

MIMO Channels in the Low SNR Regime: Communication Rate, Error Exponent and Signal Peakiness *

Xinzhou Wu and R. Srikant

Department of Electrical and Computer Engineering

and

Coordinated Science Lab

University of Illinois at Urbana-Champaign

xwu@uiuc.edu, rsrikant@uiuc.edu

December 21, 2006

Abstract

We consider MIMO fading channels and characterize the reliability function in the low-SNR regime as a function of the number of transmit and receive antennas. For the case when the fading matrix \mathbf{H} has independent entries, we show that the number of transmit antennas plays a key role in reducing the peakiness in the input signal required to achieve the optimal error exponent for a given communication rate. Further, by considering a correlated channel model, we show that the maximum performance gain (in terms of the error exponent and communication rate) is achieved when the entries of the channel fading matrix are fully correlated. The results we presented in this work in the low-SNR regime can also be applied to the infinite bandwidth regime.

1 Introduction

In this paper, we study multiple input and multiple output (MIMO) antenna channels in the low-SNR regime. Specifically, we compute the reliability function for such channels and also characterize the signal peakiness required to achieve the maximum error exponent for a given communication rate. We use a block-fading model, which is widely used in the MIMO literature [18, 25], and assume that the fading matrix \mathbf{H} remains constant

*Research supported by Vodafone Fellowship 2003-2004 at University of Illinois.

for a certain time period and then changes to an independent value for the next time period. The length of this time period is usually referred to as the *coherence time* of the channel. Throughout this paper, we use a discrete channel model which can be obtained by sampling the input and output waveform at rate $\frac{1}{B_w}$, letting B_w be the available bandwidth of the channel. We use the term *coherence time* to denote the number of samples which share the same realization of the random fading process. Thus, the coherence time is measured in units of $1/B_w$.

Letting M be the number of transmit antennas, N be the number of receive antennas and T_c be the coherence time, the channel we study in this paper is

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{W}, \quad (1)$$

where $\mathbf{Y}, \mathbf{W} \in \mathcal{C}^{N \times T_c}$, $\mathbf{X} \in \mathcal{C}^{M \times T_c}$ and $\mathbf{H} \in \mathcal{C}^{N \times M}$. Here \mathbf{W} is a proper complex Gaussian noise random matrix, which has i.i.d. entries, i.e., $w_{nt} \sim \mathcal{CN}(0, 1)$, for $1 \leq n \leq N$ and $1 \leq t \leq T_c$, where w_{nt} denotes the (n, t) th entry of \mathbf{W} . We also assume \mathbf{H} is a proper complex Gaussian matrix which has symmetric complex Gaussian entries such that $h_{nm} \sim \mathcal{CN}(0, 1)$, for $1 \leq m \leq M$ and $1 \leq n \leq N$. (We refer the reader to [11] for the definition of a proper complex matrix.) In other words, we consider Rayleigh fading MIMO channels in this paper. At this point, we do not make any assumptions regarding the correlation among the entries of \mathbf{H} . Later, we will consider two cases, one where the entries of \mathbf{H} are independent and the other where they are correlated.

Throughout this paper, we assume that neither the transmitter nor the receiver has knowledge of the realization of the random matrix \mathbf{H} . However, we always assume the distribution of \mathbf{H} is known *a priori* at both the transmitter and the receiver. We will study the reliability function of the channel (1) in the low-SNR regime under the following standard average power constraint:

$$\mathcal{E}[\|\mathbf{X}\|^2] \leq pT_c, \quad (2)$$

where $\|\mathbf{X}\|$ is the *Frobenius norm* [5] of the matrix \mathbf{X} . A special case of this model is the case when $M = N = T_c = 1$, which we will refer to as a SISO fast-fading channel.

The MIMO channel capacity in the low-SNR regime when the receiver has the channel side information has been studied by [9]. Here, we study the non-coherent case, and further, instead of just studying the capacity, we compute the low-SNR reliability function. This has been done for the coherent case in [12]. We would like to point out here our results are first-order results, unlike the capacity analysis in [22] [9]. In other words, we only study the first order behavior of the non-coherent MIMO channels as the available SNR goes to 0. The notion of *wideband slope*, which is a second order quantity, is not studied in this work.

The reliability function of a channel is defined as

$$E(R, p) = \lim_{N_c \rightarrow \infty} -\frac{\ln P_e(N_c, R)}{N_c T_c}, \quad (3)$$

where $P_e(N_c, R)$ is the minimum average probability of decoding error for any block code with block length N_c and coding rate R , which satisfies the average power constraint (2). The rate R is defined in nats per symbol and is equal to $\frac{\ln M_c}{N_c T_c}$, where M_c denotes the size of the code book, and N_c is the codeword length measured in number of channel uses, where each channel use consists of the transmission of T_c symbols.

It is well known that with multiple antennas at the receiver and/or the transmitter, the channel performance in terms of communication rate and/or decoding error probability can be improved. The extent of this improvement is now well understood in the high SNR regime. From a capacity point of view, when the elements of \mathbf{H} are i.i.d. and known at the receiver, we can achieve a degree of freedom gain (or multiplexing gain) up to $\min(M, N)$ [18]. From a probability of decoding error point of view, we can achieve a *diversity gain* up to MN with space-time coding at the input [17]. A recent work by Zheng and Tse [25] illustrated the tradeoff between these two gains in the high-SNR regime by showing the best degree of freedom gain and diversity gain that can be achieved *simultaneously*. They also extend this result to a non-coherent MIMO channel in a later work [26]. Roughly speaking, in an error-exponent theory framework, their results characterize the following curve

$$d^*(r) = \lim_{p \rightarrow \infty} \frac{E(r \ln p, p)}{\ln p},$$

where the degree of freedom gain r can be interpreted as the normalized communication rate at high SNR and the diversity gain $d^*(r)$ is nothing but the normalized reliability function at rate r .

In the low-SNR regime, i.e., the average power constraint $p \rightarrow 0$, both the reliability function and the capacity approach 0 linearly in p . Thus, it is natural to define the normalized reliability function in the low-SNR regime as follows:

$$E^{lp}(r) = \lim_{p \rightarrow 0} \frac{E(rp, p)}{p}, \quad \text{if the limit exists.} \quad (4)$$

Here the superscript stands for *low power*. For simplicity, from now on, we will refer to $E^{lp}(r)$ as the low-SNR reliability function, and r as the low-SNR communication rate. However, it should be clear that both quantities are normalized by p in the low-SNR regime.

In the low-SNR regime with non-coherent fading channels, it is well known that a “peaky” signaling scheme is required to achieve capacity [7, 19, 10, 16, 22], in the sense that, as the average power goes to zero, the optimal signaling scheme is to keep silent most of the time and use symbols of very large amplitude (going to infinity as we approach capacity) when transmitting. In other words, in the low-SNR regime, good channel performance in terms of communication rate can only be achieved at the cost of large signaling peakiness, which is undesirable in practical communication systems. Recently, much research has been devoted to studying the performance of low-SNR non-coherent channels with certain peakiness constraints [10, 16]. In this paper, we

study the relationship between *signaling peakiness*, measured by the maximum symbol amplitude in the low-SNR limit, and the channel reliability function, which is a more complete characterization of channel performance as compared to *channel capacity*. Specifically, we show that a scheme similar to the capacity achieving scheme, is required to achieve the optimal decoding error probability at an arbitrary communication rate between 0 and capacity. Further, we quantify the *minimum signal peakiness* $\tau(r)$ required to achieve the optimal decoding error probability at rate r . (We will precisely define this quantity in Section 3.) Further, in terms of the benefit brought by introducing multiple antennas in the channel, we define two gains: *performance gain* and *peakiness gain*. While the performance gain focuses on the scale of improvement in the more traditional performance measures, data rate and error exponent, the peakiness gain characterizes the reduction in the signaling peakiness due to the use of multiple antennas, which is of great importance as well for communications in the low-SNR regime.

Specifically, the performance gain and the peakiness gain are defined as follows.

Definition 1 We say a MIMO block-fading channel has a performance gain of L if

$$E^{lp}(r) = L\bar{E}^{lp}\left(\frac{r}{L}\right),$$

where \bar{E}^{lp} denotes the reliability function for a SISO fast-fading channel. ◇

Note that the performance gain as defined above provides simultaneous gains in both the error exponent and the communication rate. In other words, a performance gain of L indicates that the transmission rate has increased by a factor of L and the error exponent has increased by a factor of L as compared to a SISO channel.

