Rate-Optimal MIMO Transmission with Mean and Covariance Feedback at Low SNRs

Ramy H. Gohary, Wessam Mesbah, and Timothy N. Davidson

***Abstract*—This paper considers a multiple-input–multiple-output (MIMO) wireless communication scenario in which the channel follows a block-fading model with a general spatially correlated complex Gaussian distribution with nonzero mean. We derive an explicit characterization of the (ergodic) rate-optimal input covariance for systems that operate at low signal-to-noise ratios (SNRs). In particular, we obtain a closed-form expression for the matrix whose principal eigenvector yields the optimal beamforming direction. For the class of nonzero-mean channels with Kronecker-structured covariance, we also derive a threshold on the input signal power below which the low-SNR approximation is accurate. Our numerical results show that significant improvements in the low-SNR achievable rate can be obtained by (jointly) exploiting the mean and covariance of the channel model. Our results also show that these improvements can be extended to moderate-to-high SNRs by signaling along the low-SNR optimal eigenbasis with optimized power allocation.**

***Index Terms*—Correlated channel with nonzero mean, ergodic capacity, low signal-to-noise ratio (SNR) signaling, multiple-input–multiple-output (MIMO) systems, noncentral Wishart distribution, statistical channel state information (CSI).**

I. Introduction

The availability of channel state information (CSI) at the transmitter of a multiple-input–multiple-output (MIMO) wireless communication system that operates at low signal-to-noise ratios (SNRs) can have a fundamental impact on the ergodic achievable rate (e.g., [1]–[4]). However, it is often quite difficult to provide the transmitter with accurate information about the actual channel realization (i.e., instantaneous CSI), and it may be more appropriate to consider systems in which the transmitter has access only to statistical information about the channel (e.g., [1]). Indeed, there have been several analyses of the impact of channel correlation on low-SNR communication (e.g., [2], [3], [5], and [6]). However, these analyses have been made under rather restricted models for the channel statistics. The goal of this paper is to provide an explicit characterization of the low-SNR rate-optimal input covariance for an arbitrarily correlated channel model with nonzero mean.

In some of the early work on (ergodic) rate-optimal signaling for wireless systems with multiple antennas, optimal signaling strategies for multiple-input–single-output systems were developed [7], [8] for two classes of channels: one in which the channel mean is known at the transmitter and the channel covariance is assumed to be a scaled identity and the other in which the channel covariance is known at the transmitter and the mean is assumed to be zero. That work was extended to MIMO systems in [9] and [10], where rate-optimal transmission strategies were developed for the case in which the channel has zero mean, correlated rows, and independent columns. For that scenario, necessary and sufficient conditions under which the covariance matrix of the optimal signaling scheme is of rank 1 (and, hence, can be implemented by beamforming) were given in [9] and [10]. Using a similar technique to that in [9], rate-optimal signaling strategies and necessary and sufficient conditions for the optimality of beamforming for the case in which both the columns and the rows of the channel matrix are correlated were developed in [11]. For that case, a closed-form expression for the exact channel capacity was derived in [12]. In addition, sufficient (but not necessary) conditions under which beamforming is rate optimal for MIMO channels with nonzero mean and scaled identity covariance were derived in [13].

While MIMO channel models with zero mean, correlated columns, and correlated rows are sufficient to characterize some practical communication scenarios, the extension of the aforementioned analyses to the more general correlation models that typically occur in practice can be quite unwieldy. A characterization of the structure that the optimal input covariance matrix must possess for a zero-mean arbitrarily correlated channel model was provided in [14], but that work did not contain an explicit construction of this matrix.

In this paper, we consider the design of the rate-optimal input covariance for low-SNR signaling over an arbitrarily correlated channel with nonzero mean. We derive an explicit closed-form expression for the low-SNR ergodic capacity for this channel model, and we use that expression to derive an explicit characterization of the optimal input covariance. Similar to [2], we show that whenever the maximum eigenvalue of a certain matrix is distinct, beamforming remains optimal in the presence of a nonzero mean and an arbitrary correlation. Computing the beam direction using the approach in [2] can be quite difficult for general correlation models, and in [4], this direction was computed only for the so-called Unitary-Independent-Unitary model. The explicit characterization provided herein allows us to avoid these difficulties and enables us to obtain a closed-form expression for the matrix whose principal eigenvector yields the optimal beam direction for general correlation models with nonzero mean. The closed-form expression for this matrix also allows us to explore how the low-SNR characterization can be used to synthesize high-rate signaling strategies at moderate-to-high SNRs. In particular, we will show that by optimally allocating power to all the eigenvectors of this matrix, one can obtain higher ergodic rates than several existing schemes over a broad range of SNRs.

II. System Model

We consider MIMO systems with M transmit and N receive antennas and a channel matrix H that follows a general complex Gaussian block-fading model with mean $\bar{H}$ and covariance $\Phi = E_H\{\mathrm{vec}(H - \bar{H})\mathrm{vec}(H - \bar{H})^\dagger\}$, where $\mathrm{vec}(\cdot)$ denotes the column-stacking operator, and $(\cdot)^\dagger$ denotes the Hermitian transpose. Channel realizations from this model can be expressed as

$$\mathrm{vec}(H) = \Phi^{1/2}\mathrm{vec}(H_w) + \mathrm{vec}(\bar{H}) \tag{1}$$

where H_w is a (white) matrix with independent identically distributed circularly symmetric complex Gaussian entries with zero mean and unit variance. While the key results in this paper are derived for the general model in (1), in some cases, further insight can be obtained

Manuscript received December 30, 2007; revised December 24, 2008. First published February 24, 2009; current version published August 14, 2009. This work was supported in part by the Government of Ontario under the Premier's Research Excellence Award. The work of T. N. Davidson was supported by the Canada Research Chair Program. The review of this paper was coordinated by Prof. H.-H. Chen.

