Optimal Spatial Correlations for the Noncoherent MIMO Rayleigh Fading Channel

Shivratna Giri Srinivasan and Mahesh K. Varanasi
e-mail: {srinivsg, varanasi}@colorado.edu
Electrical & Computer Engineering Department
University of Colorado, Boulder, CO 80309

This work was supported in part by NSF grants CCF-0434410 and CCF-0431170.

Abstract

The behavior in terms of information theoretic metrics of the discrete-input, continuous-output noncoherent MIMO Rayleigh fading channel is studied as a function of spatial correlations. In the low SNR regime, the mutual information metric is considered, while at higher SNR regimes the cutoff rate expression is employed. For any fixed input constellation and at sufficiently low SNR, a fully correlated channel matrix is shown to maximize the mutual information. In contrast, at high SNR, a fully uncorrelated channel matrix (with independent identically distributed elements) is shown to be optimal, under a condition on the constellation which ensures full diversity. In the special case of the separable correlation model, it is shown that as a function of the receive correlation eigenvalues, the cutoff rate expression is a Schur-convex function at low SNR and a Schur-concave function at high SNR, and as a function of transmit correlation eigenvalues, the cutoff rate expression is Schur-concave at high SNR for full diversity constellations. Moreover, at sufficiently low SNR, the fully correlated transmit correlation matrix is optimal. Finally, for the general model, it is shown that the optimal correlation matrices at a *general* SNR can be obtained using a *difference of convex programming* formulation.

Index Terms

Block fading channels, noncoherent, MIMO, spatial correlation, mutual information, cutoff rate expression, general SNR, schur-convexity, schur-concavity, global optimization, d.c. programming, concave minimization.

I. INTRODUCTION

Practical MIMO channels exhibit correlations between path gains of the antenna elements. It is therefore important to understand the effect of spatial correlations on the channel capacity since this helps in optimally designing the transmit and receive antenna arrays. For this reason, the effect of spatial correlations on the capacity of the *coherent* MIMO Rayleigh fading channel—where the channel realizations are assumed to be known at the receiver but only the long term statistics are known at the transmitter—is studied in several papers including [1–5]. In this paper, we consider the more challenging *noncoherent* MIMO Rayleigh fading channel where the transmitter and the receiver have only knowledge of the long term statistics and neither has knowledge of the channel realizations.

The rationale for studying the noncoherent model is this. Since in practice the channel is not known to the receiver at the start of communication, an information theoretic formulation of the noncoherent problem—which implicitly accounts for the resources needed for (implicit) channel estimation without constraining the transmission scheme in any way—is more fundamental than the coherent formulation. Systems that assume coherent transmission by arguing that the channel can be acquired at the receiver by the use of pilot-symbol assisted transmission to perform explicit channel estimation either (a) do not take into account the resources (power and degrees of freedom) needed for pilot transmissions or (b) when they do (as they should), they incur a significant loss of optimality in regimes involving short coherence times and/or low SNRs, implying that explicit pilot-assisted channel estimation is highly sub-optimal in these regimes.

We study the problem of noncoherent multi-antenna communication in the context of the general spatial correlation model of [6], referred to in [2] as the Unitary-Independent-Unitary (UIU) model, which subsumes the well known *separable transmit and receive correlation* model of [1, 7] and the *virtual channel representation* model of [8]. While even the UIU Rayleigh fading model does not capture the most general form of correlations, it is viewed as a reasonable compromise between validity and analytical tractability. Justification for the UIU Rayleigh fading model is given in [9] based on physical measurements. The particular case of the separable model is justified in [1] as an approximation, while [7] incorporates physical parameters like the angle spread and antenna spacing in this model. Consequently, we often specialize our results to the separable model and obtain sharper results for it.

A summary of the main results in [2,3] which compare the coherent capacity for the separable model with transmit and receive correlations with the i.i.d. fading model, is as follows :

1. Receive correlation reduces the capacity at every SNR.
2. For $N_t \leq N_r$, transmit correlation reduces the capacity at high SNR.
3. For $N_t > N_r$, transmit correlation increases the low SNR capacity.

The application of the theory of majorization to such problems (c.f. [1,4,5]) helps in providing a more complete understanding of how the correlations affect the performance measure of interest, and it aids comparison of correlated channels. The analysis in [4] is specific to the coherent capacity of the MISO channel ($N_t > 1$ and $N_r = 1$) and it is shown that at a general SNR, the capacity is a Schur-convex function with respect to the eigenvalues of the transmit correlation matrix. This means that higher transmit correlations result in higher capacity at all SNRs for the MISO channel. In [5], it is proved that the pairwise error probability (PEP) between every pair of symbol matrices is a Schur-convex function of the receive correlation eigenvalues. This means that higher correlations at the receiver result in higher PEPs at every SNR, which indicates that higher correlations are detrimental to error probability.

With regard to the noncoherent MIMO Rayleigh fading channel, a recent paper by Wu and Srikant [10] shows that at asymptotically low SNR, a fully correlated channel maximizes the reliability function. In other SNR regimes however, little is known about the effect of spatial correlations on the noncoherent MIMO channel performance. The main stumbling block in the analysis of the noncoherent channel is the absence of a closed form expression for the capacity. Indeed, the problem of finding the noncoherent MIMO capacity is one of the longstanding open problems, partial characterizations of which may be found in for instance [11–14]. We therefore adopt the cutoff rate for our analysis and obtain useful insights in this regard. The cutoff rate is a lower bound on capacity and was previously used by [15] (and the references contained therein in different contexts), to analyze and characterize optimal constellations for the peak-power constrained noncoherent MIMO i.i.d. Rayleigh fading channel. In [16], the cutoff rate expression is used as a criterion to design constellations for the average-power constrained noncoherent MIMO Rayleigh fading channel at general SNRs.

In this paper, we maximize the cutoff rate expression with respect to the channel correlation matrices for arbitrary but fixed signal constellations. We observe that the cutoff rate at sufficiently low SNR behaves exactly the same way as the mutual information upto second order, and hence the results hold for the mutual information as well in this regime. Our main results are as follows :

1. At sufficiently low SNR, we prove that the mutual information and cutoff rate expression are maximized by a fully correlated channel matrix. For the separable model, the cutoff rate expression is thus maximized by fully correlated transmit and receive correlation matrices. In the separable case, we show the sharper result that the mutual information is in fact a Schur-convex function of the receive correlation eigenvalues. This indicates that at low SNRs, it helps to have more correlations at the receive antennas, which is in contrast to results in the coherent case.

2. At asymptotically high SNR, and under a condition that ensures that the constellation achieves full diversity, we show that the cutoff rate expression is maximized by a fully uncorrelated channel matrix. In the case of the separable model, we prove the sharper result that the cutoff rate expression is a Schur-concave function of the transmit and receive correlation matrices.

3. We show how the optimal correlation matrix may be obtained at a general SNR, using standard global optimization formulations. In particular, we transform such problems into standard

global optimization problems like difference of convex (d.c.) programming and concave minimization, and indicate algorithms that obtain the globally optimal solution.

Notation : For an integer N, $\mathbf{I}_N$ is an $N \times N$ identity matrix. Matrices are denoted by the boldfaced capital letters, and vectors by bold faced small letters. The symbol $\otimes$ denotes the Kronecker product. The matrices $\mathbf{X}^T$, $\bar{\mathbf{X}}$ and $\mathbf{X}^*$ denote the transpose, complex-conjugate, and conjugate transpose respectively. We denote the inner product between two vectors $\mathbf{x}$ and $\mathbf{y}$ by $< \mathbf{x}, \mathbf{y} > = \mathbf{x}^*\mathbf{y}$ and the norm $\|\mathbf{x}\| = \sqrt{< \mathbf{x}, \mathbf{x} >}$. $\mathrm{E}[.]$ denotes the expectation operator. diag$(a_1, a_2, \ldots, a_N)$ is an $N \times N$ diagonal matrix with diagonal elements $a_1, a_2, \ldots, a_N$. We use the notation $o(\rho)$ to mean that $\lim_{\rho \to 0} \frac{o(\rho)}{\rho} = 0$.

II. System Model

We consider a communication system with N_t transmit antennas and N_r receive antennas. We assume a block fading channel where the channel matrix $\mathbf{H} \in \mathbb{C}^{N_t \times N_r}$ is assumed to be constant for a duration of T symbols, after which it changes to an independent value. The channel matrix is assumed to be unknown to the transmitter and the receiver, while the channel *statistics* are assumed to be known at the transmitter. The received signal is

$$\mathbf{R} = \sqrt{\gamma}\mathbf{S}\mathbf{H} + \mathbf{W}, \tag{1}$$

where $\mathbf{S} \in \mathbb{C}^{T \times N_t}$ is the transmitted symbol matrix and $\mathbf{W} \in \mathbb{C}^{T \times N_r}$ is the noise matrix. Here, the symbols $\{\mathbf{S}\}$ are normalized such that $\mathrm{E}[\mathrm{tr}(\mathbf{S}\mathbf{S}^*)] = 1$, so that the average transmit power equals γ. It is assumed that $\mathbf{W}$ has i.i.d. circularly symmetric $\mathcal{CN}(0,1)$ entries. We next describe the form of the channel matrix $\mathbf{H}$, which has correlated, circularly symmetric, complex gaussian entries.