Definition 2 We say a MIMO block-fading channel, which has a performance gain of L , has a peakiness gain of G if the minimum signaling peakiness $\tau(r)$ required to achieve the optimal decoding error probability at communication rate r satisfies

$$\tau(r) = \frac{1}{G}\bar{\tau}\left(\frac{r}{L}\right),$$

where $\bar{\tau}(r)$ denotes the minimum signal peakiness at rate r for a SISO fast-fading channel.

The performance gain tells us how much we can improve the channel performance (both the capacity and the reliability function) by using multiple antennas. On the other hand, since signals with very large amplitudes are undesirable in practice, the peakiness gain quantifies the effect of using multiple antennas in reducing the peakiness necessary to achieve a certain performance level.

The main contributions of the paper are summarized below:

- (1) When \mathbf{H} is independently faded, we characterize the low-SNR reliability function for the MIMO fading channel (1) for all rates between 0 and capacity.
- (2) When \mathbf{H} is independently faded, analogous to the result in [22] which shows that only *flash signaling* schemes are capacity-achieving in the low-SNR regime, we show that only a certain kinds of signaling schemes, which we call *general flash signaling* schemes, can achieve the optimal probability of decoding error at a certain communication rate. We quantify the *minimum signal peakiness* $\tau(r)$ for a signaling scheme to achieve the optimum probability of decoding error for a certain communication rate r . As r approaches the low-SNR capacity, $\tau(r)$ goes to infinity.
- (3) Again for the case where \mathbf{H} is independently faded, by using M transmit antennas and N receive antennas, we show that one can achieve a *performance gain* of N and a *peakiness gain* of MT_c . While increasing the number of transmit antennas does not improve the error exponent, it reduces the minimum signal peakiness required to achieve a certain point on the reliability function curve. We also show that, for a MIMO channel where both the average power and peak signal amplitude are constrained, using multiple transmit antennas can improve *both* the capacity and the error exponent.
- (4) Finally for a correlated Gaussian fading matrix \mathbf{H} , an upper bound on the low-SNR reliability function is established. Further, the optimal reliability function, optimized over fading matrix \mathbf{H} , is achieved when the entries of \mathbf{H} are *fully correlated*, i.e., the entries of \mathbf{H} are (complex) multiples of the same random variable. Two such examples include the case where all the entries are identical and the case where all entries are identical except for a phase shift. For this fully correlated model, having multiple transmit (or receive) antennas can improve both the reliability function and reduce the peakiness required for a certain performance level. In fact, we show that potentially one can obtain a performance gain of MN and peakiness gain of M^2NT_c . Thus, one of the conclusions of this paper is that correlated channels are better than independent fading in the low-SNR regime which indicates that spacing antennas far away from each other to ensure independent fading may not be the right thing to do in the low-SNR regime.

Although in this paper, we are considering the low-SNR regime where the capacity and communication rate are vanishingly small, the results we obtained here can also be applied to the wideband regime, where the number of degrees of freedom is large and reliable communication is possible at a non-vanishing rate. To see this, we consider a *doubly block fading* channel model as in [24, 20], where the block fading assumption is used both in the time and the frequency dimensions. A fading channel with bandwidth BW_c , where W_c is the coherence

bandwidth of the multipath fading, can be treated as B parallel independent fading channels each with bandwidth W_c and for each sub-channel, flat fading can be assumed. The channel model we study in (1) is simply one of the sub-channels. Under these assumptions, it can be shown that¹ the reliability function of the parallel channel can be expressed in terms of the reliability function of one of the sub-channels, i.e.,

$$E(R, P, B) = BE(R/B, P/B, 1),$$

where $E(R, P, B)$ denotes the reliability function for a parallel channel with B independent sub-channels at communication rate R and average power P . From here, it can be easily seen that the wideband limit of the reliability function is

$$\lim_{B \rightarrow \infty} E(R, P, B) = \lim_{B \rightarrow \infty} BE(R/B, P/B, 1) = P \lim_{p \rightarrow 0} \frac{E(\frac{R}{P}p, p)}{p} = PE^{lp}(R/P),$$

which connects the wideband reliability function to the low-SNR reliability function we defined in (4). Further, all the implications we obtain in the low-SNR regime also find their counterpart in the wideband regime if we measure the coherence time in units of $1/W_c$. Such connections between the wideband regime and the low SNR regime are made based on the assumption of the validity of the concept of *block fading* in frequency and time. This model does not capture the correlation between the coherence blocks and is not entirely realistic. In the high-SNR regime, the behavior of the channel capacity in terms of SNR would change when a more sophisticated model is introduced [8]. However, for tractability of the computation of the reliability function in the low-SNR regime, the block fading model seems to be a good start. It is still an open problem to quantify the influence of the correlation between different coherent blocks on the reliability function in the low-SNR regime.

The rest of the paper is organized as follows. In Section 2, we start with a general channel with an arbitrary transition probability density function $f(y|x)$. Under the constraint that the input belongs to a finite and discrete alphabet, we show that the low-SNR reliability function has a closed-form expression by applying Gallager's result in [3] with a few modifications. In Section 3, we calculate the low-SNR reliability function under the assumption that \mathbf{H} is independently faded. In Section 4, correlated fading MIMO channels are considered. We provide some concluding remarks in Section 5.

Throughout this paper, we will use small letters for scalar variables, capital letters for constants, boldfaced small letters for vectors and boldfaced capital letters for matrices. We use \mathbf{A}^T to denote the transpose, \mathbf{A}^* to denote the complex conjugate and \mathbf{A}^\dagger to denote the conjugate transpose of matrix \mathbf{A} , i.e.,

$$\mathbf{A}^\dagger = (\mathbf{A}^*)^T.$$

¹See Lemma 3.1 and the proof of Theorem 3.2 of [23] for the case of B parallel SISO channels; the proofs there can be extended in a straightforward manner to parallel MIMO channels.

Finally, we use $\mathcal{E}[\cdot]$ to denote *expectations* and $E(\cdot)$ to denote reliability functions.

2 Preliminaries

In this section, we provide a quick review of the error exponent theory [2] and present a theorem characterizing the low SNR reliability function for an arbitrary memoryless channel by extending the results in [3] to allow less restrictive input distributions. This theorem will be used extensively in the rest of the paper.

We consider a general discrete-time memoryless channel with an arbitrary input alphabet A , and its output determined by the transition probability function $f(y|x)$. Further, we have a cost function $b : A \rightarrow R^+$ associated with each input symbol $x \in A$, and assume that the input to the channel is constrained by the following average cost constraint

$$\mathcal{E}[b(x)] \leq p. \quad (5)$$

Definition 3 [2] *Let $P_e(\tilde{N}_c, R)$ be the minimum probability of error for any block code of block length \tilde{N}_c (the number of channel uses) and rate R , which satisfies the average power constraint p (5), for a given channel. The reliability function $E(R, p)$ of this channel is defined as*

$$E(R, p) = \lim_{\tilde{N}_c \rightarrow \infty} -\frac{\ln P_e(\tilde{N}_c, R)}{\tilde{N}_c}. \quad (6)$$

◇

Next we specify the additional constraint on the input distributions besides the average power constraint. Throughout this paper, we only consider input distributions with a discrete and finite alphabet. Specifically, at a SNR level p , we constrain the input distributions to be in the following set

Definition 4 *Define $\mathcal{D}(p) = \{q(x) : \mathcal{E}[b(x)] = p; \text{ support of } q(x) \text{ is an arbitrary finite set of discrete points in } A\}$.*

If A itself is a finite set, this constraint is automatically satisfied. However, in many situations, A is a continuous set. For example, for our channel model (1), A is the set of complex matrices with dimension $M \times T_c$.

Note that unlike in [3], we do not require this alphabet set to be fixed as p goes to 0 and unlike [4], we have an average power constraint rather than a peak power constraint, i.e., we have the freedom to choose arbitrarily large symbols to transmit even when the power constraint goes to 0. This freedom is especially crucial for the non-coherent fading channel model since in that case, constraining the peak power to go to zero is highly sub-optimal, as we will further explore in the next sections. The only restriction we impose is the discreteness and finiteness of the support of the input distribution, which is a reasonable assumption for any practical coding scheme.

Under this constraint on the input distribution, our goal in this section is to find the low-SNR reliability function $E^{lp}(r)$ of this channel model, as defined by (4), by borrowing tools from both [3] and [21]. For simplicity, in this paper, we assume that a *unique* zero-cost symbol always exist in the input alphabet, i.e., $\exists x \in A$ satisfying $b(x) = 0$. Without loss of generality, we let 0 be the zero-cost symbol and define $A' = A \setminus \{0\}$.

