbf{H}_{\mathsf{ind}}$ for the normalized channel as

$$\mathbf{H} = \frac{1}{\sqrt{\rho_c}} \cdot \mathbf{\Sigma}_r^{1/2} \mathbf{H}_{\mathsf{iid}} \mathbf{\Sigma}_t^{1/2} \tag{8}$$

$$\Longrightarrow \mathbf{H} \sim \frac{1}{\sqrt{\rho_c}} \cdot \mathbf{U}_r \mathbf{\Lambda}_r^{1/2} \mathbf{H}_{\mathsf{iid}} \mathbf{\Lambda}_t^{1/2} \mathbf{U}_t^H \tag{9}$$

$$\Longrightarrow \mathbf{H}_{\mathsf{ind}} = \frac{1}{\sqrt{\rho_c}} \cdot \mathbf{\Lambda}_r^{1/2} \mathbf{H}_{\mathsf{iid}} \mathbf{\Lambda}_t^{1/2} \tag{10}$$

where $\mathbf{H}_{\mathsf{iid}}$ is an i.i.d. channel matrix and the correlation of the channel entries is in the form of a Kronecker product of the transmit and receive covariance matrices. Even though the separable model may be an accurate fit under certain channel conditions, deficiencies acquired by the separability property result in misleading estimates of system performance [26]. The readers are referred to [26] and [32] for more details on how the general (nonseparable) version of the canonical model fits measured data better.

B. Receiver Architecture

Under these assumptions, the optimal reception strategy corresponds to nonlinear maximum likelihood (ML) decoding. However, the exponential complexity of ML decoding in both antenna dimensions and coherence length implies that simpler receiver architectures are preferred. In this work, we assume a linear minimum mean-squared error (MMSE) receiver. With this receiver, the symbol corresponding to the k-th data-stream is recovered by projecting the received signal $\mathbf{y}$ on to the $N_r \times 1$ vector

$$\mathbf{g}_k = \sqrt{\frac{\rho}{M}} \left(\frac{\rho}{M} \mathbf{H}\mathbf{F}\mathbf{F}^H\mathbf{H}^H + \mathbf{I}_{N_r} \right)^{-1} \mathbf{H}\mathbf{f}_k \tag{11}$$

where $\mathbf{f}_k$ is the k-th column of $\mathbf{F}$. That is, the recovered symbol is $\widehat{\mathbf{s}}(k) = \mathbf{g}_k^H \mathbf{y}$, and the mean-squared error of this recovery process, MSE_k, is given by

$$\mathsf{MSE}_k = \left[\left(\mathbf{I}_M + \frac{\rho}{M} \mathbf{F}^H \mathbf{H}^H \mathbf{H} \mathbf{F} \right)^{-1} \right]_k. \tag{12}$$

III. Preliminaries

We first summarize known results on optimal precoder design in this section before proceeding onto the focus of this paper.

The metric of interest in this work is the mutual information between the input and output symbols since it captures both the achievable rate as well as reliability performance under a concatenated inner and outer code design [34] (where soft decisions are allowed at the decoder of the inner code). Under the assumption that the input symbols are Gaussian, the mutual information at an SNR (of ρ) is given as

$$I(\mathbf{s};\mathbf{y}) = \log\det\left(\mathbf{I}_M + \frac{\rho}{M} \mathbf{F}^H \mathbf{H}^H \mathbf{H} \mathbf{F} \right). \tag{13}$$

It can be seen that maximizing the mutual information in (13) can be formulated as the minimization of a Schur-concave function: the determinant of the mean-squared error matrix [9].

A. Perfect CSI Case

A unified convex programming framework for precoder optimization in the perfect CSI case, summarized in the following lemma, is proposed in [9] by studying two broad classes of functions: Schur-concave and Schur-convex functions.

[4]This model is referred to as the "eigen-beam or beamspace model" in [32] and is used in capacity analysis in [24].

Lemma 1: Let $f : \mathbb{R}^M \mapsto \mathbb{R}$ be a function such that $f(\cdot)$ is monotonically increasing in its arguments. That is, let the univariate function $f(\ldots, x_k, \ldots) : \mathbb{R} \mapsto \mathbb{R}$ be monotonically increasing for all k. If $\mathsf{MSE} = [\mathsf{MSE}_1 \cdots \mathsf{MSE}_M]$ and $f(\cdot)$ is Schur-concave over its domain, then $f(\mathsf{MSE})$ is minimized by $\mathbf{F}_{\mathsf{perf}}$ whose singular value decomposition (SVD) is given as

$$\mathbf{F}_{\mathsf{perf}} = [\mathbf{v}_1 \cdots \mathbf{v}_M] \cdot \mathbf{\Lambda}_{\mathsf{perf}}^{1/2}. \tag{14}$$

On the other hand, if $f(\cdot)$ is Schur-convex, $f(\mathsf{MSE})$ is minimized by

$$\mathbf{F}_{\mathsf{perf}} = [\mathbf{v}_1 \cdots \mathbf{v}_M] \cdot \mathbf{\Lambda}_{\mathsf{perf}}^{1/2} \cdot \mathbf{\Gamma} \tag{15}$$

for an appropriate choice of unitary matrix $\mathbf{\Gamma}$ (see [9] for its construction). In both cases, the diagonal entries of $\mathbf{\Lambda}_{\mathsf{perf}}$ are obtained via waterfilling and we assume a SVD for $\mathbf{H}$ as

$$\mathbf{H} = \mathbf{U}_{\mathbf{H}} \mathbf{\Lambda}_{\mathbf{H}}^{1/2} \mathbf{V}_{\mathbf{H}}^H, \;\; \mathbf{V}_{\mathbf{H}} = [\mathbf{v}_1 \cdots \mathbf{v}_{N_t}] \tag{16}$$

and the singular values are arranged in decreasing order. ■

Specific instantiations of the above lemma have been studied in the cases of average mean-squared error of the data-streams [1]–[4], weighted average of mean-squared error of the data-streams [5], [6], determinant of the mean-squared error matrix [7], determinant under a peak-power constraint [8], and bit-error rate [9], [10].

Lemma 2: Using the ideas of [9] and [31], Lemma 1 can be straightforwardly extended to the case of perfect CSI semiunitary precoding, where $\mathbf{\Lambda}_{\mathbf{F}}$ in (2) is constrained to be $\mathbf{\Lambda}_{\mathbf{F}} = \mathbf{I}_M$. If $f(\cdot)$ is Schur-concave over its domain, then $f(\mathsf{MSE})$ is minimized by

$$\mathbf{F}_{\mathsf{perf,semi}} = [\mathbf{v}_1 \cdots \mathbf{v}_M]. \tag{17}$$

On the other hand, if $f(\cdot)$ is Schur-convex, $f(\mathsf{MSE})$ is minimized by

$$\mathbf{F}_{\mathsf{perf,semi}} = [\mathbf{v}_1 \cdots \mathbf{v}_M] \cdot \mathbf{\Gamma} \tag{18}$$

for an appropriate choice of unitary matrix $\mathbf{\Gamma}$ (same as in the perfect CSI case). In fact, Lemma 1 can be extended to the case where $\mathbf{\Lambda}_{\mathbf{F}}$ is fixed (but is different from $\mathbf{I}_M$) by using the notion of weak super-majorization from [31]. The details are not provided here. ■

B. Statistical Case

Following Lemmas 1 and 2, since the eigen-modes of the optimal input are a function of the CSI, performance degradation with respect to CSI error is directly related to singular vector perturbations of the channel matrix. While it is true that a small perturbation in the matrix entries can only lead to a small perturbation in the singular values, a small entry-wise perturbation can result in a *large* perturbation of the singular vectors depending on the condition number of the true channel matrix [35, pp. 202–203],[36], [37]. See, for example, [38]–[40, Figs. 6 and 7] etc. that illustrate MIMO settings where losses equivalent to a 25 dB SNR penalty occur due to lack of perfect CSI.

On the other hand, it may not be possible to adapt the precoder structure to the channel optimally even if perfect CSI is available since RF design is often the fundamental bottleneck for realizing MIMO systems in practice [14, Ch. 5]. This may be because: 1) the eigenspace of the optimal input could change dramatically from one channel realization to the next, and/or 2) the efficient utilization of CSI is constrained by fundamental limits on energy per bit constraints at the computational or processing level [11]–[14]. For example, the move towards multicarrier signaling and the fast rate at which channel realizations evolve leads to computational limits on how many SVD operations can be afforded. These reasons suggest that statistical precoding where the optimal input is adapted in response to the statistical information, which evolves slowly compared with the channel realizations, is of importance. In this setting, the following lemma considers the mutual information maximization problem.

Lemma 3: Let $\mathbf{H}$ be described by the statistical model in (3) with the eigenvalues of $\mathbf{\Sigma}_t$ arranged in the decreasing order. Let $\widetilde{\mathbf{H}}_{\mathsf{ind}}$ denote the $N_r \times M$ principal submatrix of $\mathbf{H}_{\mathsf{ind}}$. The optimal precoder that maximizes the average mutual information is of the form

$$\mathbf{F}_{\mathsf{stat}} = \mathbf{V}_{\mathsf{stat}} \mathbf{\Lambda}_{\mathsf{stat}}^{1/2} \tag{19}$$

where $\mathbf{V}_{\mathsf{stat}}$ is a set of M-dominant eigenvectors of $\mathbf{\Sigma}_t$ and $\mathbf{\Lambda}_{\mathsf{stat}}$ is the unique solution to the following constrained optimization problem:

$$\mathbf{\Lambda}_{\mathsf{stat}} = \arg\max_{\mathbf{\Lambda} \in \mathcal{L}} E_{\mathbf{H}} \left[\log\det \left(\mathbf{I}_{N_r} + \frac{\rho}{M} \widetilde{\mathbf{H}}_{\mathsf{ind}} \mathbf{\Lambda} \widetilde{\mathbf{H}}_{\mathsf{ind}}^H \right) \right] \tag{20}$$

with $\mathcal{L}$ denoting the convex set of all diagonal $M \times M$ nonnegative definite matrices $\mathbf{\Lambda}$ such that $\mathsf{Tr}(\mathbf{\Lambda}) \leq M$. ■

The optimality of the dominant eigenvectors of $\mathbf{\Sigma}_t$ is not surprising; see [17]–[20], [22]–[25], and references therein for problems of a similar nature. The optimization in (20) is standard: maximizing a concave function over a convex set. A gradient descent-type approach for this is provided in [27] and Monte Carlo approaches are provided in [23] and [24].

C. Statistical Semiunitary Precoder

While Lemma 3 establishes the benchmark in the statistical case, computational constraints (as in the perfect CSI case) of Monte Carlo/gradient descent approaches could often make the computation of $\mathbf{\Lambda}_{\mathsf{stat}}$ hard, if not impossible. This motivates studying a low-complexity alternative of statistical semiunitary precoding:

$$\mathbf{F}_{\mathsf{stat,semi}} = \mathbf{V}_{\mathsf{stat}} \tag{21}$$

where $\mathbf{V}_{\mathsf{stat}}$ corresponds to the optimal choice of eigen-modes from Lemma 3.