A. Unitary-Independent-Unitary (UIU) Rayleigh fading model

In the UIU Rayleigh fading model, the channel matrix is assumed to be of the form

$$\mathbf{H} = \mathbf{U}_t \tilde{\mathbf{H}} \mathbf{U}_r^*, \tag{2}$$

where $\mathbf{U}_t$ and $\mathbf{U}_r$ are the transmit and receive unitary matrices. The elements of $\tilde{\mathbf{H}}$ are uncorrelated and zero-mean, circularly symmetric, complex gaussian, but not necessarily with the same variance. Define $\mathbf{h} = \mathrm{vec}(\mathbf{H})$. Assuming that $\tilde{\mathbf{\Sigma}} = \mathrm{E}[\mathrm{vec}(\tilde{\mathbf{H}})\mathrm{vec}(\tilde{\mathbf{H}})^*] = \mathbf{\Lambda}$, which is a non-negative diagonal matrix, we have

$$\mathbf{\Sigma} = \mathrm{E}[\mathbf{h}\mathbf{h}^*] \;=\; (\bar{\mathbf{U}}_r \otimes \mathbf{U}_t)\mathbf{\Lambda}(\bar{\mathbf{U}}_r \otimes \mathbf{U}_t)^*. \tag{3}$$

Let $\{\lambda_i\}_{i=1}^{N_t N_r}$ be the eigenvalues of $\mathbf{\Sigma}$. The normalizations in (1) are assumed to be such that $\sum_{i=1}^{N_t N_r} \lambda_i = N_t N_r$.

B. Separable Transmit and Receive Correlation model

For the separable transmit and receive correlation model, $\mathbf{H}$ is represented by

$$\mathbf{H} \;=\; \mathbf{\Sigma}_t^{1/2}\mathbf{H}_w\mathbf{\Sigma}_r^{1/2}, \tag{4}$$

where $\mathbf{H}_w$ has i.i.d. circularly symmetric $\mathcal{CN}(0,1)$ entries. The matrices $\mathbf{\Sigma}_t$ and $\mathbf{\Sigma}_r$ are the transmit and receive array correlation matrices, with eigenvalues $\{\lambda_i^t\}_{i=1}^{N_t}$ and $\{\lambda_i^r\}_{i=1}^{N_r}$, respectively. Substituting the eigenvalue decompositions for $\mathbf{\Sigma}_t = \mathbf{U}_t\mathbf{\Lambda}_t\mathbf{U}_t^*$ and $\mathbf{\Sigma}_r = \mathbf{U}_r\mathbf{\Lambda}_r\mathbf{U}_r^*$, we get that

$$\mathbf{H} \;=\; \mathbf{U}_t\tilde{\mathbf{H}}\mathbf{U}_r^*, \tag{5}$$

where $\tilde{\mathbf{H}} = \mathbf{\Lambda}_t^{1/2}\mathbf{U}_t^*\mathbf{H}_w\mathbf{U}_r\mathbf{\Lambda}_r^{1/2}$. The normalizations in (1) are assumed to be such that $\sum_{i=1}^{N_t} \lambda_i^t = N_t$ and $\sum_{i=1}^{N_r} \lambda_i^r = N_r$. Since $\mathbf{U}_t^*\mathbf{H}_w\mathbf{U}_r$ has the same distribution as $\mathbf{H}_w$, it can be seen that with $\tilde{\mathbf{h}} = \mathrm{vec}(\tilde{\mathbf{H}})$, $\tilde{\mathbf{\Sigma}} = \mathrm{E}[\tilde{\mathbf{h}}\tilde{\mathbf{h}}^*] = \mathbf{\Lambda}_r \otimes \mathbf{\Lambda}_t$. We may therefore obtain the correlation matrix of $\mathbf{h} = \mathrm{vec}(\mathbf{H})$ as

$$\mathrm{E}[\mathbf{h}\mathbf{h}^*] \;=\; (\bar{\mathbf{U}}_r \otimes \mathbf{U}_t)(\mathbf{\Lambda}_r \otimes \mathbf{\Lambda}_t)(\bar{\mathbf{U}}_r \otimes \mathbf{U}_t)^*. \tag{6}$$

Comparing (3) and (6), we see that the main difference between the separable and UIU Rayleigh fading model is that the eigenvalue matrix of the channel correlation matrix in the separable model is a Kronecker product, while there is no such restriction in the UIU Rayleigh fading model.

C. Effective channel model and output probability density function (p.d.f.)

The output of the channel in (1) using either assumption on the channel matrix $\mathbf{H}$ can be written as

$$\mathbf{R} \;=\; \sqrt{\gamma}\mathbf{S}\mathbf{U}_t\tilde{\mathbf{H}}\mathbf{U}_r^* + \mathbf{W}\,. \tag{7}$$

After post-multiplying (7) by $\mathbf{U}_r$, denoting $\mathbf{SU}_t$ by $\mathbf{X}$, and denoting $\mathbf{RU}_r$ by $\mathbf{Y}$, we get

$$\mathbf{Y} = \sqrt{\gamma}\mathbf{X}\tilde{\mathbf{H}} + \mathbf{N} \, , \tag{8}$$

which represents the sufficient statistics of the received signal. Clearly, $\mathrm{E}[\mathrm{tr}(\mathbf{SS}^*)] = \mathrm{E}[\mathrm{tr}(\mathbf{XX}^*)]$, and hence the *precoded* constellation $\{\mathbf{X}\}$ satisfies the same average-power constraint as the original constellation $\{\mathbf{S}\}$. $\mathbf{N}$ has i.i.d. circularly symmetric $\mathcal{CN}(0,1)$ entries since it has the same distribution as $\mathbf{W}$. We may hence consider (8) to be our effective channel model, and use the notation $\mathbf{X}$ to denote a constellation matrix precoded by the transmit unitary matrix. Note that this amounts to a form of statistical beamforming which exploits knowledge of the channel statistics at the transmitter and is not to be confused with channel realization dependent transmit beamforming which is of course not feasible in the noncoherent channel.

Applying a vec operation to (8), we get $\mathbf{y} = \sqrt{\gamma}(\mathbf{I}_{N_r} \otimes \mathbf{X})\tilde{\mathbf{h}} + \mathbf{n} = \sqrt{\gamma}\boldsymbol{\mathcal{X}}\tilde{\mathbf{h}} + \mathbf{n}$, where $\mathbf{y} = \mathrm{vec}(\mathbf{Y})$ and $\mathbf{n} = \mathrm{vec}(\mathbf{N})$. The pdf of $\mathbf{y}$ conditioned on $\boldsymbol{\mathcal{X}}$ being sent is given by

$$p(\mathbf{y}|\boldsymbol{\mathcal{X}}) \;=\; \frac{1}{\pi^{TN_r}|\mathbf{I} + \gamma\boldsymbol{\mathcal{X}}\tilde{\boldsymbol{\Sigma}}\boldsymbol{\mathcal{X}}^*|} e^{-\mathbf{y}^*(\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}\tilde{\boldsymbol{\Sigma}}\boldsymbol{\mathcal{X}}^*)^{-1}\mathbf{y}} \, .$$

III. Optimal Correlations at Low SNR

Throughout this paper, we assume our model to be a discrete input (of cardinality L) and continuous output channel over which a constellation $\{\mathbf{X}_i\}_{i=1}^L$ with corresponding prior probabilities $\{P_i\}_{i=1}^L$ is used. In this section we obtain the optimal correlation matrices that maximize the mutual information at sufficiently low SNR.

Rao and Hassibi [17] derive the low SNR mutual information for the continuous input and continuous output channel, when the signals are subject to average and peak power constraints. Such regularity conditions are required since otherwise the optimal signals at low SNR have very large peak-powers. With a similar analysis tailored to the discrete input and continuous output channel with spatially correlated fading, a similar expression for the mutual information can be obtained which is

$$I_{low} \;=\; \frac{\gamma^2}{2}\left\{\mathrm{E}[\mathrm{tr}\{(\boldsymbol{\mathcal{X}}\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}^*)^2\}] - \mathrm{tr}\{(\mathrm{E}[\boldsymbol{\mathcal{X}}\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}^*])^2\right\} + o(\gamma^2). \tag{9}$$

Under different regularity conditions and for more general channels, the authors in [18] also obtain the mutual information upto the second order. When the expression for mutual information at low SNR in [18] is specialized to the channel model assumed in this paper, it can be seen to be identical to (9).

The expression for I_{low} in (9) may be rewritten as follows

$$
\begin{aligned}
I_{low} &= \frac{\gamma^2}{4}\Bigg\{\sum_i P_i \mathrm{tr}(\boldsymbol{\mathcal{X}}_i \boldsymbol{\Lambda} \boldsymbol{\mathcal{X}}_i^* \boldsymbol{\mathcal{X}}_i \boldsymbol{\Lambda} \boldsymbol{\mathcal{X}}_i^*) + \sum_j P_j \mathrm{tr}(\boldsymbol{\mathcal{X}}_j \boldsymbol{\Lambda} \boldsymbol{\mathcal{X}}_j^* \boldsymbol{\mathcal{X}}_j \boldsymbol{\Lambda} \boldsymbol{\mathcal{X}}_j^*) \\
&\quad -2\mathrm{tr}(\sum_i P_i \boldsymbol{\mathcal{X}}_i \boldsymbol{\Lambda} \boldsymbol{\mathcal{X}}_i^* \sum_j P_j \boldsymbol{\mathcal{X}}_j \boldsymbol{\Lambda} \boldsymbol{\mathcal{X}}_j^*)\Bigg\} + o(\gamma^2) \\
&= \frac{\gamma^2}{4} \sum_i \sum_j P_i P_j \mathrm{tr}\{(\boldsymbol{\mathcal{X}}_i \boldsymbol{\Lambda} \mathbf{X}_i^* - \boldsymbol{\mathcal{X}}_j \boldsymbol{\Lambda} \boldsymbol{\mathcal{X}}_j^*)^2\} + o(\gamma^2)\,. \qquad (10)
\end{aligned}
$$

Let $\boldsymbol{\lambda}$ denote the vector of diagonal elements of $\boldsymbol{\Lambda}$, which are the eigenvalues of $\boldsymbol{\Sigma}$. Let λ_i denote the i^{th} element of $\boldsymbol{\lambda}$. We next maximize (10) with respect to $\boldsymbol{\lambda}$ subject to the constraint $\sum_i \lambda_i = N_t N_r$.