Before we get into the low-SNR regime, we first summarize various bounds for the reliability function $E(R)$ in the following theorem.

Theorem 1 [2] *Assume that the input distribution is constrained to be in $\mathcal{D}(p)$. The reliability function $E(R)$ satisfies*

$$\max(E_r(R), E_{ex}(R)) \leq E(R) \leq \min(E_{sp}(R), E_{sl}(R)),$$

where $E_r(R)$, $E_{ex}(R)$, $E_{sp}(R)$ and $E_{sl}(R)$ denote the random-coding bound, expurgated bound, sphere-packing bound and straight-line bound respectively. The first three bounds can be characterized by the following series of equations:

$$\begin{aligned} E_r(R) &= \sup_{0 \leq \rho \leq 1} -\rho R + E_o(p, \rho), \\ E_{sp}(R) &= \sup_{\rho \geq 0} -\rho R + E_o(p, \rho), \end{aligned} \tag{7}$$

$$\begin{aligned} E_{ex}(R) &= \sup_{\rho \geq 1} -\rho R + E_x(p, \rho), \\ E_o(p, \rho) &= \sup_{q \in \mathcal{D}(p)} \sup_{\beta \geq 0} -\ln \int \left(\int q(x) e^{\beta(b(x)-p)} f(y|x)^{\frac{1}{1+\rho}} dx \right)^{1+\rho} dy, \end{aligned} \tag{8}$$

$$E_x(p, \rho) = \sup_{q \in \mathcal{D}(p)} \sup_{\beta \geq 0} -\rho \ln \int \int q(x_1) q(x_2) e^{\beta(b(x_1)+b(x_2)-2p)} \left(\int \sqrt{f(y|x_1) f(y|x_2)} dy \right)^{\frac{1}{\rho}} dx_1 dx_2. \tag{9}$$

The straight-line bound $E_{sl}(R)$ is the smallest linear function of R which touches the curve $E_{sp}(R)$, and $E_{sl}(0)$ is equal to the zero-rate error exponent $E(0)$. \diamond

It should be noted that most of these results are originally derived for channels with discrete input and output alphabets and no average power constraint. To verify the validity of these results for a channel with continuous output alphabet and average power-constrained input, please refer to Gallager's work [3] and discussions in [24].

Our main theorem in this section is a generalization of Gallager's result in [3] and is stated below.

Theorem 2 *The low-SNR reliability function $E^{lp}(r)$ satisfies*

$$E_r^{lp}(r) \leq E^{lp}(r) \leq \min(E_{sp}^{lp}(r), E_{sl}^{lp}(r)). \tag{10}$$

Here E_{sp}^{lp} and E_r^{lp} are defined as the following

$$E_r^{lp}(r) = \sup_{0 \leq \rho \leq 1} -\rho r + \tilde{E}_o(\rho); \quad (11)$$

$$E_{sp}^{lp}(r) = \sup_{\rho \geq 0} -\rho r + \tilde{E}_o(\rho), \quad (12)$$

and $\tilde{E}_o(\rho)$ is characterized by the following equation

$$\tilde{E}_o(\rho) = \sup_{x \in A'} \frac{-(1 + \rho) \ln \int f(y|0)^{\frac{\rho}{1+\rho}} f(y|x)^{\frac{1}{1+\rho}} dy}{b(x)}. \quad (13)$$

The low-SNR straight-line bound $E_{sl}^{lp}(r)$ is the smallest linear function of r which touches $E_{sp}^{lp}(r)$ and satisfies

$$E_{sl}^{lp}(0) = \sup_q \frac{\mathcal{E} \left[-\ln \int f(y|x_1)^{\frac{1}{2}} f(y|x_2)^{\frac{1}{2}} dy \right]}{\mathcal{E}[b(x)]}, \quad (14)$$

and the sup is over all possible probability distributions with a discrete and finite alphabet set.

Proof: The proof of this theorem is similar to Gallager's proof in [3] with a few modifications to allow the input distributions to use different symbols in different SNR levels. For completeness, we provide a self-contained proof in Appendix A. \diamond

Remarks: The low-SNR bounds here have similar forms as the bounds in [3]. However, as we mentioned before, we are allowing the input distribution to use symbols as a function of p , instead of fixing them. Surprisingly, the only difference in our results as compared to Gallager's results is that in the expression of $\tilde{E}_o(\rho)$ (13), we have a sup over all possible non-zero input symbols rather than a max over the non-zero alphabet as in [3]. In our case, a maximum does not necessarily exist in (13). For example, for coherent fading channels, which we will not explore in this paper, the supremum is achieved by setting x' to be arbitrarily close to 0.

Our result in (13) has a close connection with the capacity per unit cost expression in [21], since the capacity per unit cost is always equal to the first-order derivative of capacity with respect to cost at zero cost, due to the concavity of the capacity formula. As we will show in the proof, we use bounding techniques similar to the ones used in [21] to get the low-SNR sphere-packing bound (12). The sphere-packing exponent and random-coding exponent coincide with each other for any rate larger than r_{crit} and thus provide the closed-form reliability function $E^{lp}(r)$ for this rate region. The capacity per unit cost is simply the largest rate r such that the low-SNR reliability function $E^{lp}(r) > 0$, which is the derivative of $\tilde{E}_o(\rho)$ (13) with respect to ρ at $\rho = 0$. To check this, we perform the following rough calculation:

$$C^{lp} = \lim_{\rho \rightarrow 0} \frac{\tilde{E}_o(\rho)}{\rho}$$

$$\begin{aligned}
&= \lim_{\rho \rightarrow 0} \sup_{x \in A'} \frac{-(1+\rho) \ln \int f(y|0)^{\frac{\rho}{1+\rho}} f(y|x)^{\frac{1}{1+\rho}} dy}{b(x)\rho} \\
&= \sup_{x \in A'} \lim_{\rho \rightarrow 0} \frac{-(1+\rho) \ln \int f(y|0)^{\frac{\rho}{1+\rho}} f(y|x)^{\frac{1}{1+\rho}} dy}{b(x)\rho} \\
&= \sup_{x \in A'} \left. \frac{d}{d\rho} \frac{-(1+\rho) \ln \int f(y|0)^{\frac{\rho}{1+\rho}} f(y|x)^{\frac{1}{1+\rho}} dy}{x} \right|_{\rho=0} \\
&= \sup_{x \in A'} \frac{1}{b(x)} \int f(y|x) \ln \frac{f(y|x)}{f(y|0)} dy \\
&= \sup_{x \in A'} \frac{D(f(y|x)||f(y|0))}{b(x)},
\end{aligned}$$

which is the same as the result in Theorem 3 of [21]. Note the calculations above are quite rough since we exchange the order of lim and sup, integration and differentiation without carefully checking the technical conditions. However, it is not difficult to remove all these technicalities and rigorously prove the relation between our result and Verdu's result.

If the straight-line bound at zero-rate $E_{sl}^{lp}(0)$ is upper bounded by the zero-rate random-coding exponent, i.e., if $E_{sl}^{lp}(0) \leq \tilde{E}_o(1)$, then the lower bound and upper bound in (10) will coincide with each other and thus give us the true reliability function for all rates. We state this observation as the following corollary:

Corollary 1 *If*

$$E_{sl}^{lp}(0) \leq \tilde{E}_o(1), \quad (15)$$

then $E^{lp}(r) = E_r^{lp}(r)$ for all rates. \diamond

As we will show later, for the channel models in this paper, the condition (15) can always be satisfied. However, this is not always true for a general transition probability density function $f(y|x)$.

3 Independent MIMO fading channels

In this section, we will assume that the entries of \mathbf{H} are independent and apply Theorem 2 to find the low-SNR reliability function for the MIMO fading channel (1). Further, we will characterize the asymptotic optimal signaling schemes which minimizes the average probability of decoding error for a certain communication rate.