R. H. Gohary was with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4L8, Canada. He is now with the Communications Research Centre, Ottawa, ON K2H 8S2, Canada (e-mail: rgohary@crc.ca).

W. Mesbah was with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4L8, Canada. He is now with the Department of Electrical and Computer Engineering, Texas A&M University at Qatar, Doha, Qatar (e-mail: wessam.mesbah@qatar.tamu.edu).

T. N. Davidson is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4L8, Canada (e-mail: davidson@mcmaster.ca).

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/TVT.2009.2015670

by considering the more tractable Kronecker model (e.g., [1]) and its generalization in [14], in which the channel covariance matrix is separable into transmitter and receiver components. Realizations from a channel model with nonzero mean and a general separable covariance take the form

$$H = \sum_{s=1}^{S} R_s^{1/2} H_w T_s^{1/2} + \bar{H} \tag{2}$$

where $\{R_s\}_{s=1}^S$ and $\{T_s\}_{s=1}^S$ are sets of Hermitian positive semidefinite matrices of sizes $N \times N$ and $M \times M$ that characterize the correlation among the elements of the receiver and the correlation among the elements of the transmitter, respectively. The covariance of the channel model in (2) is $\Phi = \sum_{s=1}^S T_s^T \otimes R_s$, and the Kronecker model is obtained by setting $S = 1$.

III. Low-SNR Approximation of Ergodic Capacity

Consider coherent communication over the channel model in (1) in the presence of zero-mean additive white Gaussian noise (AWGN) with normalized spatial covariance $K = I$. When only the mean $\bar{H}$ and the covariance Φ are known at the transmitter, the ergodic capacity is [15]

$$C = \max_{Q \succeq 0,\, \mathrm{Tr}(Q)=P} E_H \left\{ \log\det(I + HQH^\dagger) \right\} \tag{3}$$

where Q is the covariance matrix of the input signal, and P is the total power budget. As is well known (e.g., [1]), the expression in (3) can be applied to systems with an arbitrary nonsingular noise covariance K by simply replacing H by $\acute{H} = K^{-1/2}H$. Since $\acute{\bar{H}} = K^{-1/2}\bar{H}$ and $\acute{\Phi} = (I \otimes K^{-1/2})\Phi(I \otimes K^{-1/2})$, the work herein is immediately applicable in that case, but for simplicity, we will focus on the case in which $K = I$.

To facilitate the analysis of (3), we observe that for a positive definite matrix $X \succ 0$, $\log\det(X) = \mathrm{Tr}(\log(X))$, where the log function of a positive definite matrix is defined as the inverse of the matrix exponential (cf. [16]). Using this result, for a Hermitian matrix A with eigenvalues in $(-1, 1)$, $\log\det(I + A)$ can be expanded as

$$\begin{aligned} \log\det(I + A) &= \mathrm{Tr}\left(\log(I + A)\right) \\ &= \mathrm{Tr}(A) - \frac{1}{2}\mathrm{Tr}(A^2) + \frac{1}{3}\mathrm{Tr}(A^3) + \cdots. \end{aligned} \tag{4}$$

The matrix $HQH^\dagger$ is Hermitian and positive semidefinite, and for sufficiently low input signal power, its maximum eigenvalue, $\lambda_{\max}(HQH^\dagger)$, satisfies $\lambda_{\max}(HQH^\dagger) \leq \epsilon \ll 1$ with a high probability. Using the fact that $\lambda_{\max}(HQH^\dagger) \leq \lambda_{\max}(Q)\lambda_{\max}(H^\dagger H)$ and that the power constraint implies that $\lambda_{\max}(Q) \leq P$, it can be seen that if the expansion in (4) is substituted into (3), then for sufficiently small values of P, the resulting expression is dominated by the linear term. That is, at low SNRs, one can approximate the capacity by

$$C \approx \max_{Q \succeq 0,\, \mathrm{Tr}(Q)=P} E_H \left\{ \mathrm{Tr}(HQH^\dagger) \right\}. \tag{5}$$

Since the nth term in (4) can be thought of as the nth derivative of $\log\det(I + tA)$ with respect to the scalar t at $t = 0$, the result in [3, Sec. III] can be used to show that maximizing the term in (5) is equivalent to minimizing $(E_b/N_0)_{\min}$, which is the minimum energy per bit required to communicate at a positive rate. In Section VI, we will derive a (conservative) bound on P, below which, (5) is reasonably accurate for a nonzero-mean channel with Kronecker-structured covariance.

IV. Explicit Low-SNR Ergodic Capacity

In this section, we obtain an explicit expression for the low-SNR ergodic capacity in (5), and in Section V, we will use this explicit expression to cast the problem of optimizing the input covariance matrix as a linear program that can be solved analytically.