Theorem 1: $\lim_{\gamma\to 0} \frac{I_{low}}{\gamma^2}$ is maximized by choosing all eigenvalues of $\boldsymbol{\Sigma}$ to be zero except for one, ie. $\boldsymbol{\lambda} = [0, 0, \ldots, N_t N_r, \ldots, 0]^T$, where the position of the non-zero element $N_t N_r$ depends on the specific constellation used.

Proof: Denote the k^{th} column of $\boldsymbol{\mathcal{X}}_i$ by $\mathbf{x}_{ik}$, $k = 1, \ldots, N_t N_r$.

$$
\begin{aligned}
\lim_{\gamma\to 0} \frac{I_{low}}{\gamma^2} &= \frac{1}{4} \sum_i \sum_j \; P_i P_j \mathrm{tr}\{(\boldsymbol{\mathcal{X}}_i \boldsymbol{\Lambda} \boldsymbol{\mathcal{X}}_i^* - \boldsymbol{\mathcal{X}}_j \boldsymbol{\Lambda} \boldsymbol{\mathcal{X}}_j^*)^2\} \qquad (11) \\
&= \frac{1}{4} \sum_i \sum_j \; P_i P_j \mathrm{tr}\left\{\left\{\sum_{k=1}^{N_t N_r} \lambda_k (\mathbf{x}_{ik}\mathbf{x}_{ik}^* - \mathbf{x}_{jk}\mathbf{x}_{jk}^*)\right\}^2\right\} \qquad (12) \\
&= \frac{1}{4} \sum_i \sum_j \; P_i P_j \mathrm{tr}\left\{\sum_k \sum_l \lambda_k \lambda_l (\mathbf{x}_{ik}\mathbf{x}_{ik}^* - \mathbf{x}_{jk}\mathbf{x}_{jk}^*)(\mathbf{x}_{il}\mathbf{x}_{il}^* - \mathbf{x}_{jl}\mathbf{x}_{jl}^*)\right\} \qquad (13)
\end{aligned}
$$

Let $\mathbf{A}_{ijk} = \mathbf{x}_{ik}\mathbf{x}_{ik}^* - \mathbf{x}_{jk}\mathbf{x}_{jk}^*$, $\forall k = 1, \ldots, N_t N_r$. Also, define $a_{kl} = \lambda_k \lambda_l$. We need to solve the optimization problem $\max_{\substack{\sum_k \lambda_k = N_t N_r \\ \lambda_k \geq 0 \;\; \forall k}} \; \sum_k \sum_l \lambda_k \lambda_l \; \left\{\sum_i \sum_j \; P_i P_j \mathrm{tr}(\mathbf{A}_{ijk} \mathbf{A}_{ijl})\right\}$

$$
\begin{aligned}
&= \max_{\substack{\sum_k \sum_l \lambda_k \lambda_l = N_t^2 N_r^2 \\ \lambda_k \geq 0 \;\; \forall k}} \; \sum_k \sum_l \lambda_k \lambda_l \left\{\sum_i \sum_j \; P_i P_j \mathrm{tr}(\mathbf{A}_{ijk} \mathbf{A}_{ijl})\right\} \qquad (14) \\
&= N_t^2 N_r^2 \max_{\substack{\sum_k \sum_l \frac{a_{kl}}{N_t^2 N_r^2} = 1 \\ \lambda_k \geq 0 \;\; \forall k}} \; \sum_k \sum_l \frac{a_{kl}}{N_t^2 N_r^2} \left\{\sum_i \sum_j \; P_i P_j \mathrm{tr}(\mathbf{A}_{ijk} \mathbf{A}_{ijl})\right\}\,, \qquad (15)
\end{aligned}
$$

which can be viewed as a convex combination of terms of the form $\sum_i \sum_j \ P_i P_j \mathrm{tr}(\mathbf{A}_{ijk}\mathbf{A}_{ijl}),\ \forall k,l$, with weights $\left\{\frac{a_{kl}}{N_t^2 N_r^2}\right\}$. Therefore, the maximum should occur when all except one of the $\{a_{kl}\}$ are zero. The only non-zero weight (say) $a_{kl} = N_t^2 N_r^2$ corresponds to the indices that achieve $\max_{k,l} \ \sum_i \sum_j \ P_i P_j \mathrm{tr}(\mathbf{A}_{ijk}\mathbf{A}_{ijl})$. We cannot however have $a_{kl} = \lambda_k \lambda_l = N_t^2 N_r^2$ when $k \neq l$ since $\sum_k \lambda_k = N_t N_r$. Therefore, we first show that the maximum occurs only when $k = l$ and then conveniently obtain the maximum of the convex combination.

$$\begin{aligned}
\sum_i \sum_j P_i P_j \mathrm{tr}(\mathbf{A}_{ijk}\mathbf{A}_{ijl}) &= \sum_i \sum_j P_i P_j \sum_m < \mathbf{A}_{ijk}^{(m)}, \mathbf{A}_{ijl}^{(m)} > && (16)\\
&\leq \sum_i \sum_j P_i P_j \sum_m \|\mathbf{A}_{ijk}^{(m)}\| \|\mathbf{A}_{ijl}^{(m)}\| && (17)\\
&\leq \sum_i \sum_j P_i P_j \sqrt{\sum_m \|\mathbf{A}_{ijk}^{(m)}\|^2} \sqrt{\sum_m \|\mathbf{A}_{ijl}^{(m)}\|^2} && (18)\\
&= \sum_i \sum_j P_i P_j \sqrt{\mathrm{tr}(\mathbf{A}_{ijk}^2)\mathrm{tr}(\mathbf{A}_{ijl}^2)} && (19)\\
&\leq \sqrt{\left\{\sum_i \sum_j P_i P_j \mathrm{tr}(\mathbf{A}_{ijk}^2)\right\}\left\{\sum_i \sum_j P_i P_j \mathrm{tr}(\mathbf{A}_{ijl}^2)\right\}} && (20)\\
&\leq \max\left\{\sum_i \sum_j P_i P_j \mathrm{tr}(\mathbf{A}_{ijk}^2), \sum_i \sum_j P_i P_j \mathrm{tr}(\mathbf{A}_{ijl}^2)\right\}. && (21)
\end{aligned}$$

The inequalities in (17) and (18) are obtained by applying the Cauchy-Schwarz inequality successively. The inequality in (20) is obtained by recognizing that the geometric mean $f(\mathbf{x}) = \sqrt{x_1 x_2}$ is a concave function on $\mathbf{x} \in \Re_{++}^2$ and by applying the Jensen's inequality on (19). The square root is well defined since $\mathrm{tr}(\mathbf{A}^2) \geq 0$ whenever $\mathbf{A}$ is Hermitian, which is the case here. Finally, (21) is obtained by using the fact that $\sqrt{ab} \leq \max\{a,b\}$ for positive a,b.

The chain of inequalities leads to the conclusion that the $\max_{k,l} \ \sum_i \sum_j \ P_i P_j \mathrm{tr}(\mathbf{A}_{ijk}\mathbf{A}_{ijl})$ occurs only when $k = l$. For this maximizing index k (say), the convex combination in (15) is maximized by choosing $\lambda_k = N_t N_r$ and all other eigenvalues to be zero, ie., $\boldsymbol{\lambda} = [0 \ 0 \ \ldots \ N_t N_r \ \ldots 0]^T$. The maximum value of the mutual information would be

$$N_t^2 N_r^2 \frac{\gamma^2}{4} \ \max_k \sum_i \sum_j \ P_i P_j \mathrm{tr}(\mathbf{A}_{ijk}^2). \tag{22}$$

■

Theorem 1 implies that for *any* set of signals, at sufficiently low SNR, a channel having just a single effective eigenchannel is optimal. Having a single effective channel would imply that the effective dimensions available is one, but this enables focusing power which is more essential at low SNR, when there is no channel state information at the receiver. Placing the transmit and receive antennas densely reduces the resolvability of the different paths in the angular domain, thus resulting in $\mathbf{H}$ having more correlated entries [19]. In terms of system design therefore, Theorem 1 suggests that having closely spaced transmit and receive antenna arrays results in a higher capacity for noncoherent MIMO communications in the low SNR regime. This is in contrast to the coherent MIMO scenario at high SNR, where maximizing the number of degrees of freedom in the channel is crucial. Similar insights are also obtained in [10] while considering the reliability function at low SNR.

We will need the definitions of majorization, Schur-convex and Schur-concave functions from [20] for some of the ensuing propositions. These definitions and properties are provided in Appendix-A for the sake of completeness.

If $\mathbf{x}$ and $\mathbf{y}$ are vectors of eigenvalues of two correlation matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$, $\mathbf{x} \prec \mathbf{y}$ would mean that $\mathbf{\Sigma}_2$ is more correlated than $\mathbf{\Sigma}_1$. This notion of majorization has been used in many papers studying the effect of fading correlations on the MIMO channel. It provides a more detailed characterization of the performance, and thereby helps to compare two correlated channels whenever possible using their respective vectors of eigenvalues. It should be noted that the notion of majorization need not relate any two vectors whose entries sum up to the same value. The results obtained, hence pertain to those vectors of eigenvalues that can be compared via majorization.