Let I_{perf} and $I_{\mathsf{stat,semi}}$ denote the mutual information (random variables) achievable with $\mathbf{F}_{\mathsf{perf}}$ and $\mathbf{F}_{\mathsf{stat,semi}}$, respectively. The main goal of this paper is to compare the performance of a statistical semiunitary precoder with respect to its perfect CSI benchmark. In particular, we would like to estimate ΔI_{semi}, defined as,

$$\Delta I_{\mathsf{semi}} \triangleq \frac{E_{\mathbf{H}} \left[I_{\mathsf{perf}} - I_{\mathsf{stat,semi}} \right]}{E_{\mathbf{H}} \left[I_{\mathsf{stat,semi}} \right]}. \tag{22}$$

The reason for considering a normalized quantity in (22) in contrast to $E_{\mathbf{H}}\left[I_{\sf perf} - I_{\sf stat,semi}\right]$ is the following. For any signaling scheme, the mutual information tends to zero as $\rho \rightarrow 0$ and tends to infinity as $\rho \rightarrow \infty$. Thus, the difference in mutual information between two schemes can converge to zero as $\rho \rightarrow 0$ at a rate different from that of either scheme, and/or could blow up to infinity as $\rho \rightarrow \infty$. In this setting, a more meaningful metric would be the relative difference in mutual information between these schemes.

It is clear that $\Delta I_{\sf semi}$ is a complicated function of the SNR, channel statistics and antenna dimensions, and a general closed-form expression seems hard. To simplify further analysis, we will assume that the SNR as well as the antenna dimensions are large. In particular, we will assume that $\rho \geq \alpha \frac{M}{\mathbf{\Lambda}_t(M)}$ for some suitable $\alpha > 1$. With respect to asymptotics of antenna dimensions, four cases arise based on the correlation structure in (3) and how antenna dimensions go to infinity: i) separable correlation with $\frac{M}{N_r} \rightarrow 0$ or ∞, ii) nonseparable correlation with $\frac{M}{N_r} \rightarrow 0$ or ∞, iii) separable correlation with $\frac{M}{N_r} \rightarrow \gamma \in (0,\infty)$, and iv) nonseparable correlation with $\frac{M}{N_r} \rightarrow \gamma \in (0,\infty)$. The first two cases denote the setting of relative antenna asymptotics, where one antenna dimension increases to infinity relative to the other. The last two correspond to the case where antenna dimensions grow in *proportion*.

IV. Mutual Information Loss With Semiunitary Precoding

The difference $\Delta I_{\sf semi}$ in (22) can be expanded as

$$\Delta I_{\sf semi} = \underbrace{\frac{E_{\mathbf{H}}\left[I_{\sf perf} - I_{\sf perf,semi}\right]}{E_{\mathbf{H}}\left[I_{\sf stat,semi}\right]}}_{\Delta I_1} + \underbrace{\frac{E_{\mathbf{H}}\left[I_{\sf perf,semi} - I_{\sf stat,semi}\right]}{E_{\mathbf{H}}\left[I_{\sf stat,semi}\right]}}_{\Delta I_2} \quad (23)$$

where $I_{\sf perf,semi}$ denotes the mutual information achievable with $\mathbf{F}_{\sf perf,semi}$. Since the argument within the expectation of the numerator of ΔI_1 is not explicitly dependent on the spatial correlation model, it is straightforward to obtain a bound for ΔI_1 in the high SNR regime.

Proposition 1: Let $\mathbf{\Lambda}_{\mathbf{H}}(M) = \lambda_M(\mathbf{H}^H\mathbf{H})$ denote the Mth largest squared singular value of $\mathbf{H}$ as in (16). If ρ is such that $\rho \geq \alpha E_{\mathbf{H}}\left[\frac{M}{\mathbf{\Lambda}_{\mathbf{H}}(M)}\right]$ for some $\alpha > 1$, ΔI_1 is bounded as

$$\Delta I_1 \leq \frac{2M}{\alpha^2 E_{\mathbf{H}}\left[I_{\sf stat,semi}\right]} \cdot \frac{E_{\mathbf{H}}\left[\left(\frac{1}{\mathbf{\Lambda}_{\mathbf{H}}(M)}\right)^2\right]}{\left(E_{\mathbf{H}}\left[\frac{1}{\mathbf{\Lambda}_{\mathbf{H}}(M)}\right]\right)^2}. \quad (24)$$

Proof: See Appendix B. ■

Intuitively, as α and, hence, the SNR increases, the waterfilling power allocation of the perfect CSI scheme converges to uniform power allocation across the M modes (see [22], [23], [25], etc.) and thus, ΔI_1 decreases. The bound provided in (24) is not tight since we have not characterized the exact probability $\Pr(n_{\mathbf{H}} < M)$ (in Appendix B) that determines ΔI_1. But the above bound is sufficient to capture the performance loss with uniform power allocation. Characterization of ΔI_2, which is explicitly dependent on the spatial correlation model, is nontrivial. In the following section, we provide estimates of ΔI_2 for different correlation models and regimes.

A. Relative Antenna Asymptotics

We start with the simplest case of separable correlation.

Theorem 1: Let the channel $\mathbf{H}$ be described by the normalized separable model as in (8)–(10). Let the columns of $\mathbf{H}_{\sf iid}$ be ordered such that the eigenvalues of $\mathbf{\Lambda}_t$ are in decreasing order. For any fixed value of ρ and under the assumption of $\frac{N_t}{N_r} \rightarrow 0$, ΔI_2 is bounded as

$$\Delta I_2 \leq \kappa_1 \cdot \frac{\sqrt{\sum_{i=1}^{N_r}(\mathbf{\Lambda}_r(i))^2}}{\sum_{i=1}^{N_r}\mathbf{\Lambda}_r(i)} \cdot \frac{M}{\sum_{i=1}^{M}\log\left(1+\frac{\rho}{M}\mathbf{\Lambda}_t(i)\right)} \quad (25)$$

where κ_1 is a constant determined from an application of Lemma 6 (in Appendix A).

Proof: See Appendix C. ■

As seen from Appendix C, ΔI_2 is a function of only $\lambda_k\left(\mathbf{\Lambda}_t\mathbf{H}_{\sf iid}^H\mathbf{\Lambda}_r\mathbf{H}_{\sf iid}\right)$ and $\lambda_k\left(\widetilde{\mathbf{\Lambda}}_t\widetilde{\mathbf{H}}_{\sf iid}^H\mathbf{\Lambda}_r\widetilde{\mathbf{H}}_{\sf iid}\right)$. Since $\lambda(\mathbf{AB}) = \lambda(\mathbf{BA})$, Theorem 1 can be easily modified even when $\frac{M}{N_r} \rightarrow \infty$. Hence, this case will not be studied in considerable detail. We now consider the nonseparable case with $\frac{N_t}{N_r} \rightarrow 0$.

Theorem 2: Let $\mathbf{H}$ be described by the general model in (3) and let σ_{ij}^2 denote the variance of $\mathbf{H}_{\sf ind}(i,j)$ with the assumption that

$$\frac{\sum_{i=1}^{N_r}\sigma_{ij}^2}{N_r} = \mathcal{O}(1) \ \text{ for all } \ j = 1,\ldots,M. \quad (26)$$

There exists a constant κ_2 determined from an application of Lemma 6 (in Appendix A) such that

$$\Delta I_2 \leq \kappa_2 \cdot \sqrt{\frac{N_t}{N_r}} \cdot \sum_{j=1}^{M}\frac{\rho N_r}{M+\rho\sum_i\sigma_{ij}^2} \cdot \frac{1}{\sum_{j=1}^{M}\log\left(1+\frac{\rho}{M}\cdot\sum_i\sigma_{ij}^2\right)}. \quad (27)$$

■

The proof of Theorem 2 follows along the approach of Theorem 1 via the generalized asymptotic eigenvalue characterization in Lemma 6. Observe that ΔI_2 in both (25) and (27) converges to zero as SNR increases as $\frac{1}{\log(\sf SNR)}$. In terms of the asymptotic trend as antenna dimensions increase, since $\sum_i \mathbf{\Lambda}_r(i) = \rho_c = N_t N_r$, the typical behavior of $\mathbf{\Lambda}_r(i)$ is $\mathbf{\Lambda}_r(i) = \mathcal{O}(N_t)$, which implies that

$$\sqrt{\sum_i(\mathbf{\Lambda}_r(i))^2} = \mathcal{O}(N_t\sqrt{N_r})$$

$$\Longrightarrow \frac{\sqrt{\sum_{i=1}^{N_r}(\mathbf{\Lambda}_r(i))^2}}{\sum_{i=1}^{N_r}\mathbf{\Lambda}_r(i)} = \mathcal{O}\left(\frac{N_t}{\sqrt{N_r}}\right) \quad (28)$$

which is essentially the same trend as (27).

B. Special Case: Beamforming

We now pay attention to the beamforming case ($M = 1$), the low-complexity of which makes it an attractive signaling choice in many wireless standards. While the SNR regime where beamforming is capacity-optimal has been established in prior work [22], [23], [25], [41], the performance gap between statistical and perfect CSI beamforming is less clear. Using tools from eigenvector perturbation theory, introduced in [40], we establish the following result.

First, note that the term ΔI_1 is redundant in the beamforming case. Let I_{perf} and I_{stat} denote the mutual information achievable by beamforming with perfect CSI and statistical information alone, respectively. Define the loss term

$$\Delta I_{\mathsf{bf}} \triangleq \frac{E_{\mathbf{H}}\left[I_{\mathsf{perf}} - I_{\mathsf{stat}}\right]}{E_{\mathbf{H}}\left[I_{\mathsf{stat}}\right]}. \tag{29}$$

The following discussion complements recent work on the performance gap with the separable model [42], that has been established by exploiting some recent advances in random matrix theory. Unlike [42] which is based on exact random matrix theory results and is applicable only for $E_{\mathbf{H}}\left[I_{\mathsf{perf}} - I_{\mathsf{stat}}\right]$ in the separable case, we generalize the results to the general canonical modeling framework, but do not consider fine refinement of constants in the following result for the sake of brevity.

Proposition 2: In the regime where $\frac{N_t}{N_r} \rightarrow 0$, ΔI_{bf} can be bounded as

$$\Delta I_{\mathsf{bf}} \leq \kappa_{\mathsf{bf}} \cdot \frac{N_t \cdot \log(N_r)}{N_r - N_t} \cdot \frac{1}{\log\left(1 + \rho N_r\right)} \tag{30}$$

where κ_{bf} is a constant that depends only on the eigenvalues of $\mathbf{\Sigma}_t$ and $\mathbf{\Sigma}_r$.

Proof: See Appendix D. ■

Note that the trend of ΔI_2 in (30) is similar to that of (25) and (27) in terms of SNR behavior, whereas in terms of trend as antenna dimensions increase, we are able to leverage eigenvector perturbation theory to obtain a tighter bound, in contrast with the earlier discussion.

C. Proportional Growth of Antenna Dimensions

We now consider the more complicated asymptotic setting where $\{M, N_r\} \rightarrow \infty$ with $\frac{M}{N_r} \rightarrow \gamma$ and $\gamma \in (0, \infty)$.