3.1 Low-SNR reliability function

Since we assume that \mathbf{H} has i.i.d. entries, given an input matrix \mathbf{X} , the row vectors of the output matrix \mathbf{Y} are independent of each other. Denote $\mathbf{Y}^T = (\mathbf{y}_1^*, \mathbf{y}_2^*, \dots, \mathbf{y}_N^*)$, $\mathbf{H}^T = (\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_N)$ and $\mathbf{W}^T =$

$(\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_N)$. The covariance matrix of $\mathbf{y}_i^\dagger = \mathbf{h}_i^T \mathbf{X} + \mathbf{w}_i^T$ is

$$\Sigma_i = \mathcal{E}[\mathbf{y}_i \mathbf{y}_i^\dagger] = \mathbf{I} + \mathbf{X}^\dagger \mathbf{X}.$$

Then, the channel transition probability density function $f(\mathbf{Y}|\mathbf{X})$ is as follows

$$\begin{aligned} f(\mathbf{Y}|\mathbf{X}) &= \prod_{i=1}^N f(\mathbf{y}_i|\mathbf{X}) \\ &= \prod_{i=1}^N \frac{1}{\pi^{T_c} \det(\mathbf{I} + \mathbf{X}^\dagger \mathbf{X})} \exp\{-\mathbf{y}_i^\dagger (\mathbf{I} + \mathbf{X}^\dagger \mathbf{X})^{-1} \mathbf{y}_i\} \end{aligned}$$

Applying Theorem 2 here, we have the following result.

Theorem 3 *For the MIMO model (1) with M transmit antennas and N receive antennas and independent fading matrix \mathbf{H} , the low-SNR reliability function is given by*

$$E^{lp}(r) = \sup_{0 \leq \rho \leq 1} -\rho r + \tilde{E}_o(\rho), \quad (16)$$

where $\tilde{E}_o(\rho)$ is defined as follows

$$\tilde{E}_o(\rho) = N(1 + \rho) \sup_{x > 0} \frac{1}{x} \ln \frac{1 + \frac{\rho}{1+\rho}x}{(1+x)^{\frac{\rho}{1+\rho}}}. \quad (17)$$

Remarks: Consider the SISO fast fading model where $M = N = T_c = 1$. The reliability function curve corresponding to this model is

$$E^{lp}(r) = \sup_{0 \leq \rho \leq 1} -\rho r + (1 + \rho) \sup_{x > 0} \frac{1}{x} \ln \frac{1 + \frac{\rho}{1+\rho}x}{(1+x)^{\frac{\rho}{1+\rho}}}. \quad (18)$$

This form of reliability function first appeared in a slightly different form in [7] as the reliability function for an infinite-bandwidth waveform fading dispersive channel when the input is constrained to frequency-shift-keying. Recently, Subramanian [15] also obtained the same curve for WSSUS fading channels as an upper bound for the random-coding bound, by directly applying Gallager's results [3] for an arbitrary fixed input alphabet, which does not change as p goes to 0. As compared to these two results, our result in (18) is for a simpler discrete-time flat-fading model. However, we have a less restrictive constraint on the input and we rigorously prove that for all rates between 0 and capacity, (18) gives us the true reliability function. Further, we extend the result to a block-fading channel with multiple antennas.

Define

$$f(x) = \frac{1}{x} \ln \frac{1 + \frac{\rho}{1+\rho}x}{(1+x)^{\frac{\rho}{1+\rho}}}. \quad (19)$$

The properties of $f(x)$ were well studied in [7]. We just mention the key properties of $f(x)$ here:

1. $f(x) \geq 0, \forall x \geq 0$. When x goes to 0 or ∞ , $f(x)$ goes to 0, i.e., $f(0) = f(\infty) = 0$.
2. For a fixed $\rho > 0$, there is always a unique optimizing x , which we denote by x_ρ , i.e.,

$$x_\rho = \arg \max_{x>0} \frac{1}{x} \ln \frac{1 + \frac{\rho}{1+\rho}x}{(1+x)^{\frac{\rho}{1+\rho}}}. \quad (20)$$

3. As ρ goes to zero, x_ρ monotonically increases to infinity.

The low-SNR reliability function for a SISO fast-fading channel (18) can also be written as

$$E^{lp}(r) = \sup_{x>0} \sup_{0 \leq \rho \leq 1} -\rho r + (1+\rho) \frac{1}{x} \ln \frac{1 + \frac{\rho}{1+\rho}x}{(1+x)^{\frac{\rho}{1+\rho}}}. \quad (21)$$

Let $\bar{\tau}(r)$ be the optimizing x in (21), i.e.,

$$\bar{\tau}(r) = \arg \max_{x>0} \sup_{0 \leq \rho \leq 1} -\rho r + (1+\rho) \frac{1}{x} \ln \frac{1 + \frac{\rho}{1+\rho}x}{(1+x)^{\frac{\rho}{1+\rho}}}. \quad (22)$$

$\bar{\tau}(r)$ establishes a direct relation between rate and how large the non-zero symbol has to be for an on-off signaling scheme to achieve the optimal reliability function at rate r . As we will show in the next section, for any signaling scheme, including on-off signaling, it is a necessary condition that a symbol with energy of $\bar{\tau}(r)$ be used for a signaling scheme to be able to achieve the optimal channel performance at rate r . Thus, $\bar{\tau}(r)$ indicates the *minimum signal peakiness* for the SISO channel at rate r .

For the more general MIMO block-fading model, our result here tells us

1. Having a longer channel coherence time T_c does not improve the decoding error probability. This is a bit surprising since one would expect that making T_c large may give us more time to estimate the channel and thus help improve the non-coherent channel performance. However, such a conclusion may hold for *training*-based schemes, where a pilot or training symbol is sent to estimate the channel before data transmission. Such schemes are themselves sub-optimal in the non-coherent low-SNR communications. We will show that the optimal signaling scheme in this regime to achieve the optimal decoding error probability belongs to a set of peaky signaling schemes which we call a *generalized flash-signaling* scheme. For such signaling schemes, having longer coherence time will not help with the channel performance.
2. Having more than one input antenna does not improve the channel performance. This is natural for this independently fading model since using more than one antenna forces us to spread the little power into even smaller slices to each antenna. However, as we will see in the proof of Theorem 3, it is not *detrimental* either to spread the power across multiple antennas. In the next section, we will show that having more transmit antennas reduces the *peakiness* needed for the signaling.

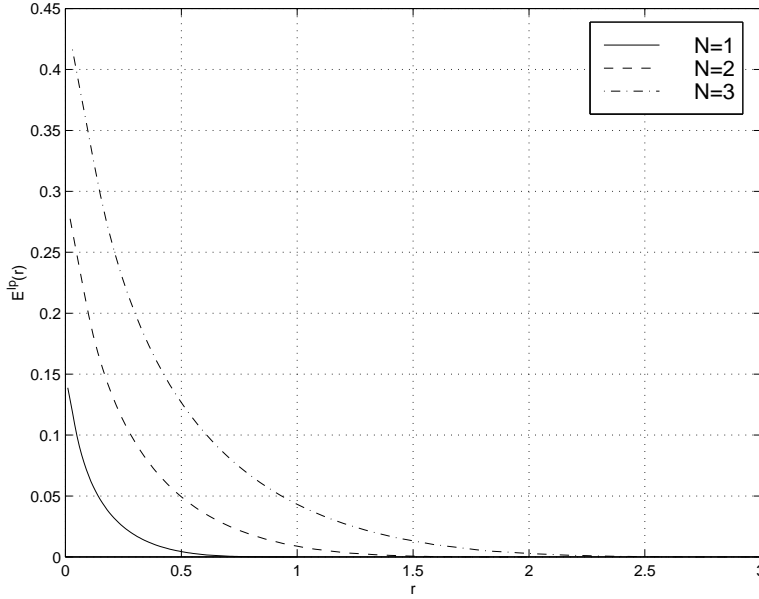


Figure 1: The low-SNR reliability functions for MIMO fading channels

3. The performance gain, as defined in Definition 1, for a MIMO fading channel is N , which is the number of receive antennas.

To get a better feel for the performance gain due to multiple antennas, we look at a numerical example. Figure 1 shows the low-SNR reliability function for MIMO channels with $N = 1$, $N = 2$ and $N = 3$. From this figure, we can see having multiple receive antennas can improve both capacity and reliability function by a factor of N .