Using the model in (1) and defining $\tilde{H} = \mathrm{unvec}(\Phi^{1/2}\mathrm{vec}(H_w))$, where $\mathrm{unvec}(\cdot)$ denotes the inverse of the $\mathrm{vec}(\cdot)$ operator, the expression $E_H\{\mathrm{Tr}(HQH^\dagger)\}$ in (5) can be written as

$$\begin{aligned} &E_H \left\{ \mathrm{Tr}(HQH^\dagger) \right\} \\ &\quad = E_{H_w} \left\{ \mathrm{Tr}\left(Q^{1/2}(\tilde{H}^\dagger + \bar{H}^\dagger)(\tilde{H} + \bar{H})Q^{1/2} \right) \right\} \\ &\quad = E_{H_w} \left\{ \mathrm{Tr}(Q^{1/2}\tilde{H}^\dagger\tilde{H}Q^{1/2}) \right\} + \mathrm{Tr}(\bar{H}^\dagger\bar{H}Q). \end{aligned} \tag{6}$$

We can write the first term on the right-hand side (RHS) of (6) as

$$\begin{aligned} &E_{\tilde{H}} \left\{ \mathrm{Tr}(Q^{1/2}\tilde{H}^\dagger\tilde{H}Q^{1/2}) \right\} \\ &\quad = E_{\tilde{H}} \left\{ \left(\mathrm{vec}(\tilde{H})\right)^\dagger (Q^{T/2} \otimes I_N)(Q^{T/2} \otimes I_N)\mathrm{vec}(\tilde{H}) \right\} \quad (7) \\ &\quad = \mathrm{Tr}\left(\Phi\left(Q^T \otimes I_N \right) \right) \quad (8) \end{aligned}$$

where A^T denotes the transpose of the matrix A, respectively, and in (7) we have used the identity $\mathrm{Tr}(A^\dagger B) = (\mathrm{vec}(A))^\dagger \mathrm{vec}(B)$. Using (8), we have that for the general model in (1), the low-SNR rate-optimal covariance matrix maximizes

$$\begin{aligned} &\mathrm{Tr}\left(\Phi(Q^T \otimes I_N) \right) + \mathrm{Tr}(\bar{H}^\dagger\bar{H}Q) \quad (9) \\ &\quad = \mathrm{Tr}\left(\Phi(Q^T \otimes I_N) \right) + \frac{1}{N}\mathrm{Tr}(\bar{H}^T\bar{H}^* \otimes I_N)(Q^T \otimes I_N) \\ &\quad = \mathrm{Tr}\left(X(Q \otimes I_N) \right) \quad (10) \end{aligned}$$

where $X = (\Phi^T + (1/N)(\bar{H}^\dagger\bar{H} \otimes I_N))$. To simplify this expression, we observe that the ijth $N \times N$ block of the $NM \times NM$ matrix $(Q \otimes I_N)$ is in the form of a scaled identity, i.e., $[Q \otimes I_N]_{(i-1)N+r,(j-1)N+s} = q_{ij}\delta_{rs}$, where q_{ij} denotes the ijth entry of Q, and δ_{rs} is the Kronecker delta. Using this expression, we have

$$\begin{aligned} \mathrm{Tr}\left(X(Q \otimes I_N) \right) &= \sum_{i,j=1}^{M} \sum_{r,s=1}^{N} q_{ij}\delta_{rs}X_{(i-1)N+r,(j-1)N+s} \\ &= \sum_{i,j=1}^{M} q_{ij}\,\mathrm{Tr}\left(X_{[i,j]} \right) \end{aligned} \tag{11}$$

where $X_{[i,j]}$ denotes the ijth $N \times N$ block of X. Now, we can write (10) as $\mathrm{Tr}(ZQ)$, where

$$[Z]_{ij} = \mathrm{Tr}\left(X_{[i,j]} \right). \tag{12}$$

Using the result in Appendix A, it can be shown that since Z is a matrix whose elements are the traces of blocks of a positive semidefinite matrix, it is positive semidefinite.

V. Optimal Input Covariance Matrix

We now proceed to find the input covariance matrix that maximizes the low-SNR ergodic achievable rate in (5). That is, we solve the optimization problem

$$\max_{Q \succeq 0,\, \mathrm{Tr}(Q) \leq P} \quad \mathrm{Tr}(ZQ) \tag{13}$$

where Z was defined in (12). For the separable channel model in (2), the expression for Z simplifies to $\sum_{r,s=1}^S \mathrm{Tr}(R_r^{1/2}R_s^{1/2})T_r^{1/2}T_s^{1/2} + \bar{H}^\dagger\bar{H}$. The first step in the solution of (13) is to use the result in [17, Example 7.4.13] to show that it is sufficient to consider matrices Q with the same eigenbasis as Z (see [9] for a related derivation). That is, if $Z = U_Z\Lambda_Z U_Z^\dagger$ denotes the ordered eigendecomposition of Z, it is sufficient to consider matrices Q of the form $U_Z\Lambda_Q U_Z^\dagger$, where

Λ_Q denotes the diagonal matrix of (nonnegative) eigenvalues of Q. In that case, the optimization problem in (13) can be cast as the linear program $\max_{\lambda_{Q_i}\geq 0, \Sigma_{i=1}^M \lambda_{Q_i}\leq P} \sum_{i=1}^M \lambda_{Z_i}\lambda_{Q_i}$, where λ_{Z_i} and λ_{Q_i} denote the ith eigenvalues of Z and Q, respectively. If the maximum eigenvalue of Z is distinct, the optimal eigenvalues of Q are $\lambda_{Q_1} = P$ and $\lambda_{Q_2} = \cdots = \lambda_{Q_M} = 0$. That is, beamforming along the principal eigenvector of Z is sufficient for rate-optimal communication at low SNRs. For the case in which the largest eigenvalue of Z has multiplicity greater than 1, any partitioning of power in the direction of the eigenvectors corresponding to these eigenvalues is optimal up to the first-order approximation. However, for up to the second-order optimality, power may have to be carefully distributed across these directions [3].