Theorem 3: Let the channel $\mathbf{H}$ be characterized by the normalized separable model. Also, let $A \triangleq \frac{N_t N_r}{M^2} = \mathcal{O}(1)$ and $B \triangleq \frac{M}{\mathbf{\Lambda}_r(M)} = \mathcal{O}(1)$. Let $G_{M,\bullet}$ denote the geometric means of the statistical eigenvalues, defined as,

$$G_{M,\mathsf{tx}} \triangleq \left(\prod_{i=1}^{M} \mathbf{\Lambda}_t(i)\right)^{1/M}, \; G_{M,\mathsf{rx}} \triangleq \left(\prod_{i=1}^{M} \mathbf{\Lambda}_r(i)\right)^{1/M}. \tag{31}$$

If $\rho = \alpha \cdot \frac{M}{\mathbf{\Lambda}_t(M)}$ for some $\alpha > 1$ and X is defined as

$$X \triangleq 1 - \frac{\sqrt{AB} \cdot \sqrt{AB + 4\alpha}}{2\alpha}, \tag{32}$$

ΔI_2 is bounded as

$$\Delta I_2 \leq \frac{\log\left(e/M\right) + \kappa_3}{\log(\rho/e\rho_c) + \log\left(G_{M,\mathsf{tx}} \cdot G_{M,\mathsf{rx}} \cdot X\right)} \tag{33}$$

$$\kappa_3 = \kappa_3' + \log\left(\frac{\min\{\mathbf{\Lambda}_t(1), \mathbf{\Lambda}_r(1)\}}{G_{M,\mathsf{tx}} \cdot G_{M,\mathsf{rx}} \cdot X}\right) \tag{34}$$

where κ_3' is a constant dependent only on the antenna dimensions.

Proof: See Appendix E. ■

In the general case of nonseparable correlation, bounding ΔI_2 is difficult due to the lack of a fundamental random matrix theory of spectral properties of random matrices with independent entries. As a result, unlike the earlier cases, we have to resort to approximations for ΔI_2.

Proposition 3: Let the channel be characterized by the nonseparable model with $\frac{M}{N_r} \rightarrow \gamma$ and $\gamma \in (0, \infty)$. Let $\delta > 0$ be a constant (appropriately small). Then, the following approximation to an upper bound of ΔI_2 holds with high probability (which converges to 1 as $\delta \rightarrow 0$):

$$\Delta I_2 \leq \Delta I_2^{\mathsf{UB}} \tag{35}$$

$$\approx \frac{\log\left(\frac{N_r e}{M}\right) + \frac{1}{M}\sum_{i=1}^{M} \log\left(1 + \frac{\delta(M-1)N_r}{\mathbf{\Lambda}_t(i)}\right)}{\log\left(\frac{\rho}{N_r e}\right) + \frac{1}{M}\log\left(\prod_{i=1}^{M} \mathbf{\Lambda}_t(i)\right)}. \tag{36}$$

Proof: See Appendix F. ■

Since

$$\sum_i \mathbf{\Lambda}_r(i) = \sum_i \mathbf{\Lambda}_t(i) = \rho_c = N_t N_r \tag{37}$$

the typical behavior of $G_{M,\mathsf{tx}}$ and $G_{M,\mathsf{rx}}$ is

$$\mathcal{O}(G_{M,\mathsf{tx}}) = \mathcal{O}(G_{M,\mathsf{rx}}) = \mathcal{O}(N_t) = \mathcal{O}(N_r). \tag{38}$$

Thus, typically, both (33) and (36) are symmetric with

$$\Delta I_2 \overset{\mathsf{SNR} \rightarrow \infty}{=} \mathcal{O}\left(\frac{1}{\log(\mathsf{SNR})}\right) \text{ and } \tag{39}$$

$$\Delta I_2 \overset{\{M, N_t, N_r\} \rightarrow \infty}{=} \mathcal{O}\left(\frac{1}{\log(N_t)}\right) = \mathcal{O}\left(\frac{1}{\log(N_r)}\right). \tag{40}$$

Also, note that while (33) and (36) are asymmetric in the sense that (33) is a function of $G_{M,\mathsf{rx}}$ whereas (36) is not. This is a deficiency of the approximation technique in the most general case and not of the trend exhibited by the tightest bound possible for ΔI_2.

Comparing the bounds between the relative antenna asymptotic and the proportional growth settings, the only difference is that $\Delta I_2 = \mathcal{O}(1/\sqrt{N_r})$ in the former case, whereas $\Delta I_2 = \mathcal{O}(1/\log(N_r))$ in the latter case. This difference arises as a consequence of the fundamental difference in asymptotic spectral properties in the two cases.

V. Discussion and Numerical Studies

We now use the bounds established in Section IV to develop a heuristic on the structure of $\mathbf{H}$ that is 'best' or 'worst' for a given precoding scheme. For this, we freeze $\mathbf{\Lambda}_r$ to be a fixed matrix so as to develop an understanding of the structure of $\mathbf{\Lambda}_t$ that minimizes the bounds to ΔI_{semi}.

Given that a constraint $\sum_{i=1}^{N_t} \mathbf{\Lambda}_t(i) = \rho_c$ has to be met, the common performance loss-minimizing $\mathbf{\Lambda}_t$ (if it exists) is the solution to the following simultaneous optimization:

$$\max \left\{ \sum_{i=1}^{M} \log\left(1 + \frac{\rho}{M}\mathbf{\Lambda}_t(i)\right), G_{M,\mathsf{tx}} \right\}, \quad \text{and} \quad \min \left\{ \mathbf{\Lambda}_t(1), \sum_{i=1}^{M} \log\left(1 + \frac{\delta_1}{\mathbf{\Lambda}_t(i)}\right) \right\} \tag{41}$$

for some $\delta_1 > 0$. The above objectives are equivalent to minimizing ΔI_2 in each of the four cases studied in Section IV. While these objectives are in general unrelated, as SNR and antenna dimensions increase, the four problems can be incorporated into the following optimization:

$$\max \prod_{i=1}^{M} \mathbf{\Lambda}_t(i) \ \text{subject to} \ \sum_{i=1}^{N_t} \mathbf{\Lambda}_t(i) = \rho_c. \tag{42}$$

The solution to the above problem is

$$\mathbf{\Lambda}_t(1) = \cdots = \mathbf{\Lambda}_t(M) = \frac{\rho_c}{M}, \tag{43}$$

$$\mathbf{\Lambda}_t(M+1) = \cdots = \mathbf{\Lambda}_t(N_t) = 0. \tag{44}$$

On the other extreme, the worst choice of $\mathbf{\Lambda}_t$ that minimizes $\prod_{i=1}^{M} \mathbf{\Lambda}_t(i)$ and hence, maximizes the upper bound to ΔI_{semi} is of the form:

$$\mathbf{\Lambda}_t(1) \approx \rho_c \ \text{ and } \ \mathbf{\Lambda}_t(i) \approx 0, i \geq 2, \tag{45}$$

but with the additional constraint that $\mathsf{rank}(\mathbf{\Lambda}_t) \geq M$. It is important to note that the largest gap[5] is *not* achieved when $\mathsf{rank}(\mathbf{\Lambda}_t) = 1$. Motivated by the above discussion, it is worthwhile defining a matching metric for the transmitter side:

$$\mathcal{M}_t \triangleq \prod_{i=1}^{M} \mathbf{\Lambda}_t(i), \tag{46}$$

that captures the closeness of a given channel from the best and worst channels. While $\mathcal{M}_t$ is defined following Section IV where bounds to ΔI_{semi} are obtained, we hope that as $\mathcal{M}_t$ increases, the channel becomes more matched on the transmitter side and the performance loss ΔI_{semi} decreases and *vice versa*.

Capturing the impact of $\mathbf{\Lambda}_r$ on performance loss in the general setting is difficult since $\mathbf{\Lambda}_r$ is hidden in the first-order analysis of Section IV. Nevertheless, in one special case, (25) suggests that a matching metric for the receiver side can be defined as

$$\mathcal{M}_r \triangleq \sum_{i=1}^{N_r} \left(\mathbf{\Lambda}_r(i)\right)^2. \tag{47}$$

Note that since $\sum_{i=1}^{N_r} \mathbf{\Lambda}_r(i) = \rho_c$, $\mathcal{M}_r$ is minimized by

$$\mathbf{\Lambda}_r = \frac{\rho_c}{N_r}\mathbf{I}_{N_r} \tag{48}$$

and maximized by

$$\mathbf{\Lambda}_r(1) \approx \rho_c \ \text{ and } \ \mathbf{\Lambda}_r(i) \approx 0, \ i \geq 2, \tag{49}$$

but with the added constraint that $\mathsf{rank}(\mathbf{\Lambda}_r) \geq M$. It can be seen that the performance loss is not maximized when $\mathsf{rank}(\mathbf{\Lambda}_r) < M$. As before, $\mathcal{M}_r$ is defined following bounds to ΔI_{semi} and the notion of matching has to be understood within this fundamental constraint.

To summarize the above discussion, we refer to a channel that is perfectly matched on both the transmitter and the receiver sides as a perfectly *matched channel* and this structure is optimal (as per the bounds established) for the given precoder structure (fixed choice of M). The structure of this channel is such that: 1) the rank of $\mathbf{\Lambda}_t$ is M with the dominant transmit eigenvalues being well-conditioned, and 2) $\mathbf{\Lambda}_r$ is also well-conditioned. A channel that is ill-conditioned on both the transmit and the receive sides such that $\mathsf{rank}(\mathbf{H}) \geq M$ (with probability 1) is said to be a perfectly *mismatched channel*.

An interesting consequence of the study in Theorems 1 and 2 is that channel hardening, that occurs as N_r increases, results in the vanishing of ΔI_{semi}. That is, statistical information is as good as perfect CSI in the receive antenna asymptotics. This behavior is peculiar of this asymptotic regime, as documented in the beamforming case [38], [40], [42]. The high SNR characterization for signaling with M spatial modes ($\rho \geq \alpha \frac{M}{\mathbf{\Lambda}_t(M)}$ for some $\alpha > 1$) has also been identified in prior work [41]. Our results can also be extended to the case of relative average error probability enhancement with the semiunitary precoder. However, these details are not provided here.

A. Numerical Studies

We now illustrate the results established so far via numerical studies.

- Conservatism of the Bounds: While Section IV has established bounds for ΔI_{semi} under certain assumptions, it is important to understand as to how conservative these bounds are and whether they capture the underlying tradeoffs accurately in the low to medium SNR regime and with reasonable antenna numbers. Fig. 1 compares the exact ΔI_{semi}, obtained via Monte Carlo averaging, with the bounds in Theorem 1 for the separable case with $N_t = 4, M = 2$ and $N_r = 4, 8, 16$ and 32. We plot $\log(\Delta I_{\mathsf{semi}})$ vs. ρ and while Fig. 1 shows that the bounds are loose (due to the lack of tight random matrix theoretic estimates) especially in the low SNR regime with small antenna numbers, they get tight in the regime where the theoretical results have been established. Nevertheless, the following study addresses the question of whether the intuition obtained via these bounds is useful in practice or not.
- Performance Gap as a Function of $\mathcal{M}_t$: In contrast to bounds on ΔI_{semi}, the focus here is on the performance gap between the perfect CSI and statistical precoders with the exact ΔI_{semi}. We consider 4×4 channels with $M = 2$ and freeze $\mathbf{U}_t$ and $\mathbf{U}_r$ to some fixed choice in our study. We also freeze $\mathbf{\Lambda}_r$ to $\mathbf{\Lambda}_r = 4\mathbf{I}_4$ so as to maintain $\rho_c = N_t N_r = 16$ and to focus on the impact of matching

[5]In fact, if $\mathsf{rank}(\mathbf{\Lambda}_t) = 1$, the statistical precoder achieves the same throughput as the optimal precoder.