Proof of Theorem 3: We first compute $\tilde{E}_o(\rho)$ using (13). The integral inside (13) can be computed as follows

$$\begin{aligned}
& \int f(\mathbf{Y}|\mathbf{0})^{\frac{\rho}{1+\rho}} f(\mathbf{Y}|\mathbf{X})^{\frac{1}{1+\rho}} d\mathbf{Y} \\
&= \int \left(\prod_{i=1}^N \frac{1}{\pi^{T_c}} e^{-\|y_i\|^2} \right)^{\frac{\rho}{1+\rho}} \left(\prod_{i=1}^N \frac{1}{\pi^{T_c} \det(\mathbf{I} + \mathbf{X}^\dagger \mathbf{X})} e^{-\mathbf{y}_i^\dagger (\mathbf{I} + \mathbf{X}^\dagger \mathbf{X})^{-1} \mathbf{y}_i} \right)^{\frac{1}{1+\rho}} d\mathbf{Y} \\
&= \prod_{i=1}^N \int \frac{1}{\pi^{T_c}} \frac{1}{\det(\mathbf{I} + \mathbf{X}^\dagger \mathbf{X})^{\frac{1}{1+\rho}}} e^{-\mathbf{y}_i^\dagger \left[\frac{\rho}{1+\rho} \mathbf{I} + \frac{1}{1+\rho} (\mathbf{I} + \mathbf{X}^\dagger \mathbf{X})^{-1} \right] \mathbf{y}_i} d\mathbf{Y} \\
&= \left[\frac{1}{\det(\mathbf{I} + \mathbf{X}^\dagger \mathbf{X})^{\frac{1}{1+\rho}} \det\left(\frac{\rho}{1+\rho} \mathbf{I} + \frac{1}{1+\rho} (\mathbf{I} + \mathbf{X}^\dagger \mathbf{X})^{-1}\right)} \right]^N \\
&= \left[\frac{\det(\mathbf{I} + \mathbf{X}^\dagger \mathbf{X})^{\frac{\rho}{1+\rho}}}{\det\left(\mathbf{I} + \frac{\rho}{1+\rho} \mathbf{X}^\dagger \mathbf{X}\right)} \right]^N.
\end{aligned}$$

Let $\sigma(\mathbf{A})$ denote the set of eigenvalues of \mathbf{A} . Assuming that the eigenvalues for the matrix $\mathbf{X}\mathbf{X}^\dagger$ are $\{\lambda_1, \lambda_2, \dots, \lambda_{T_c}\}$,

i.e.,

$$\sigma(\mathbf{X}\mathbf{X}^\dagger) = \{\lambda_1, \lambda_2, \dots, \lambda_{T_c}\},$$

we have

$$\begin{aligned} \tilde{E}_o(\rho) &= (1 + \rho) \sup_{\mathbf{X} \neq \mathbf{0}} \frac{1}{\|\mathbf{X}\|^2} \ln \left(\frac{\det(\mathbf{I} + \frac{\rho}{1+\rho} \mathbf{X}^\dagger \mathbf{X})}{\det(\mathbf{I} + \mathbf{X}^\dagger \mathbf{X})^{\frac{\rho}{1+\rho}}} \right)^N \\ &= N(1 + \rho) \sup_{\mathbf{X} \neq \mathbf{0}} \frac{1}{\sum_{t=1}^{T_c} \lambda_t} \ln \frac{\prod_{t=1}^{T_c} (1 + \frac{\rho}{1+\rho} \lambda_t)}{\prod_{t=1}^{T_c} (1 + \lambda_t)^{\frac{\rho}{1+\rho}}} \\ &= N(1 + \rho) \sup_{\mathbf{X} \neq \mathbf{0}} \frac{\sum_{t=1}^{T_c} \ln \frac{(1 + \frac{\rho}{1+\rho} \lambda_t)}{(1 + \lambda_t)^{\frac{\rho}{1+\rho}}}}{\sum_{t=1}^{T_c} \lambda_t} \\ &\leq N(1 + \rho) \sup_{\mathbf{X} \neq \mathbf{0}} \sup_t \frac{\ln \frac{(1 + \frac{\rho}{1+\rho} \lambda_t)}{(1 + \lambda_t)^{\frac{\rho}{1+\rho}}}}{\lambda_t} \end{aligned} \quad (23)$$

$$\leq N(1 + \rho) \sup_{\mathbf{X} \neq \mathbf{0}} \sup_t \sup_{x > 0} \frac{1}{x} \ln \frac{1 + \frac{\rho}{1+\rho} x}{(1 + x)^{\frac{\rho}{1+\rho}}} \quad (24)$$

$$= N(1 + \rho) \sup_{x > 0} \frac{1}{x} \ln \frac{1 + \frac{\rho}{1+\rho} x}{(1 + x)^{\frac{\rho}{1+\rho}}}. \quad (25)$$

To get (23), we use the inequality

$$\frac{\sum_l a_l}{\sum_l b_l} \leq \max_l \frac{a_l}{b_l}, \quad \text{if } a_l, b_l > 0, \text{ for all } l.$$

The inequalities in (23) and (24) can be made equalities by carefully choosing the input matrix \mathbf{X} . It is easy to see that a sufficient condition for this to happen is to choose \mathbf{X} such that the eigenvalues of $\mathbf{X}\mathbf{X}^\dagger$ are either zero or x_ρ , where x_ρ is defined as (20). A simple example of such an input matrix is $\mathbf{X} = \sqrt{x_\rho} \mathbf{I}_{M \times T_c}$. Thus, we have (17).

From Corollary 1, we know that to show that (16) is valid for all rates, we have to check that (15) is true for this channel model. From (14), it is easy to verify

$$E_{sl}^{lp}(0) = N \sup_q \frac{\mathcal{E} \left[\ln \frac{\det(\mathbf{I} + \frac{1}{2} \mathbf{X}_1^\dagger \mathbf{X}_1 + \frac{1}{2} \mathbf{X}_2^\dagger \mathbf{X}_2)}{\det(\mathbf{I} + \mathbf{X}_1^\dagger \mathbf{X}_1)^{\frac{1}{2}} \det(\mathbf{I} + \mathbf{X}_2^\dagger \mathbf{X}_2)^{\frac{1}{2}}} \right]}{\mathcal{E}[\|\mathbf{X}\|^2]}. \quad (26)$$

To get an upper bound for $E_{sl}^{lp}(0)$, we first state the following lemma.

Lemma 1 Assume $\mathbf{A}, \mathbf{B} \in \mathcal{C}^{M \times M}$ are Hermitian and semi positive definite matrices, i.e., $\mathbf{A}, \mathbf{B} \geq 0$, we have

$$\det(\mathbf{I} + \mathbf{A} + \mathbf{B}) \leq \det(\mathbf{I} + \mathbf{A}) \det(\mathbf{I} + \mathbf{B}). \quad (27)$$

Proof: See Appendix B. ◇

It is easy to see that $\frac{1}{2}\mathbf{X}_1^\dagger\mathbf{X}_1 \geq 0$ and $\frac{1}{2}\mathbf{X}_2^\dagger\mathbf{X}_2 \geq 0$. Applying the above lemma to (26), we have

$$\begin{aligned}
E_{sl}^{lp}(0) &\leq N \sup_q \frac{\mathcal{E} \left[\ln \frac{\det(\mathbf{I} + \frac{1}{2}\mathbf{X}_1^\dagger\mathbf{X}_1) \det(\mathbf{I} + \frac{1}{2}\mathbf{X}_2^\dagger\mathbf{X}_2)}{\det(\mathbf{I} + \mathbf{X}_1^\dagger\mathbf{X}_1)^{\frac{1}{2}} \det(\mathbf{I} + \mathbf{X}_2^\dagger\mathbf{X}_2)^{\frac{1}{2}}} \right]}{\mathcal{E}[\|\mathbf{X}\|^2]} \\
&= N \sup_q \frac{\mathcal{E} \left[\ln \frac{\det(\mathbf{I} + \frac{1}{2}\mathbf{X}_1^\dagger\mathbf{X}_1)}{\det(\mathbf{I} + \mathbf{X}_1^\dagger\mathbf{X}_1)^{\frac{1}{2}}} \right] + \mathcal{E} \left[\ln \frac{\det(\mathbf{I} + \frac{1}{2}\mathbf{X}_2^\dagger\mathbf{X}_2)}{\det(\mathbf{I} + \mathbf{X}_2^\dagger\mathbf{X}_2)^{\frac{1}{2}}} \right]}{\mathcal{E}[\|\mathbf{X}\|^2]} \\
&= 2N \sup_q \frac{\mathcal{E} \left[\ln \frac{\det(\mathbf{I} + \frac{1}{2}\mathbf{X}^\dagger\mathbf{X})}{\det(\mathbf{I} + \mathbf{X}^\dagger\mathbf{X})^{\frac{1}{2}}} \right]}{\mathcal{E}[\|\mathbf{X}\|^2]} \\
&= 2N \sup_q \frac{\sum_{l=1}^{T_c} \mathcal{E} \left[\ln \frac{(1 + \frac{1}{2}\lambda_l)}{(1 + \lambda_l)^{\frac{1}{2}}} \right]}{\sum_{l=1}^{T_c} \mathcal{E}[\lambda_l]} \\
&\leq 2N \sup_q \frac{\sum_{l=1}^{T_c} \mathcal{E} \left[\lambda_l \sup_{x>0} \frac{1}{x} \ln \frac{(1 + \frac{1}{2}x)}{(1+x)^{\frac{1}{2}}} \right]}{\sum_{l=1}^{T_c} \mathcal{E}[\lambda_l]} \\
&= 2N \sup_{x>0} \frac{1}{x} \ln \frac{(1 + \frac{1}{2}x)}{(1+x)^{\frac{1}{2}}} \\
&= \tilde{E}_o(1),
\end{aligned}$$

which completes the proof of this theorem. ◇

3.2 Optimal signaling schemes in the low-SNR regime

In this section we study the conditions for an input distribution to be optimal in the low-SNR regime.