Remark: For the case of channels with Kronecker-structured covariance, the optimum beam direction is along the principal eigenvector of $(\mathrm{Tr}(R)T + \bar{H}^\dagger\bar{H})$. If the channel is zero mean, the matrix R does not affect the optimal beam direction, and in agreement with the result in [11], this direction is along the principal eigenvector of T. In contrast, if the channel is nonzero mean, this expression suggests that $\mathrm{Tr}(R)$ acts as a weight that controls the relative impact of the eigenvectors of T and that of the eigenvectors of $\bar{H}^\dagger\bar{H}$ on the optimal beam direction.

VI. Accuracy of the Low-SNR Approximation

In this section, we derive a threshold on the input power below which the low-SNR approximation of the ergodic capacity in (5) is reasonably accurate. We consider the nonzero-mean channel model with Kronecker-structured covariance with positive definite T and R. Our strategy for computing this threshold is to ensure that $\lambda_{\max}(HQH^\dagger) \leq \epsilon$ with a high probability. To do that, we compute a lower bound on $\Pr(\lambda_{\max}(HQH^\dagger) \leq \epsilon)$ and use this bound to determine the required threshold.

We begin by observing that $\lambda_{\max}(HQH^\dagger) \leq \lambda_{\max}(Q)\lambda_{\max}(H^\dagger H)$. Substituting $\breve{H} = R^{1/2}H_w + \bar{H}T^{-1/2}$ into this expression, we have

$$\begin{aligned}\lambda_{\max}(HQH^\dagger) &\leq P\lambda_{\max}(T^{1/2}\breve{H}^\dagger\breve{H}T^{1/2})\\ &\leq P\lambda_{\max}(T)\lambda_{\max}(R)\lambda_{\max}(R^{-1}\breve{H}\breve{H}^\dagger) \qquad (14)\end{aligned}$$

and hence

$$\begin{aligned}&\Pr\left(\lambda_{\max}(HQH^\dagger) \leq \varepsilon\right)\\ &\qquad \geq \Pr\left(P\lambda_{\max}(T)\lambda_{\max}(R)\lambda_{\max}(R^{-1}\breve{H}\breve{H}^\dagger) \leq \varepsilon\right). \qquad (15)\end{aligned}$$

Therefore, a conservative approach for ensuring that the left-hand side of (15) is greater than some $\gamma \in [0, 1)$ is to restrict the input power so that the RHS of (15) is greater than γ. In Appendix B, we show how a result from [18] can be used to construct an explicit expression for the RHS of (15).

VII. Numerical Examples

In this section, we provide some numerical examples that illustrate the utility of the explicit characterization of the optimal low-SNR signaling strategy derived in Section V. We consider a MIMO system with $M = N = 5$ and an identity noise covariance matrix.

We first consider an instance of the general channel model in (1). The channel mean $\bar{H}$ was randomly chosen, and its Frobenius norm was set to be equal to 0.866. The matrix Φ in the general correlation model in (1) was also randomly chosen, and its trace was set to 2.5. The resulting matrix Z in (12) appears in (16), shown at the bottom of the next page. In Fig. 1, we provide a plot of the ergodic achievable rates for this channel model under different signaling strategies against the transmitted signal power. To provide a benchmark for these results, in Fig. 1, we also provide an upper bound that corresponds to the

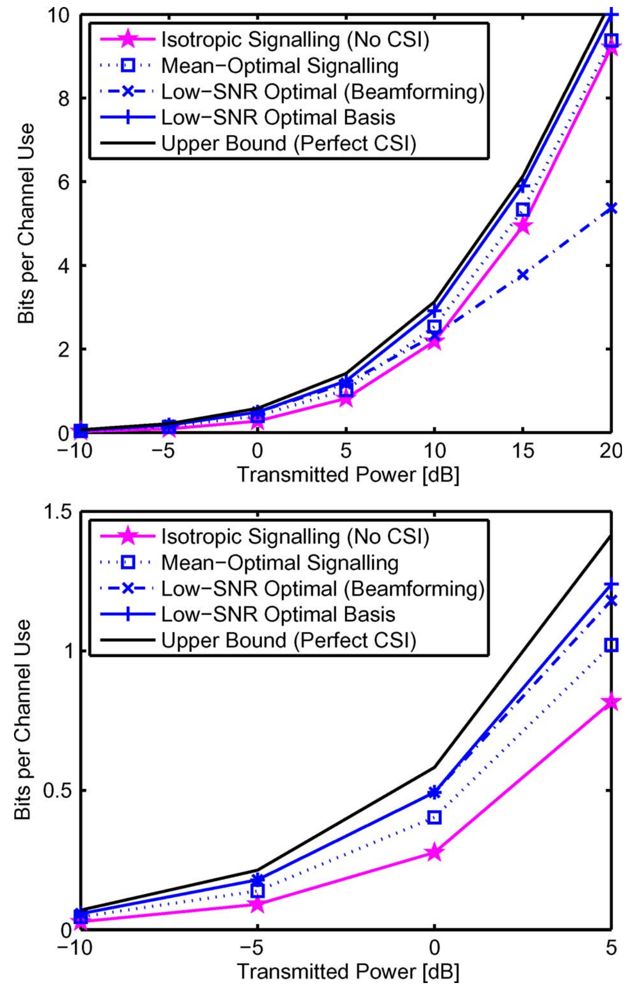

Fig. 1. Achievable rates versus transmitted signal power for different transmission schemes over an arbitrarily correlated channel model with nonzero mean. The signaling schemes are as follows: isotropic signaling (Strategy 1), mean-optimal signaling (Strategy 2), optimal signaling for low SNRs (beamforming, Strategy 3), and signaling along low-SNR optimal basis (Strategy 4). The lower figure provides the details of the low-SNR region of the upper figure.

maximum rate that would be achievable had the channel realizations been perfectly known to the transmitter (i.e., instantaneous CSI). The signaling strategies considered in Fig. 1 are the following.