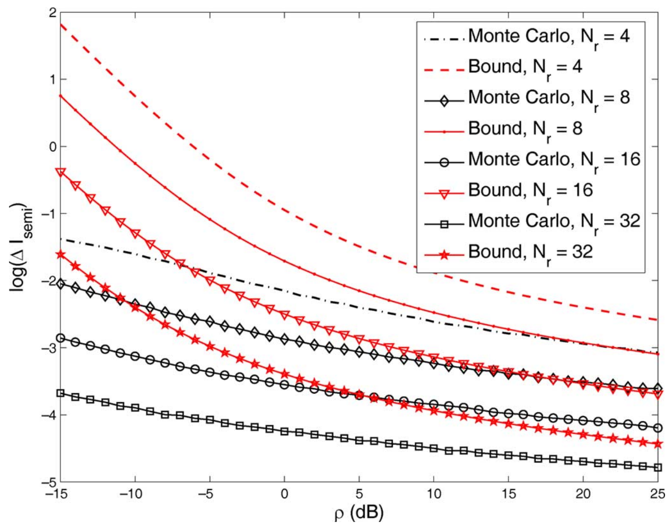

Fig. 1. Comparison of Monte Carlo estimates of ΔI_{semi} with the bounds established in Theorem 1.

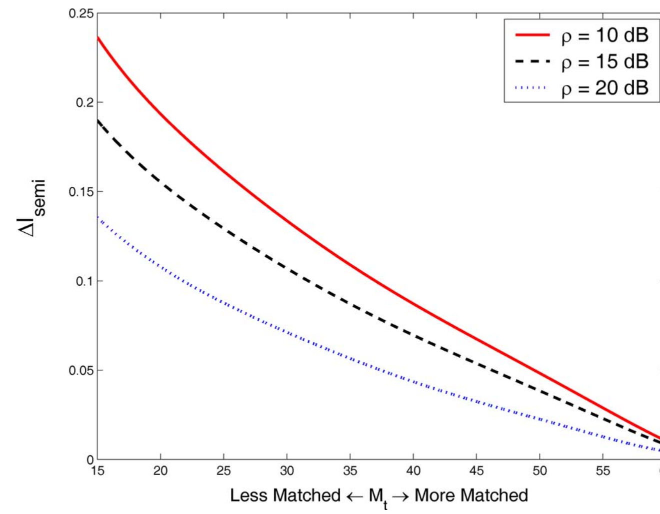

Fig. 2. Gap in mutual information performance between statistical and perfect CSI semiunitary precoding as a function of the matching metric $\mathcal{M}_t$.

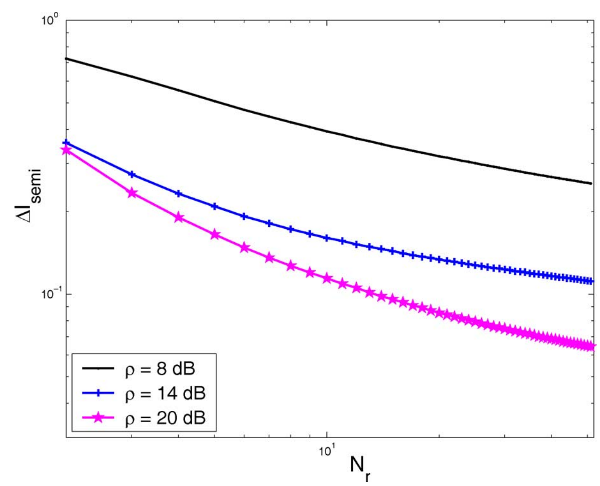

Fig. 3. Performance loss with the statistical semiunitary precoder for fixed $N_t = 4, M = 2$ as N_r increases.

on the transmitter side. Note that the matching metric, $\mathcal{M}_t = \prod_{k=1}^{M} \mathbf{\Lambda}_t(k)$, takes values in the range $(0, 64]$ in our setting. A family of ~ 1700 channels (each characterized uniquely by $\mathbf{\Lambda}_t(k),\ k = 1, \ldots, N_t$) is generated such that $\sum_{k=1}^{N_t} \mathbf{\Lambda}_t(k) = \rho_c = 16$ and $\mathcal{M}_t$ takes values over its range. The channels become more matched (on the transmitter side) to the precoder structure as $\mathcal{M}_t$ increases. While much of our study has been based on asymptotic random matrix theory, Fig. 2 illustrates that the notion of matched channels developed here is useful even in practically relevant regimes like 4×4 channels. Fig. 2 shows that the exact ΔI_{semi} decreases as $\mathcal{M}_t$ increases for three choices of ρ. Note that for a given channel as ρ increases, ΔI_{semi} decreases as $1/\log(\rho)$. It is important to note that while there exists no ordering relationship between any two matrix channels [31], when the focus is only on the mutual information performance, $\mathcal{M}_t$ (and $\mathcal{M}_r$) are sufficient to order channels.

- *Asymptotic Optimality:* The next study illustrates the asymptotic optimality of statistical precoding. Fig. 3 plots the exact ΔI_{semi} as a function of N_r with N_t and M fixed at $N_t = 4$ and $M = 2$. The channels have a separable correlation structure with $\mathbf{\Lambda}_t = \mathbf{I}_4$ whereas $\mathbf{\Lambda}_r = \frac{4}{N_r}\mathbf{I}_{N_r}$ which results in $\rho_c = 4$ for all the channels. As can be seen from the study in the previous section, channel hardening, where the eigenvectors of $\mathbf{H}^H\mathbf{H}$ converge to the eigenvectors of $\mathbf{\Sigma}_t = E[\mathbf{H}^H\mathbf{H}]$ as $\frac{N_t}{N_r} \rightarrow 0$, ensures that even statistical information is sufficient for near-perfect CSI performance as N_r increases.

VI. Concluding Remarks

The main focus of this work is on precoding for spatially correlated multiantenna channels that are often encountered in practice. Motivated and inspired by many recent wireless standardization efforts, we studied the performance of statistical semiunitary precoding in this paper. Here, the eigen-modes of the precoder are chosen to be the dominant eigenvectors of the transmit covariance matrix whereas the power allocation across the excited modes is uniform. We analytically characterized the relative average mutual information loss of the semiunitary precoder using tools from random matrix and eigenvector perturbation theories.

Our results show that given a precoder architecture (that is, fixed antenna dimensions and precoder rank), the relative difference metric is minimized by a channel that is matched to it. A matched channel is one that has: 1) the same number of dominant transmit eigen-modes as the precoder rank, and 2) the dominant transmit as well as the receive eigen-modes that are well-conditioned. Our theoretical study also characterizes *matching metrics* that enable the comparison of two channels with respect to performance loss captured by the relative difference metric. In particular, as the channel becomes more matched to the precoder structure and the matching metric changes accordingly *continuously*, the performance loss decreases monotonically and *vice versa*. As a by-product of our computations, we also showed that the semiunitary precoder is near-optimal in the relative antenna asymptotic setting for any channel. This result generalizes previous work [40], [42] on the beamforming case ($M = 1$) where the performance of the statistical beamforming scheme has been studied.

While prior works on statistical precoding exist, ours is the first attempt to transparently characterize the performance in terms of the channel statistics. Much of this study has been rendered feasible due to substantial advances in capturing the eigen-properties of random matrices with independent entries. Nevertheless, there exist many directions along which this work can be developed. We now list a few of these directions. This work is limited to the high SNR, large antenna asymptotic regime where a comprehensive random matrix theory is available to capture precoder performance [28], [29]. Even in this regime, it may be possible to refine the constants in the bounds for the relative loss terms and obtain further insights on the impact of spatial correlation on performance loss. The notion of precoder-channel matching introduced in this work can be developed further to aid in the design of low-complexity, structured and adaptive signaling schemes. In the case of mismatched channels, the construction of limited feedback schemes to bridge the gap in performance has been undertaken in recent work [39]. The question of tradeoffs between spatial versus spatio-temporal precoding and extensions to more general Ricean fading, multiuser, wideband systems are also of interest.

APPENDIX

A) Key Mathematical Results: We now introduce some key mathematical results from matrix theory that will be needed in the ensuing proofs.

Lemma 4: This lemma provides bounds for eigenvalues of sums and products of Hermitian matrices [43]. If $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ Hermitian matrices, for any $k = 1, \ldots, n$, we have

$$\lambda_k(\mathbf{A})\lambda_{\min}(\mathbf{B}) \leq \lambda_k(\mathbf{AB}) \leq \lambda_k(\mathbf{A})\lambda_{\max}(\mathbf{B}), \quad (50)$$

$$\lambda_k(\mathbf{A})+\lambda_{\min}(\mathbf{B}) \leq \lambda_k(\mathbf{A}+\mathbf{B}) \leq \lambda_k(\mathbf{A})+\lambda_{\max}(\mathbf{B}). \quad (51)$$

■

Lemma 5: This lemma extends the previous one to the complex case [31, p. 253–255]. Let $\mathbf{A}$ be an $n \times n$ complex matrix with $\{R_i, C_i\}$ defined as

$$R_i = \sum_{j=1}^{n} |\mathbf{A}(i,j)|, C_j = \sum_{i=1}^{n} |\mathbf{A}(i,j)|, i,j = 1, \ldots, n. \quad (52)$$

Let the eigenvalues of $\mathbf{A}$ be arranged in a decreasing order: $|\lambda_1(\mathbf{A})| \geq \cdots \geq |\lambda_n(\mathbf{A})|$, and let $\{R_i, C_i\}$ be rearranged such that $R_{[1]} \geq \cdots \geq R_{[n]}$ and $C_{[1]} \geq \cdots \geq C_{[n]}$. Then, we have

$$\prod_{i=1}^{k} |\lambda_i(\mathbf{A})| \leq \min\Big\{ \prod_{i=1}^{k} R_{[i]}, \prod_{i=1}^{k} C_{[i]} \Big\}. \quad (53)$$

■

Lemma 6: Let $\mathbf{X}$ be a $p \times n$ complex random matrix with i.i.d. entries of mean zero, common variance 1 and a finite fourth moment. Consider two cases: 1) p is finite and $n \to \infty$, and 2) $\{p, n\} \to \infty$ with $p/n \to 0$. In either case, in the asymptotics of n, the empirical eigenvalue distribution of $\frac{\mathbf{XX}^H - n\mathbf{I}_p}{2\sqrt{np}}$ converges pointwise with probability 1 to the semi-circular law $F(x)$ where

$$F(x) = \begin{cases} 0 & \text{if } x < -1, \\ \int_{y=-1}^{x} \frac{2}{\pi}\sqrt{1-y^2}\,\mathrm{d}y & \text{if } -1 \leq x \leq 1, \\ 1 & \text{if } x > 1. \end{cases} \quad (54)$$

In particular, with probability one, we have

$$1 - 2\sqrt{\frac{p}{n}} \leq \liminf_n \frac{\lambda_{\min}(\mathbf{XX}^H)}{n} \leq \limsup_n \frac{\lambda_{\max}(\mathbf{XX}^H)}{n} \leq 1 + 2\sqrt{\frac{p}{n}}. \quad (55)$$

Let $\mathbf{\Lambda}$ be an $n \times n$ positive definite diagonal matrix. Under the same assumptions on $\mathbf{X}, p, n$ as above, there exists a finite constant $\gamma_1 > 0$ (dependent on p and n only through $\mathbf{\Lambda}$) such that, with probability 1

$$\frac{\sum_i \mathbf{\Lambda}(i)}{n} - \gamma_1\sqrt{\frac{p}{n}} \leq \liminf_n \frac{\lambda_{\min}(\mathbf{X\Lambda X}^H)}{n} \leq \limsup_n \frac{\lambda_{\max}(\mathbf{X\Lambda X}^H)}{n} \leq \frac{\sum_i \mathbf{\Lambda}(i)}{n} + \gamma_1\sqrt{\frac{p}{n}}. \quad (56)$$

On the other hand, let $\mathbf{X}$ be a $p \times n$ complex random matrix with independent entries from a fixed probability space such that $\mathbf{X}(i,j)$ is zero mean, has variance σ_{ij}^2 and

$$\sup_{n,p} \max_{ij} E[|\mathbf{X}(i,j)|^4] \leq \gamma_2 < \infty. \quad (57)$$

Also, without loss of generality, assume that $\{ \sum_{j=1}^n \sigma_{ij}^2 \}$ are arranged in decreasing order. Then there exists a finite constant $\gamma_3 > 0$ (independent of p, n) such that, for all i

$$\frac{\sum_{j=1}^n \sigma_{ij}^2}{n} - \gamma_3\sqrt{\frac{p}{n}} \leq \liminf_n \frac{\lambda_i(\mathbf{XX}^H)}{n} \leq \limsup_n \frac{\lambda_i(\mathbf{XX}^H)}{n} \leq \frac{\sum_{j=1}^n \sigma_{ij}^2}{n} + \gamma_3\sqrt{\frac{p}{n}} \quad (58)$$

with probability 1.