Definition 5 A sequence of input distributions $\{q_p(\mathbf{X}) \in \mathcal{D}(p)\}$ is called *first-order optimal* with respect to a communication rate r if it satisfies

$$\lim_{p \rightarrow 0} \frac{E(rp, q_p, p)}{p} = E^{lp}(r), \tag{28}$$

where $E(R, q_p, p)$ is the reliability function of the channel when the input distribution is chosen to be q_p . ◇

In the above definition, to be consistent with prior literature, we use the term *first-order* optimal to indicate optimality in the limit $SNR \rightarrow 0$ [22, 24].

From the proof of Theorem 3, it immediately follows that the on-off signaling scheme given below is first-order optimal:

Corollary 2 *An on-off signaling scheme*

$$\mathbf{X} = \begin{cases} 0 & w.p. \ 1 - \frac{pT_c}{\|\mathbf{X}_o\|^2}; \\ \mathbf{X}_o & w.p. \ \frac{pT_c}{\|\mathbf{X}_o\|^2}, \end{cases} \quad (29)$$

is first-order optimal with respect to a transmission rate r if the non-zero eigenvalues of $\mathbf{X}_o\mathbf{X}_o^\dagger$ are all equal to $\kappa(r)$, where $\kappa(r)$ is defined as the following

$$\kappa(r) = \bar{\tau}\left(\frac{r}{N}\right) = \arg \max_{x>0} \sup_{0 \leq \rho \leq 1} -\rho \frac{r}{N} + (1 + \rho) \frac{1}{x} \ln \frac{(1 + \frac{\rho}{1+\rho}x)}{(1+x)^{\frac{\rho}{1+\rho}}}. \quad (30)$$

◇

For the SISO fast-fading model where $M = N = T_c = 1$, a first-order optimal signaling scheme is simply to choose the input symbol to be 0 or $\kappa(r)^{\frac{1}{2}}$. However, the peak-to-average ratio of the on-off signaling scheme goes to infinity as the mean power goes to 0. A natural question to ask here is the following: can we find less “peaky” signaling schemes which are also first-order optimal, by considering other signaling schemes other than on-off signaling? It turns out that only signaling schemes in the following category *can* be first-order optimal.

Definition 6 *A sequence of input distributions $\{q_p\}$ defined on the input matrix $\mathbf{X} \in C^{M \times T_c}$ is said to be a general flash signaling scheme with peak constraint K if it satisfies the average power constraint*

$$\mathcal{E}[\|\mathbf{X}\|^2] = pT_c,$$

and for any $\epsilon > 0$ and any $l = 1, 2, \dots, T_c$,

$$\liminf_{p \rightarrow 0} \frac{\mathcal{E} \left[\lambda_l I_{\{|\lambda_l - K| \geq \epsilon\}} \right]}{p} = 0, \quad (31)$$

where $\{\lambda_1, \lambda_2, \dots, \lambda_l\}$ are the eigenvalues of the input matrix $\mathbf{X}^\dagger \mathbf{X}$ and I_A denotes the indicator function of the event A . ◇

Remarks: This condition (31) can be interpreted as follows: *almost all probability mass is assigned to those input matrices \mathbf{X} such that the eigenvalues of $\mathbf{X}^\dagger \mathbf{X}$ are either 0, or in an arbitrarily small neighborhood of a constant K .* Now consider the case when $T_c = 1$, the input matrix \mathbf{X} is simply a vector and therefore the only eigenvalue of $\mathbf{X}^\dagger \mathbf{X}$ is $\|\mathbf{X}\|^2$. When $K = \kappa(r)$, (31) becomes

$$\liminf_{p \rightarrow 0} \frac{\mathcal{E} \left[\|\mathbf{X}\|^2 I_{\{|\|\mathbf{X}\|^2 - \kappa(r)| \geq \epsilon\}} \right]}{p} = 0. \quad (32)$$

As r approaches capacity, $\kappa(r)$ goes to infinity. Then this definition is equivalent to the definition of the *flash signaling* in [22]. Thus, the signaling scheme defined in (31) is essentially a generalized flash signaling scheme. The next lemma shows that a general flash signaling scheme with peak constraint $\kappa(r)$ is necessary to achieve the best reliability function at rate r .

Lemma 2 *A necessary condition for a sequence of input distributions $\{q_p \in \mathcal{D}(p)\}$ to be first-order optimal with respect to a coding rate r is that $\{q_p\}$ is a general flash-signaling scheme with peak constraint $\kappa(r)$, where $\kappa(r)$ is defined by (30).*

Proof: We consider two cases: $0 \leq r < r_{crit}$ and $r \geq r_{crit}$, where r_{crit} corresponds to the maximum rate r such that the optimizing ρ at rate r in (16) is 1. First we look at the case where $r \geq r_{crit}$ and show that the lemma holds.

To see this, we need to go back to the sphere-packing proof of Theorem 2. From (71), we know that for any sequence of input distributions q_p ,

$$\begin{aligned}
& \lim_{p \rightarrow 0} \frac{E(rp, q_p, p)}{p} \\
\leq & \limsup_{p \rightarrow 0} \sup_{\rho \geq 0} -\rho r + \frac{-(1+\rho)\mathcal{E} \left[\ln \int f(\mathbf{Y}|\mathbf{0})^{\frac{\rho}{1+\rho}} f(\mathbf{Y}|\mathbf{X})^{\frac{1}{1+\rho}} d\mathbf{Y} \right]}{pT_c} \\
= & \limsup_{p \rightarrow 0} \sup_{\rho \geq 0} -\rho r + N(1+\rho) \frac{\mathcal{E} \left[\ln \frac{\det(\mathbf{I} + \frac{\rho}{1+\rho} \mathbf{X}^\dagger \mathbf{X})}{\det(\mathbf{I} + \mathbf{X}^\dagger \mathbf{X})^{\frac{\rho}{1+\rho}}} \right]}{pT_c} \\
\leq & \sup_{\rho \geq 0} -\rho r + N(1+\rho) \limsup_{p \rightarrow 0} \frac{\mathcal{E} \left[\ln \frac{\det(\mathbf{I} + \frac{\rho}{1+\rho} \mathbf{X}^\dagger \mathbf{X})}{\det(\mathbf{I} + \mathbf{X}^\dagger \mathbf{X})^{\frac{\rho}{1+\rho}}} \right]}{pT_c} \\
\leq & \sup_{\rho \geq 0} -\rho r + N(1+\rho) \sup_{x>0} \frac{1}{x} \ln \frac{(1 + \frac{\rho}{1+\rho}x)}{(1+x)^{\frac{\rho}{1+\rho}}} \tag{33} \\
= & \sup_{0 \leq \rho \leq 1} -\rho r + N(1+\rho) \sup_{x>0} \frac{1}{x} \ln \frac{(1 + \frac{\rho}{1+\rho}x)}{(1+x)^{\frac{\rho}{1+\rho}}} \tag{34} \\
= & E^{lp}(r)
\end{aligned}$$

Equation (34) holds here because we assume that $r \geq r_{crit}$, which indicates the optimizing ρ in (33) is between 0 and 1.