The optimal correlations for the separable model may be obtained by solving $\mathbf{\Lambda}_r \otimes \mathbf{\Lambda}_t = \mathbf{\Lambda} = \text{diag}(0, \ldots, 0, N_t N_r, 0, \ldots, 0)$. Therefore, the jointly optimal transmit and receive correlation eigenvalues are given by $\mathbf{\Lambda}_t = \text{diag}(0, \ldots, 0, N_t, 0, \ldots, 0)$ and $\mathbf{\Lambda}_r = \text{diag}(0, \ldots, 0, N_r, 0, \ldots, 0)$ respectively. The optimal matrices $\mathbf{\Lambda}_t$ and $\mathbf{\Lambda}_r$ have exactly one non-zero value each and their positions depend on the specific constellation used. We introduce the following notation which will be used in the rest of this paper. Let $\boldsymbol{\lambda}_t$ and $\boldsymbol{\lambda}_r$ be the vectors of eigenvalues of the transmit and receive correlation matrices, with elements $\{\lambda_n^t\}_{n=1}^{N_t}$ and $\{\lambda_n^r\}_{n=1}^{N_r}$, respectively. Using the fact that for the separable model $\mathbf{\Lambda} = \mathbf{\Lambda}_r \otimes \mathbf{\Lambda}_t$, the low SNR mutual information expression can be simplified to the following form

$$I_{low} = \frac{\gamma^2}{4}\left\{\sum_n (\lambda_n^r)^2\right\} \sum_i \sum_j P_i P_j \mathrm{tr}\{(\mathbf{X}_i \mathbf{\Lambda}_t \mathbf{X}_i^* - \mathbf{X}_j \mathbf{\Lambda}_t \mathbf{X}_j^*)^2\} + o(\gamma^2). \qquad (23)$$

The following propositions describe more conclusions that can be made with respect to the low SNR mutual information as a function of $\boldsymbol{\lambda}_r$ and $\boldsymbol{\lambda}_t$, for the separable model.

Proposition 1: In the separable model, the mutual information at low SNR is a Schur-convex function of $\boldsymbol{\lambda}_r$.

Proof: Since $\{\sum_n (\lambda_n^r)^2\}$ is a Schur-convex function of $\boldsymbol{\lambda}_r$ [20], so is I_{low}. ■

Proposition 1 indicates that higher correlations at the receiver are beneficial in the low SNR regime. This result contrasts with results in the coherent scenario, where receive correlations are detrimental to the performance at any SNR. An intuitive explanation for this difference is that in the low SNR noncoherent channel, it helps the implicit channel estimation when the fading coefficients across the receive antennas are highly correlated.

The mutual information upto second order is not a Schur-convex function of the transmit eigenvalues at low SNR in general. This is because it depends on the specific signals used and the expression is not even a symmetric function of $\boldsymbol{\lambda}_t$. The following proposition can however be proved by analytically maximizing (23) with respect to $\boldsymbol{\lambda}_t$, for any fixed $\boldsymbol{\lambda}_r$.

Proposition 2: The mutual information at sufficiently low SNR for the separable model is maximized by $\boldsymbol{\lambda}_t = [0 \ \ 0 \ .. \ \ N_t \ \ 0 \ \ 0]^T$ for *any* fixed $\boldsymbol{\lambda}_r$, where the position of the non-zero element N_t depends on the specific constellation used.

Proof: The proof is along similar lines to that of Theorem 1 and the details are left to the reader. ■

IV. Optimal correlations at higher SNR regimes

At a general SNR, the mutual information is not known in closed form. As a result we use the cutoff rate expression as our design criterion.

A. The cutoff rate

Consider a constellation $\{\mathbf{X}_i\}_{i=1}^L$ with corresponding prior probabilities $\{P_i\}_{i=1}^L$. The cutoff rate for the discrete input (of cardinality L) and continuous output channel is given by

$$R_0 = \max_{\{P_i\}_{i=1}^L, \{\mathbf{X}_i\}_{i=1}^L} -\log\left\{\sum_i\sum_j P_iP_j \int \sqrt{p(\mathbf{y}|i)p(\mathbf{y}|j)}d\mathbf{y}\right\}. \quad (24)$$

For the system model given in Section II, the argument of $\max(.)$ in (24) is easily found in Appendix-B to be

$$CR = -\log\left\{\sum_i\sum_j P_iP_j \frac{\left|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*\right|^{1/2}\left|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*\right|^{1/2}}{\left|\mathbf{I}+\frac{\gamma}{2}(\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*+\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*)\right|}\right\}. \quad (25)$$

We refer to CR in (25) as the *cutoff rate expression*. It should be noted that the cutoff rate expression is a lower bound on the mutual information at any SNR.

The next proposition is an extension of a result for the i.i.d. channel by Hero and Marzetta in [15] to the correlated channel.

Proposition 3: In the low SNR regime, the cutoff rate expression upto second order in γ may be expressed as

$$CR_{low} = \frac{\gamma^2}{8}\sum_i\sum_j P_iP_j\mathrm{tr}\{(\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*-\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*)^2\}+o(\gamma^2). \quad (26)$$

Proof: Refer to Appendix-C. ■

The following simple proposition shows that the low SNR cutoff rate expression behaves *identically* to the low SNR mutual information.

Proposition 4: At sufficiently low SNR, $CR_{low} = \frac{1}{2}I_{low}+o(\gamma^2)$.

Proof: The proposition follows by inspection of (10) and (26). ■

Since CR_{low} has the same behavior as the mutual information at sufficiently low SNR, the results in Section III are valid for the cutoff rate expression as well.

B. High SNR

In this section, we optimize the cutoff rate expresssion at high SNR over the eigenvalues of the correlation matrix. In the next theorem, we assume that $[\mathbf{X}_i\ \mathbf{X}_j]$ has full column rank ie. a rank of $2N_t$, for which in turn $T \geq 2N_t$ is a necessary condition. These conditions are known to be sufficient to ensure that the constellation achieves the maximum possible diversity order in the channel [21]. If $\boldsymbol{\Lambda}$ is of full rank, the maximum possible diversity order would be N_tN_r.

Theorem 2: Let $[\mathbf{X}_i\ \mathbf{X}_j]$ be of full column rank $2N_t$ for every pair i,j of constellation matrices. Then, at asymptotically high SNR the cutoff rate expression is a Schur-concave function of $\boldsymbol{\lambda}$ for $\lambda_k \in (0, N_t N_r],\ \forall k$.

Proof: Since $[\mathbf{X}_i\ \ \mathbf{X}_j]$ is of full column rank, $[\boldsymbol{\mathcal{X}}_i\ \ \boldsymbol{\mathcal{X}}_j]$ is also of full column rank. We may write the cutoff rate expression as

$$CR = -\log\left\{\sum_i P_i^2 + \sum_{\substack{i,j\\ j\neq i}} P_i P_j \frac{\left|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*\right|^{1/2}\left|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*\right|^{1/2}}{\left|\mathbf{I}+\frac{\gamma}{2}(\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*+\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*)\right|}\right\}. \quad (27)$$

Using the identity $|\mathbf{I}+\mathbf{AB}| = |\mathbf{I}+\mathbf{BA}|$, we simplify CR as follows

$$CR = -\log\left\{\sum_i P_i^2 + \sum_{\substack{i,j\\ j\neq i}} P_i P_j \frac{\left|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_i^*\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\right|^{1/2}\left|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_j^*\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\right|^{1/2}}{\left|\mathbf{I}+\frac{\gamma}{2}[\boldsymbol{\mathcal{X}}_i\ \ \boldsymbol{\mathcal{X}}_j]^*[\boldsymbol{\mathcal{X}}_i\ \ \boldsymbol{\mathcal{X}}_j]\,(\mathbf{I}_2\ \otimes\ \boldsymbol{\Lambda})\right|}\right\} \quad (28)$$

$$\approx -\log\left\{\sum_i P_i^2 + \sum_{\substack{i,j\\ j\neq i}} P_i P_j \frac{\left|\gamma\boldsymbol{\mathcal{X}}_i^*\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\right|^{1/2}\left|\gamma\boldsymbol{\mathcal{X}}_j^*\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\right|^{1/2}}{\left|\frac{\gamma}{2}[\boldsymbol{\mathcal{X}}_i\ \ \boldsymbol{\mathcal{X}}_j]^*[\boldsymbol{\mathcal{X}}_i\ \ \boldsymbol{\mathcal{X}}_j]\,(\mathbf{I}_2\ \otimes\ \boldsymbol{\Lambda})\right|}\right\} \quad (29)$$

$$= -\log\left\{\sum_i P_i^2 + \sum_{\substack{i,j\\ j\neq i}} P_i P_j \frac{\left|\gamma\boldsymbol{\mathcal{X}}_i^*\boldsymbol{\mathcal{X}}_i\right|^{1/2}\left|\boldsymbol{\Lambda}\right|^{1/2}\left|\gamma\boldsymbol{\mathcal{X}}_j^*\boldsymbol{\mathcal{X}}_j\right|^{1/2}\left|\boldsymbol{\Lambda}\right|^{1/2}}{\left|\frac{\gamma}{2}[\boldsymbol{\mathcal{X}}_i\ \ \boldsymbol{\mathcal{X}}_j]^*[\boldsymbol{\mathcal{X}}_i\ \ \boldsymbol{\mathcal{X}}_j]\right|\left|\boldsymbol{\Lambda}\right|^2}\right\} \quad (30)$$

$$= -\log\left\{\sum_i P_i^2 + \sum_{\substack{i,j\\ j\neq i}} P_i P_j c_{ij} \frac{1}{|\boldsymbol{\Lambda}|}\right\}. \quad (31)$$

Now, $\frac{1}{|\boldsymbol{\Lambda}|}$ is a Schur-convex function of $\boldsymbol{\lambda}$ for $\lambda_k > 0$. The condition $\lambda_k > 0$ is needed so that $|\boldsymbol{\Lambda}|$, which occurs in the denominator is non-zero. Therefore, since $c_{ij} \geq 0\ \ \forall i,j$ and $\{P_i\}_{i=1}^L$ are all non-negative, and since $h(x) = -\log(x)$ is a decreasing function in $\Re_{++}$, CR at high SNR is a Schur-concave function of $\boldsymbol{\lambda}$ for $\lambda_k \in (0, N_t N_r],\ \ \forall k$. ■

By the theory of majorization and Theorem 2, we conclude that $\boldsymbol{\lambda} = [1\ 1\ \ \ldots\ \ 1]^T$ is the optimal choice of the eigenvalue vector. At high SNR, Theorem 2 indicates that the channel matrix should be made as close to i.i.d. as possible. This has the beneficial effect of creating as many independent paths as possible.