Proof: We provide an elementary proof of the claim when p is finite, $n \to \infty$ and $\mathbf{X}(i,j)$ are standard, complex Gaussian. Define the set

$$A_n \triangleq \left\{ \omega : \frac{\lambda_{\max}(\mathbf{X}(\omega)\mathbf{\Lambda}\mathbf{X}(\omega)^H)}{n} > 1 + \epsilon_1 + \epsilon_2 \right\}. \quad (59)$$

If we can show that $\sum_n \Pr(A_n) < \infty$, it follows from the Borel-Cantelli lemma [44] that $\Pr(\limsup A_n) = 0$. By choosing ϵ_1 and ϵ_2 appropriately (as a function of n), we can establish strict bounds on the eigenvalues.

Breaking $\mathbf{X\Lambda X}^H$ into a diagonal component and an off-diagonal component and using Lemma 4, it follows via a union bound that

$$\Pr(A_n) \leq p\Pr\left(\frac{\sum_{i=1}^n \left(|\mathbf{X}(1,i)|^2 - 1\right)\mathbf{\Lambda}(i)}{n} > \epsilon_1 \right) + p^2\Pr\left(\frac{|\sum_{i=1}^n \mathbf{X}(1,i)\mathbf{\Lambda}(i)\mathbf{X}(2,i)^\star|}{n} > \epsilon_2 \right). \quad (60)$$

Using a Chernoff-type bound [44], we have the following:

$$\Pr(A_n) \leq p \exp\left(-\frac{\epsilon_1^2 n^2}{2\sum_{i=1}^{n}(\mathbf{\Lambda}(i))^2}\right) + 2p^2 \exp\left(-\frac{\epsilon_2^2 n^2 c}{\sum_{i=1}^{n}(\mathbf{\Lambda}(i))^2}\right) \quad (61)$$

for some $c > 0$. The smallest value of ϵ_1 and ϵ_2 that can still result in $\Pr(\limsup A_n) = 0$ is such that

$$\epsilon_1 = \mathcal{O}(\epsilon_2) = \sqrt{\frac{\sum_{i=1}^{n}(\mathbf{\Lambda}(i))^2}{n}} \cdot \frac{1}{n^{1/2-\eta}},\ \eta > 0. \quad (62)$$

Letting $\eta \downarrow 0$, we have

$$\limsup \frac{\lambda_{\max}(\mathbf{X\Lambda X}^H)}{n} \leq \frac{\sum_{i=1}^{n}\mathbf{\Lambda}(i)}{n} + \gamma_4\sqrt{\frac{\sum_{i=1}^{n}\mathbf{\Lambda}(i)^2}{n}} \cdot \frac{1}{\sqrt{n}}, \quad (63)$$

where $\gamma_4 > 0$ is a constant independent of p and n. The expression for $\lambda_{\min}(\cdot)$ is symmetric with that of $\lambda_{\max}(\cdot)$ and can be obtained similarly. The extension to the case where $\mathbf{X}$ has only independent entries (not necessarily complex Gaussian) also proceeds via the same logic.

Since $p \to \infty$ in Case 2), the above technique is not useful in establishing the claim of the lemma. Here, the result follows from [28], [45, Th. 2.9, p. 623]. The generalizations with $\mathbf{\Lambda}$ and independent entries follow via the same proof technique as in [45] and hence no proofs are provided. The readers are referred to [28] for a brief summary of the general technique. ■

Stochastic Approximation for Random Determinants: In the case of an $N \times N$ matrix $\mathbf{H}_{\mathsf{iid}}$, stochastic properties of $\det(\mathbf{H}_{\mathsf{iid}}\mathbf{H}_{\mathsf{iid}}^H)$ can be studied using the *Bartlett decomposition* (or bidiagonalization) of a sample covariance matrix [46], [47], which states that there exist independent random variables $\mathbf{Z}_i$ on some probability space such that

$$\mathbf{Z} \triangleq \det\left(\mathbf{H}_{\mathsf{iid}}\mathbf{H}_{\mathsf{iid}}^H\right) \sim \prod_{i=1}^{N}\mathbf{Z}_i, \quad (64)$$

$$\mathbf{Z}_i \sim \sum_{j=i}^{N}|\mathbf{H}_{\mathsf{iid}}(i,j)|^2 \sim \frac{1}{2}\chi^2\left(2(N-i+1)\right) \quad (65)$$

where $\chi^2(2k)$ is a central chi-squared random variable with $2k$ degrees of freedom. In the non-i.i.d. case, performing this task is difficult as an equivalent decomposition is not known. Nevertheless, a tight stochastic approximation for the random determinant is still possible.

Lemma 7 (Girko): Let $\widetilde{\mathbf{H}}_{\mathsf{ind}}$ be an $N_r \times M$ random matrix with $N_r \geq M$ and independent entries that are distributed as $\mathcal{CN}(0, \sigma_{ij}^2)$. There exist independent random variables $\widetilde{\mathbf{Z}}_i, i = 1, \ldots, M$ on some probability space such that $\det\left(\widetilde{\mathbf{H}}_{\mathsf{ind}}^H\widetilde{\mathbf{H}}_{\mathsf{ind}}\right)$ can be well approximated as

$$\det\left(\widetilde{\mathbf{H}}_{\mathsf{ind}}^H\widetilde{\mathbf{H}}_{\mathsf{ind}}\right) \approx \prod_{i=1}^{M}\widetilde{\mathbf{Z}}_i, \widetilde{\mathbf{Z}}_i \sim i \cdot \frac{\sum_{k=1}^{N_r}|\widetilde{\mathbf{H}}_{\mathsf{ind}}(k,i)|^2}{N_r}. \quad (66)$$

Proof: See [47, Ch. 2, p. 104] and [48, pp. 35, 39] for a version of the above statement on random determinant approximation. The justifications for the approximation are found in [26, Lemma 5]. ■

B) Proof of Proposition 1: To characterize the behavior of ΔI_1, recall the structure of $\mathbf{F}_{\mathsf{perf}}$ and $\mathbf{F}_{\mathsf{perf,semi}}$ from Lemmas 1 and 2. Using these facts, we have

$$\Delta I_1 \cdot E_{\mathbf{H}}\left[I_{\mathsf{stat,semi}}\right] = E_{\mathbf{H}}\left[\sum_{i=1}^{n_{\mathbf{H}}}\log\left(1 + \mathbf{\Lambda_H}(i)\mathbf{\Lambda}_{\mathsf{wf}}(i)\right)\right] - E_{\mathbf{H}}\left[\sum_{i=1}^{M}\log\left(1 + \frac{\rho}{M}\mathbf{\Lambda_H}(i)\right)\right] \quad (67)$$

where given a channel realization $\mathbf{H}$, $\left\{\mathbf{\Lambda_H}(i), i = 1, \ldots, N_t\right\}$ are the squared singular values of $\mathbf{H}$, $n_{\mathbf{H}}$ modes of the channel are excited ($1 \leq n_{\mathbf{H}} \leq M$) with power $\mathbf{\Lambda}_{\mathsf{wf}}(i) \triangleq \left(\mu_{\mathbf{H}} - \frac{1}{\mathbf{\Lambda_H}(i)}\right)^+$ and the water level $\mu_{\mathbf{H}}$ is chosen such that $\sum_{i=1}^{n_{\mathbf{H}}}\mathbf{\Lambda}_{\mathsf{wf}}(i) = \rho$. It can be easily checked that $\mathbf{\Lambda}_{\mathsf{wf}}(i)$ can be written as

$$\mathbf{\Lambda}_{\mathsf{wf}}(i) = \frac{\rho}{n_{\mathbf{H}}} + \frac{1}{n_{\mathbf{H}}}\sum_{j=1}^{n_{\mathbf{H}}}\frac{1}{\mathbf{\Lambda_H}(j)} - \frac{1}{\mathbf{\Lambda_H}(i)} \quad (68)$$

and $n_{\mathbf{H}}$ satisfies

$$n_{\mathbf{H}} = \arg\max k \ \text{ s.t. } 1 \leq k \leq M,\ \sum_{i=1}^{k}\frac{\mathbf{\Lambda_H}(i) - \mathbf{\Lambda_H}(k)}{\mathbf{\Lambda_H}(i)\mathbf{\Lambda_H}(k)} \leq \rho. \quad (69)$$