For $\{q_p\}$ to be first-order optimal, all three inequalities above should hold as equalities. For (33) to hold as an equality, we need

$$\limsup_{p \rightarrow 0} \frac{\mathcal{E} \left[\ln \frac{\det(\mathbf{I} + \frac{\rho^*}{1+\rho^*} \mathbf{X}^\dagger \mathbf{X})}{\det(\mathbf{I} + \mathbf{X}^\dagger \mathbf{X}) \frac{\rho^*}{1+\rho^*}} \right]}{pT_c} = \sup_{x>0} \frac{1}{x} \ln \frac{(1 + \frac{\rho^*}{1+\rho^*}x)}{(1+x) \frac{\rho^*}{1+\rho^*}}, \quad (35)$$

where $\rho^* \in [0, 1]$ is the optimizing ρ for (33). We know for any rate $r \in [0, N)$, $\rho^* > 0$. Further, $\kappa(r)$ is the optimizing x for (35). Given $\epsilon > 0$, let $\xi(\epsilon)$ be such that

$$\sup_{x>0} \frac{1}{x} \ln \frac{1 + \frac{\rho^*}{1+\rho^*}x}{(1+x) \frac{\rho^*}{1+\rho^*}} - \xi(\epsilon) \geq \sup_{|x-\kappa(r)|>\epsilon} \frac{1}{x} \ln \frac{1 + \frac{\rho^*}{1+\rho^*}x}{(1+x) \frac{\rho^*}{1+\rho^*}}.$$

Note that such a $\xi(\epsilon)$ exists since $\kappa(r) = x_{\rho^*}$ maximizes $f(x)$. Now, the left side of (35) (without the limit) can be further written as

$$\begin{aligned} & \frac{\mathcal{E} \left[\ln \frac{\det(\mathbf{I} + \frac{\rho^*}{1+\rho^*} \mathbf{X}^\dagger \mathbf{X})}{\det(\mathbf{I} + \mathbf{X}^\dagger \mathbf{X}) \frac{\rho^*}{1+\rho^*}} \right]}{pT_c} \\ &= \frac{\sum_{l=1}^{T_c} \mathcal{E} \left[\ln \frac{(1 + \frac{\rho^*}{1+\rho^*} \lambda_l)}{(1+\lambda_l) \frac{\rho^*}{1+\rho^*}} \right]}{pT_c} \\ &= \frac{\sum_{l=1}^{T_c} \mathcal{E} \left[\ln \frac{(1 + \frac{\rho^*}{1+\rho^*} \lambda_l)}{(1+\lambda_l) \frac{\rho^*}{1+\rho^*}} I_{\{|\lambda_l - \kappa(r)| < \epsilon\}} \right] + \mathcal{E} \left[\ln \frac{(1 + \frac{\rho^*}{1+\rho^*} \lambda_l)}{(1+\lambda_l) \frac{\rho^*}{1+\rho^*}} I_{\{|\lambda_l - \kappa(r)| \geq \epsilon\}} \right]}{pT_c} \\ &\leq \frac{\sum_{l=1}^{T_c} \mathcal{E} \left[\frac{1}{\kappa(r)} \ln \frac{(1 + \frac{\rho^*}{1+\rho^*} \kappa(r))}{(1+\kappa(r)) \frac{\rho^*}{1+\rho^*}} \lambda_l I_{\{|\lambda_l - \kappa(r)| < \epsilon\}} \right] + \mathcal{E} \left[\left(\frac{1}{\kappa(r)} \ln \frac{(1 + \frac{\rho^*}{1+\rho^*} \kappa(r))}{(1+\kappa(r)) \frac{\rho^*}{1+\rho^*}} - \xi \right) \lambda_l I_{\{|\lambda_l - \kappa(r)| \geq \epsilon\}} \right]}{pT_c} \quad (36) \\ &= \frac{\sum_{l=1}^{T_c} \frac{1}{\kappa(r)} \ln \frac{(1 + \frac{\rho^*}{1+\rho^*} \kappa(r))}{(1+\kappa(r)) \frac{\rho^*}{1+\rho^*}} \mathcal{E} [\lambda_l] - \xi \mathcal{E} [\lambda_l I_{\{|\lambda_l - \kappa(r)| \geq \epsilon\}}]}{pT_c} \\ &= \sup_{x>0} \frac{1}{x} \ln \frac{(1 + \frac{\rho^*}{1+\rho^*}x)}{(1+x) \frac{\rho^*}{1+\rho^*}} - \xi \sum_{l=1}^{T_c} \frac{\mathcal{E} [\lambda_l I_{\{|\lambda_l - \kappa(r)| \geq \epsilon\}}]}{pT_c} \quad (37) \end{aligned}$$

The inequality in (36) follows from the definition of ξ .

By taking lim sup of the expression in (37) and comparing it with (35), we have

$$\sum_{l=1}^{T_c} \liminf_{p \rightarrow 0} \frac{\mathcal{E} [\lambda_l I_{\{|\lambda_l - \kappa(r)| \geq \epsilon\}}]}{p} = 0.$$

Since $\lambda_l \geq 0$, (31) follows and we know $\{q_p\}$ must be a flash signaling scheme with peak constraint $\kappa(r)$.

Next we look at the other case where $r \in [0, r_{crit})$. We claim that a necessary condition for a sequence of input distributions to be first-order optimal at any rate $r < r_{crit}$ is that

$$E_{sp}^{lp}(r_{crit}, q_p) = E_{sp}^{lp}(r_{crit}). \quad (38)$$

To see this, we bound the straight-line exponent of $E(rp, q_p, p)$, where the input distribution is chosen to be q_p , in a similar way as in the proof of the straight-line part of Theorem 2:

$$\begin{aligned} E^{lp}(r, q_p) &= \lim_{p \rightarrow 0} \frac{E(rp, q_p, p)}{p} \\ &\leq \lim_{p \rightarrow 0} \frac{E_{sl}(rp, q_p, p)}{p} \\ &= \lim_{p \rightarrow 0} \frac{\inf_{r' \geq r} (1 - \frac{r}{r'}) E(0, q_p, p) + \frac{r}{r'} E_{sp}(r'p, q_p, p)}{p} \end{aligned} \quad (39)$$

$$\leq \lim_{p \rightarrow 0} \frac{(1 - \frac{r}{r_{crit}}) E(0, p) + \frac{r}{r_{crit}} E_{sp}(r_{crit}p, q_p, p)}{p} \quad (40)$$

$$\leq (1 - \frac{r}{r_{crit}}) E_{sl}^{lp}(0) + \frac{r}{r_{crit}} E_{sp}^{lp}(r_{crit}, q_p) \quad (41)$$

$$\leq (1 - \frac{r}{r_{crit}}) E_{sl}^{lp}(0) + \frac{r}{r_{crit}} E_{sp}^{lp}(r_{crit}) \quad (42)$$

$$\leq (1 - \frac{r}{r_{crit}}) E_r^{lp}(0) + \frac{r}{r_{crit}} E_r^{lp}(r_{crit}) \quad (43)$$

$$= E_r^{lp}(r) = E^{lp}(r). \quad (44)$$

The first three steps in the sequence of equations above are straightforward. For the inequality (40), since we are considering the case when $r < r_{crit}$, we simply choose $r' = r_{crit}$ which will yield an upper bound for (39). We also bound $E(0, q_p, p)$ by $E(0, p)$, which is obviously true. For the inequality (41), we apply the bound $E^{lp}(0) = \lim_{p \rightarrow 0} \frac{E(0, p)}{p} \leq E_{sl}^{lp}(0)$ from Theorem 2. The inequality (42) is trivial since $E_{sp}^{lp}(r_{crit})$ is equal to $E_{sp}^{lp}(r_{crit}, q_p)$ optimized over all choices of q_p . We have proved that for the MIMO block fading channel in the proof of Theorem 3, $E_{sl}^{lp}(0) \leq E_r^{lp}(0)$. Thus, (43) must be true. (44) is trivial since (43) is simply the straight-line part in the random-coding bound.

For $\{q_p\}$ to be first-order optimal for $r < r_{crit}$, we need all the inequalities in the above calculations to hold as equalities. As a necessary condition, we look at the condition for (42) to hold as an equality, which is equivalent to requiring that (38) must hold. Applying the result in the first part of the proof, a necessary condition for (38) is that $\{q_p\}$ is a general flash signaling scheme with peak constraint $\kappa(r_{crit})$. In other words, a necessary condition for $\{q_p\}$ to be first-order optimal for rate $r < r_{crit}$ is that $\{q_p\}$ is a general flash signaling scheme with peak constraint $\kappa(r_{crit})$. However, it is easy to check that $\kappa(r) = \kappa(r_{crit})$ for $r \leq r_{crit}$. Thus, we complete the proof of this lemma. \diamond

The necessary condition (31) gives us a constraint on the eigenvalues of $\mathbf{X}^\dagger \mathbf{X}$ for a signaling scheme to be first-order optimal. However, usually it is not that convenient to look at the eigenvalues and try to make these eigenvalues satisfy certain peakiness conditions. For this purpose, we define the following quantity, which indicates the level of peakiness we need at the entries of the input matrix \mathbf{X} , rather than the eigenvalues of $\mathbf{X}^\dagger \mathbf{X}$, to achieve the optimal decoding error performance for a certain rate r . Let $\mathcal{F}_1(r)$ be the set of sequences of first-order optimal distributions with respect to rate r .