The optimal correlations for the separable model may be obtained by solving $\boldsymbol{\Lambda}_r \otimes \boldsymbol{\Lambda}_t = \boldsymbol{\Lambda} = \text{diag}(1, 1, \ldots, 1)$ under the assumption that $\sum_n \lambda_n^t = N_t$ and $\sum_n \lambda_n^r = N_r$. This implies that the jointly optimal transmit and receive correlation eigenvalues are given by $\boldsymbol{\Lambda}_t = \text{diag}(1, 1, \ldots, 1)$

and $\mathbf{\Lambda}_r = \text{diag}(1, 1, \ldots, 1)$, respectively. More insightful conclusions pertaining to the separable model can be made using the theory of majorization, which are stated in the following proposition.

Proposition 5: (i) At asymptotically high SNR and for *any* $\boldsymbol{\lambda}_t$, the cutoff rate expression is a Schur-concave function of $\boldsymbol{\lambda}_r$.

(ii) Let $[\mathbf{X}_i \;\; \mathbf{X}_j]$ be of full column rank $2N_t$ for every pair i,j of constellation matrices. Then, at asymptotically high SNR and for *any* $\boldsymbol{\lambda}_r$, the cutoff rate expression is a Schur-concave function of $\boldsymbol{\lambda}_t$ for $\lambda_k^t \in (0, N_t] \; \forall k$.

Proof: Let $N = \text{rank}(\mathbf{\Sigma}_r)$.

(i) We may write the cutoff rate expression as

$$CR \;=\; -\log\left\{\sum_i P_i^2 \;+\; \sum_{\substack{i,j \\ j\neq i}} P_i P_j \prod_{n=1}^{N} \frac{\left|\mathbf{I} + \gamma \mathbf{X}_i \mathbf{\Lambda}_t \mathbf{X}_i^* \lambda_n^r\right|^{1/2} \left|\mathbf{I} + \gamma \mathbf{X}_j \mathbf{\Lambda}_t \mathbf{X}_j^* \lambda_n^r\right|^{1/2}}{\left|\mathbf{I} + \frac{\gamma}{2}(\mathbf{X}_i \mathbf{\Lambda}_t \mathbf{X}_i^* + \mathbf{X}_j \mathbf{\Lambda}_t \mathbf{X}_j^*)\lambda_n^r\right|}\right\} \tag{32}$$

Let the non-zero eigenvalues of $\mathbf{A}_i = \frac{1}{2}\mathbf{X}_i \mathbf{\Lambda}_t \mathbf{X}_i^*$ be $\{\mu_{iq}\}_{q=1}^{Q_i}$ and those of $\mathbf{A}_i + \mathbf{A}_j = \frac{1}{2}(\mathbf{X}_i \mathbf{\Lambda}_t \mathbf{X}_i^* + \mathbf{X}_j \mathbf{\Lambda}_t \mathbf{X}_j^*)$ be $\{\theta_{ijs}\}_{s=1}^{S_{ij}}$. Then the cutoff rate expression may be written as

$$CR = -\log\left\{\sum_i P_i^2 + \sum_{\substack{i,j \\ j\neq i}} P_i P_j \prod_{n=1}^{N} \frac{\left[\prod_{q=1}^{Q_i}(1+\gamma\lambda_n^r \mu_{iq})\right]^{1/2} \left[\prod_{r=1}^{Q_j}(1+\gamma\lambda_n^r \mu_{jr})\right]^{1/2}}{\prod_{s=1}^{S_{ij}}(1+\gamma\lambda_n^r \theta_s)}\right\}, \tag{33}$$

which, for asymptotically high SNR ($\gamma \to \infty$), may be written as

$$CR = -\log\left\{\sum_i \; P_i^2 + \sum_{\substack{i,j \\ j\neq i}} P_i P_j \left(\frac{1}{\gamma^N \prod_{n=1}^{N} \lambda_n^r}\right)^{S_{ij} - Q_i/2 - Q_j/2} \frac{\prod_{q=1}^{Q_i} \mu_{iq}^{\frac{N}{2}} \prod_{r=1}^{Q_j} \mu_{jr}^{\frac{N}{2}}}{\prod_{s=1}^{S_{ij}} \theta_s^N}\right\}. \tag{34}$$

Now, $\left(\prod_{n=1}^{N} \lambda_n^r\right)^{-1}$ is a Schur-convex function of $\boldsymbol{\lambda}^r$ by (3.E.1) [20] , since $1/\lambda_n^r$ is a log-convex function of λ_n^r. Since $Q_i = \text{rank}(\mathbf{A}_i^{1/2})$, $Q_j = \text{rank}(\mathbf{A}_j^{1/2})$, and $S_{ij} = \text{rank}([\mathbf{A}_i^{1/2} \;\; \mathbf{A}_j^{1/2}])$, clearly $S_{ij} \geq Q_i$ and $S_{ij} \geq Q_j$. This makes $S_{ij} - Q_i/2 - Q_j/2 \geq 0$. Hence $\left(\prod_{n=1}^{N} \lambda_n^r\right)^{-(S_{ij} - Q_i/2 - Q_j/2)}$ is also a Schur-convex function of $\boldsymbol{\lambda}^r$ by (3.B.1) [20]. Grouping all terms within the summation which multiply this term and denoting it by $c_{ij} \geq 0$, we get that $\sum_{i,j} c_{ij} \times \left(\prod_{n=1}^{N} \lambda_n^r\right)^{-(S_{ij} - Q_i/2 - Q_j/2)}$ is also a Schur-convex function, since a non-negative weighted combination of Schur-convex functions is also Schur-convex. Finally, since $h(x) = -\log(x)$ is a decreasing function in $\Re_{++}$, and $g(x) = \sum_{i,j} c_{ij} \times \left(\prod_{n=1}^{N} \lambda_n^r\right)^{-(S_{ij}-Q_i/2-Q_j/2)}$ is Schur-convex, the composition $h \circ g$ is Schur-concave. Hence, CR is a Schur-concave function of $\boldsymbol{\lambda}_r$ at high SNR.

(ii) The proof of this part is similar to the proof of Theorem 2. The proof is omitted and details are left to the reader. ■

Under input peak constraints, since the noncoherent capacity scales as $O(\mathrm{SNR}^2)$ [17], the energy per bit increases without bound as the SNR tends to zero. This observation, which was first made in [22], indicates that it is very energy inefficient to operate at vanishingly small SNRs. Nevertheless, numerical results in [22] indicate that the minimum energy per bit typically occurs in the non-asymptotic low SNR regime. Since the noncoherent capacity is not known at general SNR, the insights that we obtain at asymptotically low SNR offer engineering guidelines which may still hold at the SNR where the energy-efficiency is maximum. In any case, we next propose a technique to find the optimal correlations at a *general* SNR.

C. General SNR

It can be seen that the cutoff rate expression is non-convex in general with respect to the transmit and receive eigenvalues, and hence this problem comes under the realm of *deterministic global optimization* [23]. In order to maximize the cutoff rate at a general SNR, a globally optimal solution can be obtained by formulating it as a *difference of convex programming* (d.c. programming) problem [23].

We give some definitions from [23] that we will need in the theorem that follows.

Definition 1: A *polyhedron* is defined to be the set of points $C = \{\mathbf{x} \in \Re^n : \mathbf{A}\mathbf{x} \leq \mathbf{b}$, where $\mathbf{A} \in \Re^{m \times n}$ and $\mathbf{b} \in \Re^m$. A bounded polyhedron is called a polytope.

Definition 2: A real valued function f defined on a convex set $\boldsymbol{\mathcal{A}} \subseteq \Re^n$ is called d.c. (difference of convex) on $\boldsymbol{\mathcal{A}}$ if, for all $\mathbf{x} \in \boldsymbol{\mathcal{A}}$, f can be expressed in the form

$$f(\mathbf{x}) = p(\mathbf{x}) - q(\mathbf{x}) \,, \tag{35}$$

where p, q are convex functions on $\boldsymbol{\mathcal{A}}$. The representation (35) is said to be a d.c. decomposition of f.

Definition 3: A global optimization problem is called a d.c. programming problem or d.c. program if it has the form

$$\begin{aligned} \min \quad & f_0(\mathbf{x}) \\ \text{s.t} \quad & \mathbf{x} \in \boldsymbol{\mathcal{A}}\,, \\ & f_i(\mathbf{x}) \leq 0 \quad (i = 1, \ldots, m)\,, \end{aligned} \tag{36}$$

where $\boldsymbol{\mathcal{A}}$ is a closed convex subset of $\Re^n$ and all functions f_i , $(i = 0, 1, \ldots, m)$ are d.c on $\boldsymbol{\mathcal{A}}$. If the set of all constraints form a polytope, then the problem is called a d.c. program over a polytope.

There are a number of algorithms given in Chapter 4 of [23] to find the global minimum of a d.c. program if the d.c. decomposition is known.

Definition 4: A *concave minimization* problem is an optimization problem in the following form :

$$\min_{\mathbf{x} \,\in\, X} \;\; f(\mathbf{x})\,, \tag{37}$$

where $f(\mathbf{x})$ is a concave function and $X \subset \Re^n$ is a convex set.

Definition 5: A reverse convex set (or concave set) is a set whose complement is an open convex set.

We state the next lemma which is needed to prove the ensuing theorem.