Hence, as stated in (70) and (71) at the bottom of the page, we have a bound on ΔI_1. In the second inequality, we have used

$$\Delta I_1 \cdot E_{\mathbf{H}}\left[I_{\mathsf{stat,semi}}\right] \leq E_{\mathbf{H}}\left[\sum_{i=1}^{n_{\mathbf{H}}}\log\left(1 + \frac{\frac{\rho\mathbf{\Lambda_H}(i)(M-n_{\mathbf{H}})}{n_{\mathbf{H}}M} - 1 + \frac{\mathbf{\Lambda_H}(i)}{n_{\mathbf{H}}}\sum_{j=1}^{n_{\mathbf{H}}}\frac{1}{\mathbf{\Lambda_H}(j)}}{1 + \frac{\rho\mathbf{\Lambda_H}(i)}{M}}\right)\right] \quad (70)$$

$$\leq E_{\mathbf{H}}\left[\sum_{i=1}^{n_{\mathbf{H}}}\frac{\frac{\rho\mathbf{\Lambda_H}(i)(M-n_{\mathbf{H}})}{n_{\mathbf{H}}M} - 1 + \frac{\mathbf{\Lambda_H}(i)}{n_{\mathbf{H}}}\sum_{j=1}^{n_{\mathbf{H}}}\frac{1}{\mathbf{\Lambda_H}(j)}}{1 + \frac{\rho\mathbf{\Lambda_H}(i)}{M}}\right]. \quad (71)$$

the fact that $\log(1+x) \le x$ for all $x > -1$. The following simplifications follow routinely:

$$\begin{aligned}\Delta I_1 \cdot E_{\mathbf{H}}\left[I_{\text{stat,semi}}\right] - E_{\mathbf{H}}\left[M - n_{\mathbf{H}}\right] &\le E_{\mathbf{H}}\left[\frac{1}{n_{\mathbf{H}}}\sum_{i=1}^{n_{\mathbf{H}}}\left(\frac{\sum_j \frac{\mathbf{\Lambda_H}(i)}{\mathbf{\Lambda_H}(j)}}{1+\frac{\rho\mathbf{\Lambda_H}(i)}{M}} - \frac{M}{1+\frac{\rho\mathbf{\Lambda_H}(i)}{M}}\right)\right] && (72)\\ &\le E_{\mathbf{H}}\left[\frac{M}{n_{\mathbf{H}}}\sum_{i=1}^{n_{\mathbf{H}}}\left(\sum_j \frac{1}{\rho\mathbf{\Lambda_H}(j)} - \frac{1}{1+\frac{\rho\mathbf{\Lambda_H}(i)}{M}}\right)\right] && (73)\\ &= E_{\mathbf{H}}\left[\frac{M}{n_{\mathbf{H}}}\sum_{i=1}^{n_{\mathbf{H}}}\left(\frac{n_{\mathbf{H}}}{\rho\mathbf{\Lambda_H}(i)} - \frac{1}{1+\frac{\rho\mathbf{\Lambda_H}(i)}{M}}\right)\right] && (74)\\ &= E_{\mathbf{H}}\left[\frac{M}{n_{\mathbf{H}}}\sum_{i=1}^{n_{\mathbf{H}}}\frac{n_{\mathbf{H}} + \rho\mathbf{\Lambda_H}(i)\left(\frac{n_{\mathbf{H}}}{M}-1\right)}{\rho\mathbf{\Lambda_H}(i)\left(1+\frac{\rho\mathbf{\Lambda_H}(i)}{M}\right)}\right] && (75)\\ &\le \frac{M^2}{\rho^2}\cdot E_{\mathbf{H}}\left[\sum_{i=1}^{M}\frac{1}{\mathbf{\Lambda_H}(i)^2}\right]. && (76)\end{aligned}$$

From (69), it is easily recognized that $n_{\mathbf{H}} \ge k$ if

$$\rho \ge \frac{k}{\mathbf{\Lambda_H}(k)} - \sum_{i=1}^{k}\frac{1}{\mathbf{\Lambda_H}(i)}. \quad (77)$$

Thus, if $\rho > \alpha E_{\mathbf{H}}\left[\frac{M}{\mathbf{\Lambda_H}(M)}\right]$ for some $\alpha > 1$ as in the statement of the theorem, both the terms in the expansion of $E_{\mathbf{H}}\left[I_{\text{stat,semi}}\right]$ in (76) can be bounded by constants that depend only on the channel statistics. For this note that

$$\begin{aligned}E_{\mathbf{H}}\left[M - n_{\mathbf{H}}\right] &\le M\cdot \Pr(n_{\mathbf{H}} < M) && (78)\\ &\le M\cdot\Pr\left(\frac{M}{\mathbf{\Lambda_H}(M)} - \sum_{i=1}^{M}\frac{1}{\mathbf{\Lambda_H}(i)} > \rho\right) && (79)\\ &\le M\cdot\Pr\left(\frac{1}{\mathbf{\Lambda_H}(M)} > \alpha E\left[\frac{1}{\mathbf{\Lambda_H}(M)}\right]\right) && (80)\\ &\overset{(a)}{\le} \frac{M}{\alpha^2}\cdot\frac{E\left[\left(\frac{1}{\mathbf{\Lambda_H}(M)}\right)^2\right]}{\left(E\left[\frac{1}{\mathbf{\Lambda_H}(M)}\right]\right)^2} && (81)\end{aligned}$$

where (a) follows from Chebyshev's inequality. A trivial upper bound for the other term gives the desired result. ■

C) Proof of Theorem 1: It can be checked that the numerator, $\mathcal{N}$, of ΔI_2 can be written as

$$\begin{aligned}\mathcal{N} = E_{\mathbf{H}}&\left[\sum_{k=1}^{M}\log\left(1+\frac{\rho}{M\rho_c}\lambda_k\left(\mathbf{\Lambda}_t\mathbf{H}_{\text{iid}}^H\mathbf{\Lambda}_r\mathbf{H}_{\text{iid}}\right)\right)\right]\\ &- E_{\mathbf{H}}\left[\log\left(1+\frac{\rho}{M\rho_c}\lambda_k\left(\widetilde{\mathbf{\Lambda}}_t\widetilde{\mathbf{H}}_{\text{iid}}^H\mathbf{\Lambda}_r\widetilde{\mathbf{H}}_{\text{iid}}\right)\right)\right] \quad (82)\end{aligned}$$

where $\widetilde{\mathbf{H}}_{\text{iid}}$ is the $N_t \times M$ principal submatrix of $\mathbf{H}_{\text{iid}}$ and $\widetilde{\mathbf{\Lambda}}_t$ is the $M \times M$ principal submatrix of $\mathbf{\Lambda}_t$. An application of Lemma 4 shows that

$$\begin{aligned}\mathcal{N} \le E_{\mathbf{H}}&\left[\sum_{k=1}^{M}\log\left(1+\frac{\rho\mathbf{\Lambda}_t(k)}{M\rho_c}\lambda_{\max}\left(\mathbf{H}_{\text{iid}}^H\mathbf{\Lambda}_r\mathbf{H}_{\text{iid}}\right)\right)\right]\\ &- E_{\mathbf{H}}\left[\log\left(1+\frac{\rho\mathbf{\Lambda}_t(k)}{M\rho_c}\lambda_{\min}\left(\widetilde{\mathbf{H}}_{\text{iid}}^H\mathbf{\Lambda}_r\widetilde{\mathbf{H}}_{\text{iid}}\right)\right)\right]. \quad (83)\end{aligned}$$

Following an application of Lemma 6, we have

$$\begin{aligned}\mathcal{N} \le \sum_{k=1}^{M}&\log\left(1+\frac{\rho\mathbf{\Lambda}_t(k)}{M}\left(1+\gamma\frac{\sqrt{\sum_i(\mathbf{\Lambda}_r(i))^2}}{\sum_i\mathbf{\Lambda}_r(i)}\right)\right)\\ &- \log\left(1+\frac{\rho\mathbf{\Lambda}_t(k)}{M}\left(1-\gamma'\frac{\sqrt{\sum_i(\mathbf{\Lambda}_r(i))^2}}{\sum_i\mathbf{\Lambda}_r(i)}\right)\right) \quad (84)\end{aligned}$$

where γ and γ' follow from the corresponding bounds in Lemma 6. After some straightforward simplifications, we have

$$\begin{aligned}\mathcal{N} \le &\sum_{k=1}^{M}\log\left(1+\frac{\gamma\cdot\rho\mathbf{\Lambda}_t(k)}{M+\rho\mathbf{\Lambda}_t(k)}\cdot\frac{\sqrt{\sum_i(\mathbf{\Lambda}_r(i))^2}}{\sum_i\mathbf{\Lambda}_r(i)}\right)\\ &- \sum_{k=1}^{M}\log\left(1-\frac{\gamma'\cdot\rho\mathbf{\Lambda}_t(k)}{M+\rho\mathbf{\Lambda}_t(k)}\cdot\frac{\sqrt{\sum_i(\mathbf{\Lambda}_r(i))^2}}{\sum_i\mathbf{\Lambda}_r(i)}\right). \quad (85)\end{aligned}$$

If $x \le \frac{1}{2}$, we have

$$\begin{aligned}-\log(1-x) &= \log\left(1+\frac{x}{1-x}\right)\\ &\le \log\left(1+x(1+2x)\right) \le \log(1+2x) \quad (86)\end{aligned}$$

and this in combination with the log-inequality results in

$$\begin{aligned}\mathcal{N} &\le (\gamma+2\gamma')\cdot\frac{\sqrt{\sum_i(\mathbf{\Lambda}_r(i))^2}}{\sum_i\mathbf{\Lambda}_r(i)}\cdot\sum_{k=1}^{M}\frac{\rho\mathbf{\Lambda}_t(k)}{M+\rho\mathbf{\Lambda}_t(k)} && (87)\\ &\le (\gamma+2\gamma')\cdot M\cdot\frac{\sqrt{\sum_i(\mathbf{\Lambda}_r(i))^2}}{\sum_i\mathbf{\Lambda}_r(i)}. && (88)\end{aligned}$$

A lower bound to the denominator term, $E_{\mathbf{H}}\left[I_{\text{stat,semi}}\right]$, can be obtained via the same logic and combining these two bounds result in the statement of the theorem. ■

D) Proof of Proposition 2: We have the following well-known facts:

$$I_{\text{perf}} = \log\left(1+\rho\lambda_1\right) \quad (89)$$

$$I_{\text{stat}} = \log\left(1+\rho\sum_{k=1}^{N_t}\lambda_k|\mathbf{v}_k^H\mathbf{u}_{\text{stat}}|^2\right) \quad (90)$$

where $\lambda_1 = \lambda_{\max}(\mathbf{H}\mathbf{H}^H)$, $\mathbf{u}_{\text{stat}}$ is an eigenvector corresponding to the dominant eigenvalue of $\mathbf{\Sigma}_t$ and an eigen-decomposition of $\mathbf{H}^H\mathbf{H}$ is of the form

$$\mathbf{H}^H\mathbf{H} = \sum_{k=1}^{N_t}\lambda_k\mathbf{v}_k\mathbf{v}_k^H. \quad (91)$$

The following simplifications can be made:

$$\Delta I_{\mathsf{bf}} \cdot E_{\mathbf{H}}\left[I_{\mathsf{stat}}\right] \leq E_{\mathbf{H}}\left[\log(1+\rho\lambda_1)\right] - E_{\mathbf{H}}\left[\log(1+\rho\lambda_1|\mathbf{v}_k^H\mathbf{u}_{\mathsf{stat}}|^2)\right] \quad (92)$$

$$E_{\mathbf{H}}\left[\log(1+\rho\lambda_1|\mathbf{v}_k^H\mathbf{u}_{\mathsf{stat}}|^2)\right] \geq E\left[\log\left(1+\rho\lambda_1(1-\delta)\right)\cdot\chi\left(|\mathbf{v}_k^H\mathbf{u}_{\mathsf{stat}}|^2>1-\delta\right)\right] \quad (93)$$

$$\stackrel{(a)}{=} E\left[\log\left(1+\rho\lambda_1(1-\delta)\right)\right]\Pr\left(|\mathbf{v}_k^H\mathbf{u}_{\mathsf{stat}}|^2>1-\delta\right) \quad (94)$$

$$\stackrel{(b)}{\geq} E\left[\log\left(1+\rho\lambda_1(1-\delta)\right)\right]\cdot\left(1-2N_t e^{-\frac{\delta\kappa N_r}{N_t-1}}\right) \quad (95)$$

where the bounds are optimized over the choice of δ, (a) follows from the independence between singular values and singular vectors of random matrices with independent entries [29], [47], [48], (b) follows from the distortion bound computed in [40, Th. 1] via eigenvector perturbation theory, and κ is a constant that depends only on the eigenvalues of $\mathbf{\Sigma}_t$ and $\mathbf{\Sigma}_r$. We thus have

$$\Delta I_{\mathsf{bf}} \cdot E_{\mathbf{H}}\left[I_{\mathsf{stat}}\right] \leq E_{\mathbf{H}}\left[\log\left(1+\frac{\rho\lambda_1\delta}{1+\rho\lambda_1(1-\delta)}\right)\right] + 2N_t \cdot e^{-\frac{\delta\kappa N_r}{N_t-1}} \cdot E_{\mathbf{H}}\left[I_{\mathsf{perf}}\right]. \quad (96)$$