1) *Uniform power loading* [15]. The transmitter ignores all channel information and transmits isotropically, i.e., $Q = (P/M)I$.
2) *Mean-optimal signaling* [15]. The transmitter ignores the channel covariance information and treats the mean as if it were the actual channel. Hence, transmission takes place along the eigenvectors of $\bar{H}^\dagger\bar{H}$ with the eigenvalues of Q chosen to "water-fill" over those of $\bar{H}^\dagger\bar{H}$. (At low SNRs, this results in beamforming along the principal eigenvector of $\bar{H}^\dagger\bar{H}$.)
3) *Low-SNR optimal beamforming (proposed herein)*. The transmitter employs the low-SNR rate-optimal strategy derived herein, that is, beamforming along the principal eigenvector of Z [cf. (13)].
4) *Signaling along the low-SNR optimal basis (proposed here)*. The transmitter signals along the eigenvectors of Z without restricting the input covariance matrix to be rank 1.[1] For each

[1]The development of this strategy was motivated by the fact that for zero-mean channels with Kronecker-structured covariance, the set of optimal signaling directions is the same for all SNRs [11] and can be obtained from the low-SNR analysis.

input power constraint, the eigenvalues of Q are found by solving the optimization problem in (3) (with $U_Q = U_Z$) using a gradient-based stochastic optimization technique [19]. In Appendix C, we provide a brief description of the technique and an analytic expression for the stochastic gradient. (At low input powers, this strategy reduces to Strategy 3.)

As shown in Fig. 1, at low SNRs, Strategies 3 and 4 provide higher achievable rates than Strategies 1 and 2. In particular, at an input power of 0 dB, Strategies 3 and 4 yield near identical rates[2] of about 0.5 bits per channel use (bpcu), whereas Strategies 1 and 2 yield rates of about 0.27 and 0.4 bpcu, respectively. At higher SNRs, Strategy 4 yields higher rates than those yielded by Strategies 1–3. For instance, at an input power of 20 dB, Strategies 1 and 2 yield approximately the same rate of 9 bpcu, whereas Strategy 4 yields a rate of 10 bpcu. For this input power, Strategy 3 yields about 5.8 bpcu, which illustrates the impact of restricting the input covariance to be rank 1 at high SNRs.

While we cannot claim optimality of Strategy 4, we have performed numerous experiments, all of which suggest that Strategy 4 is capable of achieving rates that are higher than those that can be achieved under the considered signaling strategies. It is worth noting that since Strategy 4 is based on stochastic optimization, it involves greater computational effort than Strategies 1–3. However, because this optimization is only over the M eigenvalues of Q, it requires significantly less computational effort than computing the optimal input covariance, which involves stochastic optimization over the M^2 real parameters that define Q.

We now provide a second numerical example that illustrates the range of input powers for which the low-SNR expansion is guaranteed to be accurate. We consider a nonzero-mean Kronecker model for the channel, in which the matrices T, R, and $\bar{H}$ were randomly chosen. The resulting matrices are given in (17a)–(17c), shown at the bottom of the page. In Fig. 2, we have plotted the cumulative distribution function (cdf) of $P\lambda_{\max}(T)\lambda_{\max}(R)\lambda_{\max}(R^{-1}\breve{H}\breve{H}^\dagger)$ [cf. (14)] for different values of P. From this plot, it can seen that the condition $\lambda_{\max}(HQH^\dagger) \ll 1$ is satisfied with a high probability for input powers of up to about -15 dB. In Fig. 3, we have plotted the achievable rates using Strategies 1–4. Since, in this figure, we are considering the Kronecker model, we also consider two additional strategies.

5) *Covariance-optimal signaling* [11]. The transmitter ignores the mean information and signals along the eigenvectors of the transmit covariance matrix. Similar to Strategy 4, in this strategy, the transmitter uses stochastic optimization to determine the eigenvalues of Q.
6) *Capacity-achieving signaling* [20]. The optimal input covariance is generated using the fixed-point algorithm in [20]. As each iteration involves an expectation that is numerically evaluated using Monte Carlo integration, this algorithm is rather computationally expensive.

The low-SNR beamforming curve in Fig. 3 suggests that the guidance provided by Fig. 2 is rather conservative. While Fig. 2 suggests that the low-SNR approximation of the ergodic capacity will be accurate for up to -15 dB, Fig. 3 suggests that the designs based on the low-SNR approximation are at least as good as the other considered strategies for up to about 0 dB. Fig. 3 also shows that, similar to the case in Fig. 1, using Strategy 4 yields rates that are consistently higher than the rates yielded by the other signaling strategies. In fact, Fig. 3 shows that the rate yielded by Strategy 4 almost coincides with the ergodic channel capacity obtained using the (computationally expensive) technique in [20] (see Strategy 6).