Lemma 1: The function $f(\mu, \mathbf{D}_1, \mathbf{D}_2) = \det\left\{(\mathbf{I} + \mu(\mathbf{A}\mathbf{D}_1\mathbf{A}^* + \mathbf{B}\mathbf{D}_2\mathbf{B}^*))^{-1}\right\}$ defined over positive semidefinite diagonal matrices $\mathbf{D}_1$, $\mathbf{D}_2$ and non-negative μ is a jointly log-convex function of $\mathbf{D}_1$ and $\mathbf{D}_2$ for fixed μ.

Proof: The functions indicated are all compositions of the function $h(\mathbf{C}) = -\log\det(\mathbf{I} + \mathbf{C})$ and linear functions of the form $g(\mathbf{D}_1, \mathbf{D}_2) = \mathbf{A}\mathbf{D}_1\mathbf{A}^* + \mathbf{B}\mathbf{D}_2\mathbf{B}^*$. Since $h(\mathbf{C})$ is convex over positive semidefinite $\mathbf{C}$, the composition $f = h \circ g$ is also convex [24]. ■

Definition 6: A function $f(\mathbf{x})$ is log-convex if $\log(f(\mathbf{x}))$ is convex.

A lemma that will be found useful in obtaining d.c. decompositions of complicated functions is next proved. It will be invoked in the ensuing theorem to get d.c. decompositions.

Lemma 2: Let $h_i(\mathbf{x})$ and $g_i(\mathbf{x})$ be log-convex functions $\forall\, i = 1. \;\ldots\;, L$ over $\Re^n$ and c_i be non-negative constants. Then $f(\mathbf{x}) = \log\left(\sum_i \; c_i \frac{g_i(\mathbf{x})}{h_i(\mathbf{x})}\right)$ is d.c. and has a d.c. decomposition

$$\log\left(\sum_i \; c_i g_i(\mathbf{x}) \prod_{j \neq i} \; h_j(\mathbf{x})\right) - \sum_i \; \log h_i(\mathbf{x}) \tag{38}$$

Proof:

$$f(\mathbf{x}) = \log\left(\sum_i c_i \frac{g_i(\mathbf{x})}{h_i(\mathbf{x})}\right) \tag{39}$$

$$= \log\left(\frac{\sum_i c_i g_i(\mathbf{x}) \prod_{j\neq i} h_j(\mathbf{x})}{\prod_i h_i(\mathbf{x})}\right) \tag{40}$$

The product of log-convex functions is log-convex, the sum of log-convex is log-convex and a positive constant times a log-convex function is log-convex. Hence the argument of $\log(.)$ in (40) is the ratio of log-convex functions. Therefore, a d.c. decomposition for $f(\mathbf{x})$ is $\log\left(\sum_i c_i g_i(\mathbf{x}) \prod_{j\neq i} h_j(\mathbf{x})\right) - \sum_i \log h_i(\mathbf{x})$. ■

Theorem 3: For a general SNR, the problem of maximizing the cutoff rate with respect to $\boldsymbol{\lambda}$ can be obtained through either

(i) a d.c. program over a polytope, or

(ii) a concave minimization program , or

(iii) a convex minimization program with an additional reverse convex constraint.

Proof: The constraint set is $\sum_n \lambda_n = N_t N_r$, which is a closed and convex set. We can instead use the inequality constraint $\sum_n \lambda_n \leq N_t N_r$, since the cutoff rate expression is an increasing function in γ and hence a solution has to lie on the boundary.

The cutoff rate expression may be written as

$$CR = -\log\left\{\sum_i P_i^2 + \sum_{\substack{i,j \\ j\neq i}} P_i P_j \frac{\left|\mathbf{I} + \frac{\gamma}{2}(\boldsymbol{\mathcal{X}}_i \boldsymbol{\Lambda} \boldsymbol{\mathcal{X}}_i^* + \boldsymbol{\mathcal{X}}_j \boldsymbol{\Lambda} \boldsymbol{\mathcal{X}}_j^*)\right|^{-1}}{\left|\mathbf{I} + \gamma \boldsymbol{\mathcal{X}}_i \boldsymbol{\Lambda} \boldsymbol{\mathcal{X}}_i^*\right|^{-1/2} \left|\mathbf{I} + \gamma \boldsymbol{\mathcal{X}}_j \boldsymbol{\Lambda} \boldsymbol{\mathcal{X}}_j^*\right|^{-1/2}}\right\} \tag{41}$$

Maximizing the expression in (41) is equivalent to maximizing the following expression due to the monotonicity of $\log(c+x)$ and $\log(x)$.

$$-\log\left\{2\sum_{\substack{i \\ j>i}} P_i P_j \frac{\left|\mathbf{I} + \frac{\gamma}{2}(\boldsymbol{\mathcal{X}}_i \boldsymbol{\Lambda} \boldsymbol{\mathcal{X}}_i^* + \boldsymbol{\mathcal{X}}_j \boldsymbol{\Lambda} \boldsymbol{\mathcal{X}}_j^*)\right|^{-1}}{\left|\mathbf{I} + \gamma \boldsymbol{\mathcal{X}}_i \boldsymbol{\Lambda} \boldsymbol{\mathcal{X}}_i^*\right|^{-1/2} \left|\mathbf{I} + \gamma \boldsymbol{\mathcal{X}}_j \boldsymbol{\Lambda} \boldsymbol{\mathcal{X}}_j^*\right|^{-1/2}}\right\}. \tag{42}$$

By Lemma 1, we have that $|\mathbf{I}+\frac{\gamma}{2}(\boldsymbol{\mathcal{X}}_i \boldsymbol{\Lambda} \boldsymbol{\mathcal{X}}_i^*+\boldsymbol{\mathcal{X}}_j \boldsymbol{\Lambda} \boldsymbol{\mathcal{X}}_j^*)|^{-1}$, $|\mathbf{I}+\gamma \boldsymbol{\mathcal{X}}_i \boldsymbol{\Lambda} \boldsymbol{\mathcal{X}}_i^*|^{-1}$ and $|\mathbf{I}+\gamma \boldsymbol{\mathcal{X}}_j \boldsymbol{\Lambda} \boldsymbol{\mathcal{X}}_j^*|^{-1}$ are log-convex functions of $\boldsymbol{\Lambda}$. A log-convex function raised to a positive index is still log-convex. Also, a positive constant times a log-convex function is log-convex. Therefore, the expression (42) can be seen to be in the form needed in Lemma 2, and hence a d.c. decomposition can be obtained. We can further simplify this d.c. decomposition and transform it into other standard global optimization problems as follows. The argument of $-\log(.)$ in (42) can be written as

$$\frac{2\sum_{\substack{i\\j>i}} P_iP_j \left|\mathbf{I}+\frac{\gamma}{2}(\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*+\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*)\right|^{-1}\prod_{\substack{k\neq i\\ l\neq j,k>l}}\left|(\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_k\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_k^*)\left(\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_l\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_l^*\right)\right|^{-1/2}}{\prod_{\substack{i\\j>i}}\left|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*\right|^{-1/2}\left|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*\right|^{-1/2}}.$$

By the properties of log-convex functions, and since the sum and product of log-convex functions is still log-convex, the last expression is a ratio of two log-convex functions. Therefore, taking $-\log(.)$ gives a d.c. decomposition for CR.

We will next express the maximization of (42) as a concave minimization problem. We need to maximize the $-\log(.)$ of the last expression, which is

$$= \max_{\mathrm{tr}(\boldsymbol{\Lambda})\leq N_tN_r} \frac{1}{2}\log\prod_{i,j>i}\left(\left|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*\right|\left|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*\right|\right)^{-1} - q(\boldsymbol{\Lambda}), \tag{43}$$

where $q(\boldsymbol{\Lambda})$ is convex over $\boldsymbol{\Lambda}$. An additional variable t is now introduced to get the equivalent optimization problem

$$\max_{\substack{\mathrm{tr}(\boldsymbol{\Lambda})\leq N_tN_r\\ q(\boldsymbol{\Lambda})\leq t}} \frac{1}{2}\log\prod_{i,j>i}\left(\left|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*\right|\left|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*\right|\right)^{-1} - t \tag{44}$$

$$= \min_{\substack{\mathrm{tr}(\boldsymbol{\Lambda})\leq N_tN_r\\ q(\boldsymbol{\Lambda})\leq t}} t + \sum_{i,j>i}\frac{1}{2}\log\left(\left|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*\right|\left|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*\right|\right) \tag{45}$$

Since $q(\boldsymbol{\Lambda})-t$ is a convex function, $q(\boldsymbol{\Lambda})-t\leq 0$ is a convex set. Since the intersection of convex sets is convex, the constraint set is convex. Further, since the objective function in (45) is concave, the optimization is a concave minimization problem over $\boldsymbol{\Lambda}$ and t by definition.

The additional variable t may also be introduced in place of the other convex function in (43), to get the equivalent problem

$$\max_{\substack{\mathrm{tr}(\boldsymbol{\Lambda})\leq N_tN_r\\ t\leq r(\boldsymbol{\Lambda})}} t-q(\boldsymbol{\Lambda}) \;=\; \min_{\substack{\mathrm{tr}(\boldsymbol{\Lambda})\leq N_tN_r\\ t\leq r(\boldsymbol{\Lambda})}} q(\boldsymbol{\Lambda})-t, \tag{46}$$

where $r(\boldsymbol{\Lambda}) = \sum_{i,j>i} -\frac{1}{2}\log\left(\left|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*\right|\left|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*\right|\right)$. Since $r(\boldsymbol{\Lambda})$ is a convex function of $\boldsymbol{\Lambda}$, $t\leq r(\boldsymbol{\Lambda})$ is a reverse convex constraint of $\boldsymbol{\Lambda}$. Since $q(\boldsymbol{\Lambda})-t$ is a convex function of $\boldsymbol{\Lambda}$, this form of the optimization problem is a convex minimization problem with an additional reverse convex constraint (or concave constraint). ■

All the three forms indicated are global optimization problems and algorithms are available to solve them. An example of an algorithm that solves the d.c. program is the Simplical Branch and Bound algorithm(Section 4.6) [23]. Alternatively, one could convert the problem into a canonical d.c. program and then use the Edge Following Algorithm (Section 4.5) [23]. More algorithms may be found in [23] and another reference [25]. The concave minimization problem is known [23] to have a solution at an extreme point of the constraint region and this fact is exploited in algorithms to solve it. Several algorithms to solve this problem are given in Chapter 3 of [23] and in [25]. The formulation involving a convex minimization problem with an additional reverse convex constraint can be solved by the branch and bound algorithm given in [26].