Upon applying Jensen's inequality and noting that $E_{\mathbf{H}}[\lambda_1] \leq \rho_c = N_t N_r$, we have

$$E_{\mathbf{H}}\left[I_{\mathsf{perf}}\right] \leq \log(1+\rho N_t N_r) \quad (97)$$

which when used with the choice

$$\delta = \frac{N_t}{N_r\cdot\kappa}\cdot\left[\log(2N_r)+\log\left(\log\left(1+\rho N_t N_r\right)\right)\right] \quad (98)$$

results in

$$\Delta I_{\mathsf{bf}} \leq \frac{\log\left(1+\frac{\delta}{1-\delta}\right)+\frac{N_t}{N_r}\cdot\frac{E_{\mathbf{H}}[I_{\mathsf{perf}}]}{\log(1+\rho N_t N_r)}}{E_{\mathbf{H}}\left[I_{\mathsf{stat}}\right]}. \quad (99)$$

In the regime where $\frac{N_t}{N_r}\to 0$, both the terms in the above equation are on the same order and thus, we have

$$\Delta I_{\mathsf{bf}} \cdot E_{\mathbf{H}}\left[I_{\mathsf{stat}}\right] \leq \frac{N_t\cdot\log(N_r)}{N_r}\cdot\kappa_1 \quad (100)$$

where κ_1 is an appropriate condition number-dependent quantity. Using (95) with the choice of δ in (98) followed by an application of Lemma 6 leads to the statement of the proposition. ■

E) Proof of Theorem 3: As in Appendix C, we can write ΔI_2 as

$$1+\Delta I_2 = \frac{E_{\mathbf{H}}[I_{\mathsf{perf,semi}}]}{E_{\mathbf{H}}[I_{\mathsf{stat,semi}}]} \quad (101)$$

$$= \frac{E_{\mathbf{H}}\left[\sum_{k=1}^{M}\log\left(1+\frac{\rho}{M}\lambda_k(\mathbf{H}^H\mathbf{H})\right)\right]}{E_{\mathbf{H}}\left[\sum_{k=1}^{M}\log\left(1+\frac{\rho}{M\rho_c}\lambda_k(\widetilde{\mathbf{\Lambda}}_t\widetilde{\mathbf{H}}_{\mathsf{iid}}^H\mathbf{\Lambda}_r\widetilde{\mathbf{H}}_{\mathsf{iid}})\right)\right]}. \quad (102)$$

The denominator of (102) can be computed following the method in [30, Theorem 1] and equals

$$E_{\mathbf{H}}[I_{\mathsf{stat,semi}}] = \sum_{k=1}^{M}\log\left(1+\frac{\rho}{\rho_c}\mu\mathbf{\Lambda}_t(k)\right) + \sum_{k=1}^{M}\log\left(1+\frac{\rho}{\rho_c}\widetilde{\mu}\mathbf{\Lambda}_r(k)\right) - \frac{\rho M}{\rho_c}\mu\widetilde{\mu} \quad (103)$$

where μ and $\widetilde{\mu}$ satisfy the recursive equations

$$\mu = \frac{1}{M}\sum_{k=1}^{M}\frac{\mathbf{\Lambda}_r(k)}{1+\frac{\rho}{\rho_c}\widetilde{\mu}\mathbf{\Lambda}_r(k)} \quad (104)$$

$$\widetilde{\mu} = \frac{1}{M}\sum_{k=1}^{M}\frac{\mathbf{\Lambda}_t(k)}{1+\frac{\rho}{\rho_c}\mu\mathbf{\Lambda}_t(k)}. \quad (105)$$

A simple lower bound for $E_{\mathbf{H}}[I_{\mathsf{stat,semi}}]$ is obtained by using the facts that $\log(1+x)\geq\log(x)$ for $x>0$ and $\frac{\rho}{\rho_c}\mu\widetilde{\mu}\leq 1$ resulting in

$$E_{\mathbf{H}}[I_{\mathsf{stat,semi}}] \geq \sum_{k=1}^{M}\log\left(\frac{\rho^2}{\rho_c^2 e}\mu\widetilde{\mu}\mathbf{\Lambda}_t(k)\mathbf{\Lambda}_r(k)\right). \quad (106)$$

We now establish that the above bound is order-optimal as α increases (with $\rho=\alpha\frac{M}{\mathbf{\Lambda}_t(M)}$), by lower bounding $\mu\widetilde{\mu}$. For this, note that $\frac{x}{1+ax}, a>0$ is monotonically increasing in x and hence

$$\mu \geq \frac{\mathbf{\Lambda}_r(M)}{1+\frac{\rho}{\rho_c}\widetilde{\mu}\mathbf{\Lambda}_r(M)}, \widetilde{\mu} \geq \frac{\mathbf{\Lambda}_t(M)}{1+\frac{\rho}{\rho_c}\mu\mathbf{\Lambda}_t(M)} \quad (107)$$

combining both of which results in the quadratic inequality

$$\frac{\rho^2}{\rho_c^2}\mathbf{\Lambda}_t(M)\mathbf{\Lambda}_r(M)\left(\mu\widetilde{\mu}\right)^2 - \left(\frac{2\rho}{\rho_c}\mathbf{\Lambda}_t(M)\mathbf{\Lambda}_r(M)+1\right)\cdot\mu\widetilde{\mu} + \mathbf{\Lambda}_t(M)\mathbf{\Lambda}_r(M) \leq 0. \quad (108)$$

It is straightforward to check that

$$\frac{2\rho^2}{\rho_c^2}\cdot\mathbf{\Lambda}_t(M)\mathbf{\Lambda}_r(M)\cdot\mu\widetilde{\mu} \geq \frac{2\rho}{\rho_c}\cdot\mathbf{\Lambda}_t(M)\mathbf{\Lambda}_r(M)+1 - \sqrt{\frac{4\rho}{\rho_c}\cdot\mathbf{\Lambda}_t(M)\mathbf{\Lambda}_r(M)+1}. \quad (109)$$

Letting A and B denote $A=\frac{N_t N_r}{M^2}$ and $B=\frac{M}{\mathbf{\Lambda}_r(M)}$, and noting that both are $\mathcal{O}(1)$ according to the assumption of the theorem, elementary computation shows that

$$\frac{\rho}{\rho_c}\mu\widetilde{\mu} \geq 1 - \frac{\sqrt{AB}\cdot\sqrt{AB+4\alpha}}{2\alpha} \quad (110)$$

with $\rho = \alpha \cdot \frac{M}{\Lambda_t(M)}$. Combining these facts, we have

$$E_{\mathbf{H}}[I_{\mathsf{stat,semi}}] \geq M \log\left(1 - \frac{\sqrt{AB}\cdot\sqrt{AB+4\alpha}}{2\alpha}\right) + \sum_{k=1}^{M} \log\left(\frac{\rho}{e\rho_c}\cdot \mathbf{\Lambda}_t(k)\mathbf{\Lambda}_r(k)\right). \quad (111)$$

Proceeding in the same way, one can obtain an upper bound for $E_{\mathbf{H}}[I_{\mathsf{perf,semi}}]$. Since the main goal here is to obtain the trends of ΔI_2, we find it convenient and less cumbersome[6] to replace the upper bound with an approximation ($\log(1+x) \approx \log(x)$) by ignoring the term that decays as $\frac{1}{x}$. Thus, we have

$$\begin{aligned} &E_{\mathbf{H}}[I_{\mathsf{perf,semi}}] \\ &\approx M\log\left(\frac{\rho}{M}\right) + E_{\mathbf{H}}\left[\sum_{k=1}^{M}\log\left(\frac{\lambda_k(\mathbf{\Lambda}_t\mathbf{H}_{\mathsf{iid}}^H\mathbf{\Lambda}_r\mathbf{H}_{\mathsf{iid}})}{\rho_c}\right)\right] && (112)\\ &\overset{(a)}{\leq} M\log\left(\frac{\rho}{M}\right) + \min(A,B) && (113) \end{aligned}$$

$$A = ME_{\mathbf{H}}\left[\log\left(\frac{\lambda_{\max}(\mathbf{H}_{\mathsf{iid}}^H\mathbf{\Lambda}_r\mathbf{H}_{\mathsf{iid}})}{\rho_c}\right)\right] + \sum_{k=1}^{M}\log\left(\mathbf{\Lambda}_t(k)\right) \quad (114)$$

$$B = ME_{\mathbf{H}}\left[\log\left(\frac{\lambda_{\max}(\mathbf{H}_{\mathsf{iid}}\mathbf{\Lambda}_t\mathbf{H}_{\mathsf{iid}}^H)}{\rho_c}\right)\right] + \sum_{k=1}^{M}\log\left(\mathbf{\Lambda}_r(k)\right) \quad (115)$$

where in (a) we have used Lemma 4. Combining (111) and (113), we have

$$\Delta I_2 \leq \frac{\log(e/M) + \kappa_3}{\log(\rho/e) + \log(X/\rho_c) + \log(G_{M,\mathsf{tx}}\cdot G_{M,\mathsf{rx}})} \quad (116)$$

$$\kappa_3 = \min\left\{E_{\mathbf{H}}\left[\log\left(\lambda_{\max}(\mathbf{H}_{\mathsf{iid}}^H\mathbf{\Lambda}_r\mathbf{H}_{\mathsf{iid}})\right)\right] \; E_{\mathbf{H}}\left[\log\left(\lambda_{\max}(\mathbf{H}_{\mathsf{iid}}\mathbf{\Lambda}_t\mathbf{H}_{\mathsf{iid}}^H)\right)\right]\right\} - \log(G_{M,\mathsf{tx}}) - \log(G_{M,\mathsf{rx}}) - \log(X) \quad (117)$$

where X and $G_{M,\bullet}$ are as defined in the statement of the theorem. Noting that [28]

$$\limsup \frac{\lambda_{\max}(\mathbf{H}_{\mathsf{iid}}^H\mathbf{H}_{\mathsf{iid}})}{N_r} \leq K \quad (118)$$

for some appropriate constant K that only depends on N_t and N_r, we have the statement of the theorem. ■

[6]The approximation can be made precise, but we will not bother with this technicality here.