VIII. Conclusion

We have provided an explicit characterization of the low-SNR rate-optimal input covariance for nonzero-mean arbitrarily correlated

[2]This is consistent with the low-SNR optimality of Strategy 3.

$$Z = \begin{bmatrix} 0.0970 & 0.0067 - j0.0163 & -0.0041 + j0.0091 & 0.0054 - j0.0125 & -0.0215 + j0.0244 \\ 0.0067 + j0.0163 & 0.0397 & -0.0143 + j0.0088 & 0.0047 - j0.0173 & 0.0284 + j0.0036 \\ -0.0041 - j0.0091 & -0.0143 - j0.0088 & 0.0595 & -0.0035 - j0.0114 & -0.0140 - j0.0436 \\ 0.0054 + j0.0125 & 0.0047 + j0.0173 & -0.0035 + j0.0114 & 0.1163 & -0.0535 + j0.0177 \\ -0.0215 - j0.0244 & 0.0284 - j0.0036 & -0.0140 + j0.0436 & -0.0535 - j0.0177 & 0.1765 \end{bmatrix} \quad (16)$$

$$T = \begin{bmatrix} 0.4150 & -0.0606 - j0.0305 & -0.0535 - j0.0178 & 0.0404 - j0.0184 & -0.2058 + j0.1730 \\ -0.0606 + j0.0305 & 0.0865 & -0.0172 + j0.0356 & 0.0184 - j0.0598 & 0.0962 + j0.0047 \\ -0.0535 + j0.0178 & -0.0172 - j0.0356 & 0.2249 & -0.0088 - j0.0056 & -0.0258 - j0.2195 \\ 0.0404 + j0.0184 & 0.0184 + j0.0598 & -0.0088 + j0.0056 & 0.3733 & -0.3320 + j0.1602 \\ -0.2058 - j0.1730 & 0.0962 - j0.0047 & -0.0258 + j0.2195 & -0.3320 - j0.1602 & 0.7874 \end{bmatrix} \quad (17a)$$

$$R = \begin{bmatrix} 0.2197 & -0.1465 - j0.1521 & -0.0121 - j0.0365 & 0.0445 - j0.1153 & -0.0381 - j0.0692 \\ -0.1465 + j0.1521 & 0.7203 & 0.1110 - j0.1539 & 0.1769 + j0.2591 & 0.0204 + j0.1715 \\ -0.0121 + j0.0365 & 0.1110 + j0.1539 & 0.3686 & -0.0231 + j0.0377 & 0.1018 + j0.0161 \\ 0.0445 + j0.1153 & 0.1769 - j0.2591 & -0.0231 - j0.0377 & 0.2663 & 0.0668 + j0.0577 \\ -0.0381 + j0.0692 & 0.0204 - j0.1715 & 0.1018 - j0.0161 & 0.0668 - j0.0577 & 0.1648 \end{bmatrix} \quad (17b)$$

$$\bar{H} = \begin{bmatrix} -0.0242 - j0.1315 & -0.0921 - j0.2263 & 0.0056 + j0.1380 & 0.0057 + j0.0212 & -0.1058 - j0.1823 \\ -0.0359 + j0.0077 & 0.0179 + j0.0287 & -0.0009 - j0.0549 & 0.1140 - j0.2008 & 0.0827 + j0.0295 \\ 0.0335 + j0.1153 & 0.0580 - j0.0103 & -0.1038 + j0.0373 & -0.1933 + j0.1363 & 0.1324 + j0.2397 \\ -0.0109 - j0.0186 & -0.0051 - j0.0679 & -0.0829 + j0.0435 & 0.0887 - j0.0296 & 0.0892 + j0.1579 \\ 0.0982 + j0.2349 & 0.1348 + j0.0578 & -0.0655 + j0.1762 & 0.2438 - j0.1231 & -0.0164 - j0.0169 \end{bmatrix} \quad (17c)$$

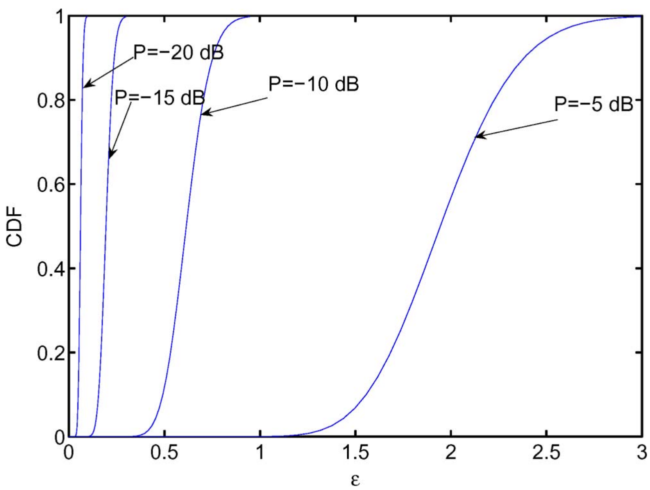

Fig. 2. CDF of $P\lambda_{\max}(T)\lambda_{\max}(R)\lambda_{\max}(R^{-1}\breve{H}\breve{H}^{\dagger})$ for different values of P.

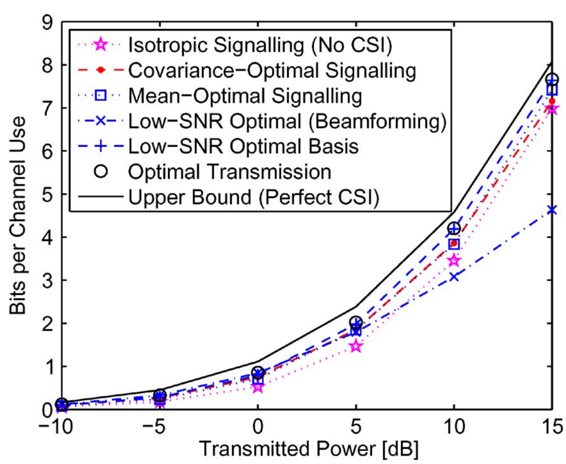

Fig. 3. Achievable rates versus transmitted signal power for different transmission schemes over a correlated channel model with Kronecker-structured covariance and nonzero mean.

MIMO channels. We have also provided a method for determining a threshold on the input power below which the low-SNR approximation is accurate. For higher input powers, we proposed a signaling strategy whereby the transmitter retains the low-SNR signaling directions but optimizes the power allocation to these directions to maximize the ergodic rate. We have shown that this novel strategy can lead to achievable rates that are close to the ergodic capacity.