Since the number of transmit and receive antennas is relatively small in practice, the problem of finding the optimal $\mathbf{\Lambda}$ which involves $N_t N_r$ variables, can be solved numerically with tractable complexity in many practical cases of interest.

D. A Numerical Example

In Figure 1, we give a numerical example using the UIU Rayleigh fading model. In this figure, we compare the simulated (via Monte-Carlo simulations) mutual informations of a systematic unitary constellation at different values of γ on a fully correlated channel, an i.i.d. channel and a channel using the optimal correlations. The constellation used has 8 points and the parameters $N_t = 3, N_r = 3$ and $T = 6$. At relatively low SNR, the performance with optimal correlations coincides with that of the fully correlated channel. At high SNRs, the performance with optimal correlations coincides with that of the i.i.d. fading channel. These simulations are hence in concordance with the analytical results in Theorems 1 and 2. At moderate SNRs, gains of upto $\approx$ 2.5 dB are observed when using the optimal correlations as compared to the better of the i.i.d. or fully correlated case. Significant improvements are observed for the optimal correlations over the i.i.d. fading case.

V. CONCLUSIONS

We considered the problem of finding the optimal correlation matrices of a noncoherent spatially correlated MIMO Rayleigh fading channel at different SNR regimes. In the low SNR regime, we use the mutual information as our design criterion, while at higher SNR regimes we use the cutoff rate expression. At sufficiently low SNR, we showed that a fully correlated channel matrix maximizes the mutual information. This indicates that it is best to focus power along one effective channel in the low SNR regime. Therefore, systems with more densely packed antenna arrays that result in high spatial correlations have a higher capacity at low SNR. At asymptotically high SNR, we showed that a fully uncorrelated channel matrix is optimal under a condition on the constellation which ensures full diversity. This indicates that in the high SNR regime, it helps to create as many independent parallel channels as possible. In the case of separable correlations, we showed that the cutoff rate expression is Schur-convex with respect to the receive correlation eigenvalues at sufficiently low SNR and Schur-concave at high SNR. This indicates that it is beneficial to have high receive correlations at sufficiently low SNR, while it helps to have the receive correlation matrix as close to i.i.d. as possible at high SNR. At sufficiently low SNR, the fully correlated transmit correlation matrix is optimal for any fixed receive correlation matrix. We show that the cutoff rate expression is Schur-concave with respect to the transmit correlation eigenvalues at high SNR. This indicates that it helps to have the transmit correlation matrix as close to i.i.d. as possible at high SNR. We also show how the problem of finding the eigenvalues of the optimal correlation matrix at a general SNR can be formulated and solved by using standard global optimization algorithms.

APPENDIX

A. *Majorization, Schur-convex and Schur-concave functions*

The following two definitions are from [20].

Definition 7: For $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$, $\mathbf{x}$ is said to be majorized by $\mathbf{y}$, denoted by $\mathbf{x} \prec \mathbf{y}$, if

$$\begin{aligned} \sum_{i=1}^{k} x_{[i]} &\leq \sum_{i=1}^{k} y_{[i]}, \quad k = 1, ..., n-1, \\ \text{and} \qquad \sum_{i=1}^{n} x_{[i]} &= \sum_{i=1}^{n} y_{[i]} \end{aligned}$$

where $x_{[i]}$ and $y_{[i]}$ denote the i^{th} largest components of $\mathbf{x}$ and $\mathbf{y}$ respectively.

Definition 8: A real valued function f defined on a set $\boldsymbol{\mathcal{A}} \subseteq \mathbf{R}^n$ is said to be Schur-convex on $\boldsymbol{\mathcal{A}}$ if for any $\mathbf{x}, \mathbf{y} \in \boldsymbol{\mathcal{A}}$, $\mathbf{x} \prec \mathbf{y} \quad \Rightarrow f(\mathbf{x}) \leq f(\mathbf{y})$. Similarly, f is defined to be Schur-concave on $\boldsymbol{\mathcal{A}}$ if for any $\mathbf{x}, \mathbf{y} \in \boldsymbol{\mathcal{A}}$, $\mathbf{x} \prec \mathbf{y} \quad \Rightarrow f(\mathbf{x}) \geq f(\mathbf{y})$.

Since the vector $[n \; 0 \; \ldots \; 0]^T$(with the n occuring at any position) majorizes every other non-negative vector whose elements add up to n, every Schur-convex function of such vectors attains its maximum at $[n\ 0\ \dots\ 0]^T$. Similarly, every Schur-concave function attains its maximum at $[1\ 1\ \dots\ 1]^T$ among all non-negative vectors whose elements add up to n.

B. Derivation of cutoff rate

The integral $\int \sqrt{p(\mathbf{y}|i)\ p(\mathbf{y}|j)}d\mathbf{y}$ in (24) is known as the Bhattacharya coefficient ρ_{ij} between hypotheses i and j. For the noncoherent MIMO Rayleigh fading channel , ρ_{ij} is

$$\begin{aligned}\rho_{ij} &= \int_{\Gamma} \left[\frac{p_j(\mathbf{y})}{p_i(\mathbf{y})}\right]^{1/2} p_i(\mathbf{y})d\mathbf{y} \\ &= \frac{|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*|^{1/2}}{\left|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*\right|^{1/2}} E_{\boldsymbol{\mathcal{X}}_i}[\exp\left(-\mathbf{y}^*\mathbf{F}_{ji}\mathbf{y}\right)],\end{aligned} \tag{47}$$

where $\mathbf{F}_{ji} = \frac{1}{2}(\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*)^{-1} - \frac{1}{2}(\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*)^{-1}$. The expectation in (47) can be evaluated using the main result in [27] to get

$$\begin{aligned}\rho_{ij} &= \frac{|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*|^{1/2}}{\left|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*\right|^{1/2}\left|\frac{1}{2}\mathbf{I}+\frac{1}{2}(\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*)(\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*)^{-1}\right|} \\ &= \frac{|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*|^{1/2}\left|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*\right|^{1/2}}{\left|\frac{1}{2}(\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*)+\frac{1}{2}(\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*)\right|}\end{aligned} \tag{48}$$

Substituting these expressions in (24) we get (25).

In the special case of separable correlations, we may simplify (48) further to obtain the following expression :

$$\rho_{ij} = \prod_{n=1}^{N_r} \frac{|\mathbf{I}+\gamma\mathbf{X}_i\boldsymbol{\Lambda}\mathbf{X}_i^*\lambda_n^r|^{1/2}\left|\mathbf{I}+\gamma\mathbf{X}_j\boldsymbol{\Lambda}\mathbf{X}_j^*\lambda_n^r\right|^{1/2}}{\left|\mathbf{I}+\frac{\gamma}{2}(\mathbf{X}_i\boldsymbol{\Lambda}\mathbf{X}_i^*+\mathbf{X}_j\boldsymbol{\Lambda}\mathbf{X}_j^*)\lambda_n^r\right|} \tag{49}$$

Equation (49) follows from (48) using the relations

$$\begin{aligned}\mathbf{I}+\gamma\lambda_n^r\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^* &= \mathbf{I}_{TN_r}+(\mathbf{I}_{N_r}\otimes\gamma\mathbf{X}_i)(\boldsymbol{\Lambda}_r\otimes\boldsymbol{\Lambda})(\mathbf{I}_{N_r}\otimes\lambda_n^r\mathbf{X}_i)^* \\ &= \mathbf{I}_{N_r}\otimes\mathbf{I}_T+(\mathbf{I}_{N_r}\otimes\gamma\mathbf{X}_i\boldsymbol{\Lambda}\mathbf{X}_i^*\lambda_n^r) \\ &= \mathbf{I}_{N_r}\otimes(\mathbf{I}_T+\gamma\mathbf{X}_i\boldsymbol{\Lambda}\mathbf{X}_i^*\lambda_n^r),\end{aligned}$$

and simplifying.