F) Proof of Proposition 3: We first apply Lemma 5 with $\mathbf{A} = \mathbf{H}_{\mathsf{ind}}^H\mathbf{H}_{\mathsf{ind}}$, $n = N_t$ and $k = M$ to bound the product of eigenvalues of $\mathbf{A}$, resulting in

$$\prod_{i=1}^{M}\lambda_i(\mathbf{H}_{\mathsf{ind}}^H\mathbf{H}_{\mathsf{ind}}) \leq \prod_{i=1}^{M} C_{[i]} \quad (119)$$

where

$$\frac{C_i}{N_r} = \frac{\sum_{k=1}^{N_r}|\mathbf{H}_{\mathsf{ind}}(k,i)|^2}{N_r} + \sum_{j=1, j\neq i}^{M}\frac{\left|\sum_{k=1}^{N_r}\mathbf{H}_{\mathsf{ind}}(k,j)\mathbf{H}_{\mathsf{ind}}^{\star}(k,i)\right|}{N_r}. \quad (120)$$

Using the law of large numbers, we know that the first term converges to $\frac{\mathbf{\Lambda}_t(i)}{N_r}$ whereas each of the terms in the second sum is small with high probability. More precisely, for every $\delta > 0$, there exists an $\epsilon > 0$ such that

$$C_i \leq \mathbf{\Lambda}_t(i) + \delta(M-1)N_r \text{ with prob. } \geq 1-(M-1)\epsilon. \quad (121)$$

Thus, we have

$$\begin{aligned} &E_{\mathbf{H}}\left[I_{\mathsf{perf,semi}}\right] \\ &= E_{\mathbf{H}}\left[\sum_{k=1}^{M}\log\left(1+\frac{\rho}{M}\lambda_k(\mathbf{H}_{\mathsf{ind}}\mathbf{H}_{\mathsf{ind}}^H)\right)\right] && (122)\\ &\overset{(a)}{\approx} M\log\left(\frac{\rho}{M}\right) + \sum_{k=1}^{M}\log\left(\lambda_k(\mathbf{H}_{\mathsf{ind}}\mathbf{H}_{\mathsf{ind}}^H)\right) && (123)\\ &\overset{(b)}{\leq} M\log\left(\frac{\rho}{M}\right) + \sum_{k=1}^{M}\log\left(\mathbf{\Lambda}_t(i) + \epsilon(M-1)N_r\right) && (124) \end{aligned}$$

where the approximation in (a) is using the high SNR assumption and (b) follows from (119) and has to be read as an approximation with high probability (following the earlier discussion).

For $E_{\mathbf{H}}[I_{\mathsf{stat,semi}}]$, we have the following high SNR approximation:

$$\begin{aligned} &E_{\mathbf{H}}\left[I_{\mathsf{stat,semi}}\right] \\ &\approx M\log\left(\frac{\rho}{M}\right) + E_{\mathbf{H}}\left[\log\det\left(\widetilde{\mathbf{H}}_{\mathsf{ind}}^H\widetilde{\mathbf{H}}_{\mathsf{ind}}\right)\right] && (125)\\ &\overset{(a)}{\approx} M\log\left(\frac{\rho}{M}\right) + \sum_{i=1}^{M}\log\left(\frac{i}{N_r}\cdot\mathbf{\Lambda}_t(i)\right) && (126)\\ &\xrightarrow{(b)} M\log\left(\frac{\rho}{M}\right) + M\log\left(\frac{M}{N_r e}\right) + \sum_{i=1}^{M}\log\left(\mathbf{\Lambda}_t(i)\right) && (127) \end{aligned}$$

where (a) follows from Lemma 7 and (b) follows from Stirling approximation as $\{M, N_r\} \to \infty$. Combining (124) and (127), we obtain the statement in (36). ■

ACKNOWLEDGMENT

The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Andrea Goldsmith. The authors would like to thank the anonymous reviewers and the Associate Editor, Prof. Andrea Goldsmith, for their detailed and careful review that helped in improving the presentation of this paper.

REFERENCES

[1] K. H. Lee and D. P. Petersen, "Optimal linear coding for vector channels," *IEEE Trans. Commun.*, vol. 24, 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.-Aug. 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, pp. 3221–3231, Dec. 1994.

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

[5] H. Sampath, P. Stoica, and A. J. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, pp. 2198–2206, Dec. 2001.

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

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

[8] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Sig. Proc.*, vol. 50, no. 5, pp. 1051–1064, May 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] S. Zhou and B. Li, "BER criterion and codebook construction for finite-rate precoded spatial multiplexing with linear receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1653–1665, May 2006.

[11] R. W. Keyes, "Physical limits in digital electronics," *Proc. IEEE*, vol. 63, pp. 740–767, May 1975.

[12] J. D. Meindl and J. A. Davis, "The fundamental limit on binary switching energy for terascale integration (TSI)," *IEEE J. Solid-State Circuits*, vol. 35, no. 10, pp. 1515–1516, Oct. 2000.

[13] J. Rabaey, A. Chandrakasan, and B. Nikolic, *Digitial Integrated Circuits: A Design Perspective*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 2003.

[14] B. Razavi, *RF Microelectronics*. Upper Saddle River, NJ: Prentice—Hall, 1998.

[15] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[16] A. S. Y. Poon, R. W. Brodersen, and D. N. C. Tse, "Degrees of freedom in multiple-antenna channels: A signal space approach," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 523–536, Feb. 2005.

[17] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2632–2639, Sep. 2001.

[18] S. A. Jafar and A. J. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems with imperfect feedback," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, Jul. 2004.

[19] E. Jorswieck and H. Boche, "Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback," *IEEE Trans. Wireless Commun.*, vol. 52, no. 10, pp. 1654–1657, Oct. 2004.

[20] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "MIMO capacity through correlated channels in the presence of correlated interferers and noise," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2545–2561, Oct. 2003.

[21] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space-time block coding based on channel mean feedback," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 1599–1613, Oct. 2002.

[22] A. J. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.

[23] V. V. Veeravalli, Y. Liang, and A. M. Sayeed, "Correlated MIMO Rayleigh fading channels: Capacity, optimal signaling and asymptotics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2058–2072, Jun. 2005.

[24] A. M. Tulino, A. Lozano, and S. Verdú, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2491–2509, Jul. 2005.

[25] D. Gesbert, H. Bolcskei, D. A. Gore, and A. J. Paulraj, "Outdoor MIMO wireless channels: Models and performance prediction," *IEEE Trans. Commun.*, vol. 50, pp. 1926–1934, Dec. 2002.

[26] 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.

[27] X. Zhang, D. P. Palomar, and B. Ottersten, "Statistically robust design of linear MIMO transceivers," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3678–3689, Aug. 2008.

[28] Z. D. Bai, "Methodologies in spectral analysis of large dimensional random matrices, A review," *Statistica Sinica*, vol. 9, pp. 611–677, 1999.

[29] A. M. Tulino and S. Verdù, "Random matrices and wireless communications," *Found. Trends in Commun. Inf. Theory*, vol. 1, no. 1, Jun. 2004.

[30] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, "A new approach for capacity analysis for large dimensional multiantenna channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 3987–4004, Sep. 2008.

[31] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*. New York: Academic, 1979.

[32] 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.

[33] C.-N. Chuah, J. M. Kahn, and D. N. C. Tse, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

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

[35] J. W. Demmel, *Applied Numerical Linear Algebra*, 1st ed. Philadelphia, PA: SIAM, 1997.

[36] R. Mathias, "Spectral perturbation bounds for positive definite matrices," *SIAM J. Matrix Anal. Appl.*, vol. 18, no. 4, pp. 959–980, Oct. 1997.

[37] R. Bhatia, *Matrix Analysis*. New York: Springer-Verlag, 1997.

[38] V. Raghavan, A. M. Sayeed, and N. Boston, "When is limited feedback for transmit beamforming beneficial?," in *Proc. IEEE Int. Symp. Inf. Theory*, Sep. 2005, pp. 1544–1548.

[39] V. Raghavan, V. V. Veeravalli, and A. M. Sayeed, "Quantized multimode precoding in spatially correlated multiantenna channels," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 6017–6030, Dec. 2008.

[40] V. Raghavan, R. W. Heath Jr., and A. M. Sayeed, "Systematic codebook designs for quantized beamforming in correlated MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1298–1310, Sep. 2007.

[41] V. Raghavan, V. V. Veeravalli, and R. W. Heath Jr., "Reduced rank signaling in spatially correlated MIMO channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2007, pp. 1081–1085.

[42] A. Forenza, M. Mckay, A. Pandharipande, R. W. Heath Jr., and I. B. Collings, "Adaptive MIMO transmission for exploiting the capacity of spatially correlated MIMO channels," *IEEE Trans. Veh. Technol.*, vol. 56, no. 2, pp. 619–630, Mar. 2007.

[43] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[44] R. A. Durrett, *Probability: Theory and Examples*, 2nd ed. New York: Duxbury, 1995.

[45] Z. D. Bai and Y. Q. Yin, "Convergence to the semi-circle law," *Ann. Prob.*, vol. 16, no. 2, pp. 863–875, 1988.

[46] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 1st ed. New York: Wiley, 1960.

[47] V. L. Girko, *Theory of Random Determinants*. Boston, MA: Kluwer, 1990.

[48] V. L. Girko, *Theory of Linear Algebraic Equations with Random Coefficients*. New York: Allerton, 1996.

Vasanthan Raghavan (S'01–M'06) received the B.Tech degree in electrical engineering from the Indian Institute of Technology, Madras, in 2001, the M.S. and Ph.D. degrees in electrical and computer engineering in 2004 and 2006, respectively, and the M.A. degree in mathematics in 2005, all from the University of Wisconsin-Madison.

He is currently a Research Fellow with The University of Melbourne, Parkville, Australia. He was with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, from 2006 to 2009. His research interests span multiantenna communication techniques, quickest change detection, information theory, multihop networking, robust control, and random matrix theory.

Akbar M. Sayeed (S'89–M'97–SM'02) received the B.S. degree from the University of Wisconsin-Madison in 1991, and the M.S. and Ph.D. degrees from the University of Illinois at Urbana-Champaign in 1993 and 1996, all in electrical engineering.

He is currently Professor of Electrical and Computer Engineering, University of Wisconsin-Madison. He was a Postdoctoral Fellow at Rice University, Houston, TX, from 1996 to 1997. His current research interests include wireless communications, statistical signal processing, multidimensional communication theory, information theory, learning theory, time-frequency analysis, and applications in wireless communication networks and sensor networks.

Dr. Sayeed is a recipient of the Robert T. Chien Memorial Award (1996) for his doctoral work at Illinois, the NSF CAREER Award (1999), the ONR Young Investigator Award (2001), and the UW Grainger Junior Faculty Fellowship (2003). He is currently serving on the Signal Processing for Communications Technical Committee of the IEEE Signal Processing Society. He also served as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS from 1999 to 2002, and as the Technical Program Co-Chair for the 2007 IEEE Statistical Signal Processing Workshop and the 2008 IEEE Communication Theory Workshop.

Venugopal V. Veeravalli (S'86–M'92–SM'98–F'06) received the B.Tech degree in 1985 from the Indian Institute of Technology, Bombay, (Silver Medal Honors), the M.S. degree in 1987 from Carnegie-Mellon University, Pittsburgh, PA, and the Ph.D. degree in 1992 from the University of Illinois at Urbana-Champaign, all in electrical engineering.

He joined the University of Illinois at Urbana-Champaign in 2000, where he is currently a Professor in the Department of Electrical and Computer Engineering, and a Research Professor with the Coordinated Science Laboratory. He served as a Program Director for Communications Research at the U.S. National Science Foundation, Arlington, VA, from 2003 to 2005. He has previously held academic positions with Harvard University, Rice University, and Cornell University. His research interests include distributed sensor systems and networks, wireless communications, detection and estimation theory, and information theory.

He was on the Board of Governors of the IEEE Information Theory Society from 2004 to 2007. He was an Associate Editor for Detection and Estimation for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2000 to 2003, and an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 1999 to 2000. Among the awards he has received for research and teaching are the IEEE Browder J. Thompson Best Paper Award, the National Science Foundation CAREER Award, and the Presidential Early Career Award for Scientists and Engineers (PECASE).