Appendix A
Positive Definiteness of Z in (12)

To show that Z is positive semidefinite, it is sufficient to prove the following lemma.

Lemma 1: Let A be an $MN \times MN$ positive semidefinite matrix, and partition it into $M \times M$ blocks, each of size $N \times N$. The matrix constructed by replacing each block by its main diagonal (see the illustration in Fig. 4) is positive semidefinite. Moreover, the $M \times M$ matrix constructed by replacing each block by its trace is also positive semidefinite.

Proof: Let E_i denote the $MN \times MN$ all-zero matrix with the ith entry on the main diagonal replaced by 1, and let the matrix constructed by replacing each block by its main diagonal be denoted by G. Then, G can be expressed as $G = \sum_{i=1}^{N}(\sum_{j=1}^{M} E_{(j-1)N+i})A(\sum_{j=1}^{M} E_{(j-1)N+i}^{T})$. Since A is positive

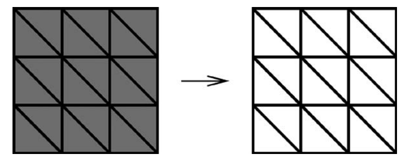

Fig. 4. Pictorial view of the replacement of each $N \times N$ block of A by its main diagonal.

semidefinite, then so is $(\sum_{j=1}^{M} E_{(j-1)N+i})A(\sum_{j=1}^{M} E_{(j-1)N+i}^{T})$. Invoking the fact that the sum of positive semidefinite matrices is also positive semidefinite completes the proof of the first statement of the lemma. To prove the second part, let F denote the $M \times M$ matrix constructed by replacing each block by its trace, and let $\mathbf{1}_N$ be the $N \times 1$ vector in which all entries are equal to unity. Then, $F = (I_M \otimes \mathbf{1}_N^T)G(I_M \otimes \mathbf{1}_N)$. Since we now know that G is positive semidefinite, then F is also positive semidefinite. ■

Appendix B
Computation of the Bound in (15)

To provide an explicit expression for the RHS of (15), we will make use of the follow lemma.

Lemma 2 [18, Th. 1]: Let X be an $m \times n$ matrix whose columns are independent m-variate complex Gaussian vectors with covariance matrix Σ and mean $E\{X\} = \bar{X}$. Let $s = \min\{m, n\}$, $t = \max\{m, n\}$, and $0 < \phi_1 \cdots < \phi_s$ be s nonzero eigenvalues of $S_X = \Sigma^{-1}XX^{\dagger}$. If $\bar{X}^{\dagger}\Sigma^{-1}\bar{X}$ has s nonzero distinct eigenvalues $0 < \theta_1 \cdots < \theta_s$, then the cdf of the largest eigenvalue ϕ_s of S_X is given by

$$\Pr(\phi_s \leq x) = \frac{e^{-\Sigma_i \theta_i}}{\det(V)\,(\Gamma(t-s+1))^{s}} \det\left(\Psi(x)\right) \tag{18}$$

where $\Gamma(\cdot)$ denotes the Gamma function, $\det(V)$ is the determinant of the Vandermonde matrix

$$V = \begin{bmatrix} \theta_1^{s-1} & \theta_1^{s-2} & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ \theta_s^{s-1} & \theta_s^{s-2} & \cdots & 1 \end{bmatrix} \tag{19}$$

and $\Psi(x)$ is an $s \times s$ matrix function of $x \in (0, \infty)$, whose entries are given in the following equation:[3]

$$\begin{aligned} [\Psi(x)]_{ij} =& \Gamma(t-i+1)\,{}_1F_1(t-i+1; t-s+1; \theta_j) \\ & - e^{-\theta_j}\Gamma(t-s+1)\sum_{\ell=1}^{s-i+1}\frac{(s-i)!}{(\ell-1)!} \\ & \times \binom{t-i}{s-i-\ell+1}\theta_j^{\ell-1}Q_{t-s+\ell}(\sqrt{2\theta_j}, \sqrt{2x}) \\ & - e^{-x}\Gamma(t-s+1)\sum_{\ell=1}^{s-i}\sum_{k=0}^{s-i-\ell} 2^{-\ell-k} \\ & \times \frac{(s-i-1-k)!}{(\ell-1)!}\binom{t-i}{s-i-\ell-k} \\ & \times (\sqrt{2\theta_j})^{s+\ell-t-1}(\sqrt{2x})^{t-s+2k+\ell+1} \\ & \times I_{t-s+\ell-1}(2\sqrt{\theta_j x}), \qquad i, j = 1, \ldots, s. \end{aligned} \tag{20}$$

[3]The expression for $\Psi(x)$ given in [18, eq. (4)] contains a number of typographical errors that have been corrected herein.

In (20), ${}_1F_1(\cdot;\cdot;\cdot)$ is the confluent hypergeometric function, $Q_p(\cdot,\cdot)$ is the pth-order generalized Marcum Q-function, and $I_p(\cdot)$ is the pth-order modified Bessel function of the first kind. ■

To use Lemma 2 to derive a bound on $\Pr(\lambda_{\max}(HQH^\dagger) \le \epsilon)$, we observe that $E\{\breve{H}\} = \bar{H}T^{-1/2}$. Furthermore, the matrix $\breve{H}$ has independent columns and correlated rows. The covariance matrix of the ith column of $\breve{H}$, $[\breve{H}]_i$ is given by

$$E\left\{\left(\breve{H}_i - \bar{H}[T^{-1/2}]_i\right)\left(\breve{H}_i - \bar{H}[T^{-1/2}]_i\right)^\dagger\right\} = E\left\{R^{1/2}[H_w]_i[H_w]_i^\dagger R^{1/2}\right\} = R. \quad (21)$$

Using this observation, it is now apparent that $\breve{H}$ satisfies the conditions of Lemma 2 with $\Sigma = R$ and $\bar{X} = \bar{H}T^{-1/2}$. Hence, one can use (14) along with the result in (18) to obtain an explicit expression for the bound in (15).