C. Derivation of low SNR cutoff rate

In this appendix, we derive the low SNR cutoff rate. The cutoff rate expression may be written as

$$CR = -\log\left\{\sum_i\sum_j P_iP_j\exp\left\{-\log\frac{\left|\mathbf{I}+\frac{\gamma}{2}(\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*+\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*)\right|}{\left|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*\right|^{1/2}\left|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*\right|^{1/2}}\right\}\right\} \quad (50)$$

$$= -\log\left\{\sum_{i,j} P_iP_j\,\mathrm{e}^{\{\frac{1}{2}\log|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*|+\frac{1}{2}\log\left|\mathbf{I}+\gamma\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*\right|-\log\left|\mathbf{I}+\frac{\gamma}{2}(\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*+\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*)\right|\}}\right\} \quad (51)$$

Now apply the formula $\log|\mathbf{I}+\gamma\mathbf{A}| = \gamma\mathrm{tr}(\mathbf{A}) - \frac{\gamma^2}{2}\mathrm{tr}(\mathbf{A}^2) + o(\gamma^2)$, which is valid for any Hermitian matrix $\mathbf{A}$ and small γ. With this approximation and some simplification, we get that

$$CR_{low} = -\log\left\{\sum_i\sum_j P_iP_j\,\mathrm{e}^{-\frac{\gamma^2}{8}\mathrm{tr}\{(\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*-\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*)^2\}\,+\,o(\gamma^2)\}}\right\} \quad (52)$$

$$= -\log\left\{\sum_i\sum_j P_iP_j\left(1-\frac{\gamma^2}{8}\mathrm{tr}\{(\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*-\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*)^2\}\,+\,o(\gamma^2)\right)\right\} \quad (53)$$

$$= -\log\left\{1-\left\{\sum_i\sum_j P_iP_j\frac{\gamma^2}{8}\mathrm{tr}\{(\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*-\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*)^2\}+\,o(\gamma^2)\right\}\right\} \quad (54)$$

$$= \frac{\gamma^2}{8}\sum_i\sum_j P_iP_j\mathrm{tr}\{(\boldsymbol{\mathcal{X}}_i\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_i^*-\boldsymbol{\mathcal{X}}_j\boldsymbol{\Lambda}\boldsymbol{\mathcal{X}}_j^*)^2\}+o(\gamma^2). \quad (55)$$

In (53), we have used the approximation $\exp(-x) = 1-x+o(x)$ which holds for small x. In (55), we have used the approximation $-\log(1-x) = x-o(x)$ which is true for small x.

REFERENCES

[1] C. Chuah, D. Tse, J. Kahn, and R. Valenzuela, "Capacity Scaling in MIMO Wireless Systems under Correlated Fading," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

[2] A. M. Tulino, A. Lozano, and S. Verdu, "The Impact of Antenna Correlation on the Capacity of Multiantenna Channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2491–2509, July 2005.

[3] A. Lozano, A. M. Tulino, and S. Verdu, "Multiple-Antenna Capacity in the Low-Power regime," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2527–2544, Oct. 2003.

[4] E. A. Jorswieck and H. Boche, "Optimal Transmission Strategies and Impact of Correlation in Multiantenna Systems with Different Types of Channel State Information," *IEEE Trans. Signal Processing*, vol. 52, no. 12, pp. 3440–3453, Dec. 2004.

[5] J. Wang, M. K. Simon, M. P. Fitz, and K. Yao, "On the performance of space-time codes over spatially correlated Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 52, no. 6, pp. 877–881, June 2004.

[6] J. H. Kotecha and A. M. Sayeed, "Transmit Signal Design for Optimal Estimation of Correlated MIMO Channels," *IEEE Trans. Wireless Commun.*, vol. 52, no. 2, pp. 546–557, Feb. 2004.

[7] H. Bolcskei and A. J. Paulraj, “On the performance of space-time codes in the presence of spatial fading correlations,” in *Proc. Asilomar Conference on Signals, Systems and Computers*, California, USA, Sept. 2000.

[8] A. M. Sayeed, “Deconstructing multiantenna fading channels,” *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[9] W. Weichselberger, M. Herdin, H. Ozcelik, and E. Bonek, “A stochastic MIMO channel model with joint correlation of both link ends,” to appear *IEEE Trans. Wireless Commun.*

[10] X. Wu and R. Srikant, “MIMO Channels in the Low SNR Regime: Communication Rate, Error Exponent and Signal Peakiness,” submitted to *IEEE Trans. Inform. Theory*, Apr. 2004.

[11] T. L. Marzetta and B. M. Hochwald, “Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading,” *IEEE Trans. Inform. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[12] L. Zheng and D. N. C. Tse, “Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel,” *IEEE Trans. Inform. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[13] B. Hassibi and T. L. Marzetta, “Multiple-antennas and isotropically-random unitary inputs: The received signal density in closed-form,” *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1473–1484, June 2002, Special Issue on Shannon Theory: Perspective, Trends, and Applications.

[14] S. A. Jafar and A. Goldsmith, “Multiple-antenna capacity on correlated Rayleigh fading with channel covariance information,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 990–997, May 2005.

[15] A. O. Hero, III and T. L. Marzetta, “Cut-off rate and signal design for the Rayleigh fading space–time channel,” *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2400–2416, Sept. 2001.

[16] S. G. Srinivasan and M. K. Varanasi, “Constellation Design for the Noncoherent MIMO Rayleigh Fading Channel at General SNR,” submitted to *IEEE Trans. Inform. Theory*, Mar. 2006.

[17] C. Rao and B. Hassibi, “Analysis of multiple-antenna wireless links at Low SNR,” *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 2123 – 2130, Sept. 2004.

[18] V. Prelov and S. Verdú, “Second-order asymptotics of mutual information,” *IEEE Trans. Inform. Theory*, vol. 50, no. 8, pp. 1567–1580, Aug. 2004.

[19] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005.

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

[21] M. Brehler and M. K. Varanasi, “Asymptotic error probability analysis of quadratic receivers in Rayleigh fading channels with applications to a unified analysis of coherent and noncoherent space–time receivers,” *IEEE Trans. Inform. Theory*, vol. 47, no. 5, pp. 2383–2399, Sept. 2001.

[22] M. C. Gursoy, H. V. Poor, and S. Verdú, “Noncoherent Rician Fading channel - Part II: Spectral Efficiency in the Low-Power Regime,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2207–2221, Sept. 2005.

[23] R. Horst, P. M. Pardalos, and N. V. Thoai, *Introduction to Global Optimization*, Kluwer, 2000.

[24] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, U.K., 2004.

[25] R. Horst and H. Tuy, *Global Optimization*, Springer, 1996.

[26] H. Tuy, “Convex programs with an additional reverse convex constraint,” *Journal of Optimization Theory and Application*, vol. 52, no. 3, pp. 463 – 486, Mar. 1987.

[27] G. L. Turin, “The characteristic function of hermitian quadratic forms in complex normal variables,” *Biometrika*, vol. 47, pp. 199–201, June 1960.

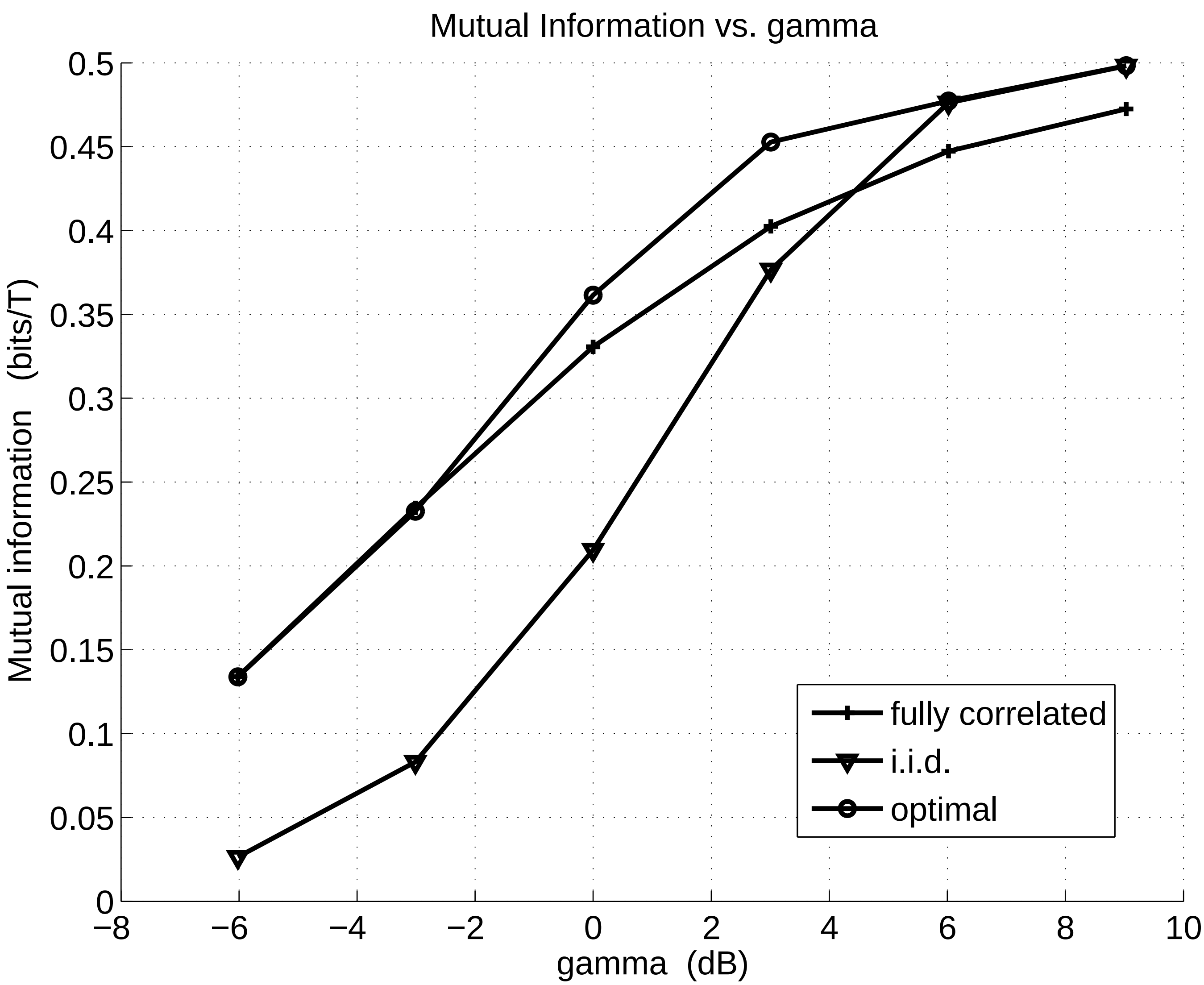

Fig. 1. Mutual information plot for systematic unitary design with $L = 8, T = 6, N_t = 3$ and $N_r = 3$.