APPENDIX C
STOCHASTIC OPTIMIZATION OF THE EIGENVALUES OF Q IN STRATEGY 4

The optimization problem to be solved in Strategy 4 is $\max_{\Lambda_Q} E_H\{\log\det(I + \breve{H}\Lambda_Q\breve{H}^\dagger)\}$, where $Q = U_Q\Lambda_Q U_Q^\dagger$, $\breve{H} = HU_Q$, and $U_Q = U_Z$. We will solve this problem using stochastic gradient ascent [19], in which the iteration is updated by taking a step in the direction of the gradient associated with a single realization of the channel, i.e., the stochastic gradient. The realization is independently chosen at each iteration, and the algorithm converges if the step size is progressively reduced using the techniques outlined in [19].

To efficiently implement the stochastic gradient-ascent algorithm, we now derive an explicit expression for the stochastic gradient. The kth entry of the stochastic gradient is

$$\begin{aligned}\frac{\partial}{\partial\lambda_{Q_k}}\log\left(\det(I + \breve{H}\Lambda_Q\breve{H}^\dagger)\right) &= \sum_{i,j=1}^{N} \frac{\partial[\breve{H}\Lambda_Q\breve{H}^\dagger]_{ij}}{\partial\lambda_{Q_k}} \left.\frac{\partial\log\det(W)}{\partial w_{ij}}\right|_{W=I+\breve{H}\Lambda_Q\breve{H}^\dagger} \\ &= \mathrm{Tr}\left(\frac{\partial(\breve{H}\Lambda_Q\breve{H}^\dagger)}{\partial\lambda_{Q_k}} \left.\frac{\partial\log\det(W)}{\partial W}\right|_{W=I+\breve{H}\Lambda_Q\breve{H}^\dagger}\right) \quad (22)\end{aligned}$$

where λ_{Q_k} is the k diagonal entry of Λ_Q, and w_{ij} is the ijth entry of W. Now, if we use $\breve{h}_r$ to denote the rth column of $\breve{H}$, the partial derivative of the quadratic term in the trace can be expressed as $(\partial(\sum_{r=1}^{N}\lambda_{Q_r}\breve{h}_r\breve{h}_r^\dagger)/\partial\lambda_{Q_k}) = \breve{h}_k\breve{h}_k^\dagger$. The partial derivative in the second term in the trace evaluates to $(I + \breve{H}\Lambda_Q\breve{H}^\dagger)^{-1}$. Substituting these expressions into (22) and simplifying, we can write the kth entry of the stochastic gradient as $\breve{h}_k^\dagger(I + \breve{H}\Lambda_Q\breve{H}^\dagger)^{-1}\breve{h}_k$.

REFERENCES

[1] A. 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.

[2] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.

[3] A. Lozano, A. M. Tulino, and S. Verdú, "Multiple-antenna capacity in the low-power regime," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2527–2544, Oct. 2003.

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

[5] M. Vu and A. Paulraj, "On the capacity of MIMO wireless channels with dynamic CSIT," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1269–1283, Sep. 2007.

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

[7] 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.

[8] A. L. Moustakas and S. H. Simon, "Optimizing multiple-input single-output MISO communication systems with general Gaussian channels: Nontrivial covariance and nonzero mean," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2770–2780, Oct. 2003.

[9] S. A. Jafar, S. Vishwanath, and A. Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," in *Proc. IEEE Int. Conf. Commun.*, Helsinki, Finland, Jun. 2001, pp. 2266–2270.

[10] S. H. Simon and A. L. Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 406–417, Apr. 2003.

[11] E. A. 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. 3, no. 5, pp. 1543–1553, Sep. 2004.

[12] M. Kießling, "Unifying analysis of ergodic MIMO capacity in correlated Rayleigh fading environment," *Eur. Trans. Telecommun.*, vol. 16, no. 1, pp. 17–35, Jan. 2005.

[13] S. H. Simon and A. L. Moustakas, "Optimality of beamforming in multiple transmitter multiple receiver communication systems with partial channel knowledge," in *Proc. DIMACS Workshop Signal Process. Wireless Commun.*, Piscataway, NJ, Oct. 2002, p. 57.

[14] I. Bjelaković and H. Boche, "Structure of optimal input covariance matrices for MIMO systems with covariance feedback under general correlated fading," in *Proc. IEEE Int. Symp. Inf. Theory*, Seattle, WA, Jul. 2006, pp. 1041–1045.

[15] I. E. Telatar, "Capacity of multiantenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[16] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1994.

[17] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1999.

[18] M. Kang and M.-S. Alouini, "Largest eigenvalue of complex Wishart matrices and performance analysis of MIMO MRC systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 418–426, Apr. 2003.

[19] A. Gaivoronski, "Stochastic quasigradient methods and their implementation," in *Numerical Techniques for Stochastic Optimization*. New York: Springer-Verlag, 1988, ch. 16, pp. 313–351.

[20] L. W. Hanlen and A. Grant, *Optimal Transmit Covariance for Ergodic MIMO Channels*, Oct. 2005. [Online]. Available: http://arxiv.org/abs/cs/0510060