Definition 7 We define $\tau(r)$ to be the minimum signal peakiness for a communication rate r on the MIMO block-fading channel (1):

$$\tau(r) = \inf_{\{q_p\} \in \mathcal{F}_1(r)} \limsup_{p \rightarrow 0} \sup_{\mathbf{X} \in \mathcal{A}_p} \|\mathbf{X}\|_\infty^2, \quad (45)$$

where \mathcal{A}_p is the alphabet for the input distribution q_p and $\|\mathbf{X}\|_\infty$ is the l_∞ norm of matrix \mathbf{X} . \diamond

From a practical point of view, it is unreasonable to allow input signals with arbitrarily large peakiness. Thus, $\tau(r)$ is an important parameter which shows how difficult in practice it is to achieve the best performance at rate r . There is a rather simple relation between $\tau(r)$ and $\kappa(r)$, which is presented in the following theorem.

Theorem 4 For a MIMO block-fading channel (1), the minimum signal peakiness to achieve the optimum decoding error probability at coding rate r is

$$\tau(r) = \frac{\kappa(r)}{MT_c} = \frac{\bar{\tau}(\frac{r}{N})}{MT_c}, \quad (46)$$

where $\kappa(r)$ is defined by (30) and $\bar{\tau}(r)$ is defined by (22). Thus, the peakiness gain achieved by the MIMO block-fading channel is MT_c .

Remarks: This result is actually quite straightforward if we take a close look at Definition 6 and Lemma 2. It can be seen there that the only condition required for a signaling scheme to be first-order optimal is that for all input matrices \mathbf{X} with a non-trivial probability, the eigenvalues of $\mathbf{X}^\dagger \mathbf{X}$ have to be either 0 or some large number, determined by $\kappa(r)$. In other words, there is a peakiness requirement for a signaling scheme to be first-order optimal. However, we have the freedom to construct the input matrix to satisfy this condition and minimize the peakiness requirement in each entry of the matrix. When we have multiple transmit antennas and coherence time T_c , one obvious way to do this is to transmit over only one of the eigenmode, i.e., make all the entries identical and equal to $\frac{\kappa(r)}{MT_c}$, which gives us a peakiness of $\frac{\kappa(r)}{MT_c}$. Now we show rigorously that this is actually the best peakiness level we can obtain to achieve the optimal channel performance at communication rate r .

Proof: We first show $\tau(r) \geq \frac{\kappa(r)}{MT_c}$. Suppose $\tau(r) < \frac{\kappa(r)}{MT_c}$, we can find $\xi > 0$, such that

$$\tau(r) \leq \frac{\kappa(r)}{MT_c} - \xi.$$

Thus, we know for some sequence of first-order input distributions $\{q_p\}$, we have for any input alphabet (matrix) \mathbf{X} ,

$$\limsup_{p \rightarrow 0} (\|\mathbf{X}\|_\infty)^2 \leq \frac{\kappa(r)}{MT_c} - \xi/2.$$

Thus, we must have for any l ,

$$\limsup_{p \rightarrow 0} \lambda_l \leq \limsup_{p \rightarrow 0} \sum_{l=1}^{T_c} \lambda_l = \limsup_{p \rightarrow 0} \text{Tr}(\mathbf{X}^\dagger \mathbf{X}) \leq \limsup_{p \rightarrow 0} MT_c (\|\mathbf{X}\|_\infty)^2 \leq \kappa(r) - \frac{MT_c \xi}{2}. \quad (47)$$

Choosing $\xi = \frac{2\epsilon}{MT_c}$, we have,

$$\limsup_{p \rightarrow 0} \frac{\mathcal{E}[\lambda_l I_{\{|\lambda_l - \kappa(r)| < \epsilon\}}]}{p} \leq \limsup_{p \rightarrow 0} \max_{\mathbf{X} \in \mathcal{A}_p \setminus \{0\}} \frac{\lambda_l I_{\{|\lambda_l - \kappa(r)| < \epsilon\}}}{\|\mathbf{X}\|^2}$$

From (47), we know when p is sufficiently small, $I_{\{|\lambda_l - \kappa(r)| < \epsilon\}} = 0$. Hence,

$$\lim_{p \rightarrow 0} \frac{\mathcal{E}[\lambda_l I_{\{|\lambda_l - \kappa(r)| < \epsilon\}}]}{p} = 0$$

and further

$$\sum_{l=1}^{T_c} \lim_{p \rightarrow 0} \frac{\mathcal{E}[\lambda_l I_{\{|\lambda_l - \kappa(r)| < \epsilon\}}]}{p} = 0.$$

Thus,

$$\sum_{l=1}^{T_c} \liminf_{p \rightarrow 0} \frac{\mathcal{E}[\lambda_l I_{\{|\lambda_l - \kappa(r)| \geq \epsilon\}}]}{p} = \sum_{l=1}^{T_c} \liminf_{p \rightarrow 0} \frac{\mathcal{E}[\lambda_l] - \mathcal{E}[\lambda_l I_{\{|\lambda_l - \kappa(r)| < \epsilon\}}]}{p} = T_c > 0,$$

which violates the necessary condition (31) for first-order optimality and thus contradicts the assumption that $\{q_p\}$ is first-order optimal. Hence, we must have $\tau(r) \geq \frac{\kappa(r)}{MT_c}$.

For the other direction, what we need to show is that there exists a sequence of first-order optimal input distributions $\{q_p\}$ such that the signal peakiness is $\frac{\kappa(r)}{MT_c}$.

Consider the following on-off signaling scheme in the form of (29): for any p , we choose each entry of \mathbf{X}_o to be identical and equal to $\frac{\kappa(r)}{MT_c}$. It is easy to check that for this input matrix, only one non-zero eigenvalue exists and equal to $\kappa(r)$. Thus the first-order sufficient condition in Corollary 2 is satisfied and $\{q_p\}$ is first-order optimal. Also it is trivial to check that the signal peakiness for this signaling scheme is $\frac{\kappa(r)}{MT_c}$. \diamond

Remarks: When $M = N = T_c = 1$, it is easy to see that $\tau(r) = \bar{\tau}(r)$. Thus, $\bar{\tau}(r)$ tells us how peaky the signaling has to be to achieve a certain point on the reliability curve for the SISO fast-fading channel. $\bar{\tau}(r)$ can

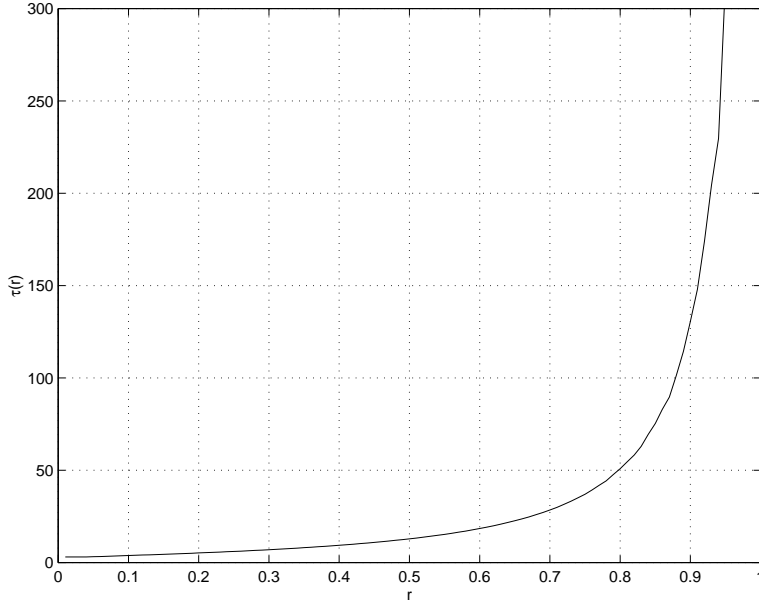


Figure 2: The required peakiness at communication rate r

be computed numerically and is shown in Figure 2. From Figure 2, we see that as we get closer and closer to capacity, we have to use more and more peaky signaling schemes.

This theorem also tells us that the role of having more transmit antennas and longer coherence time in the low-SNR regime is to reduce the minimum signal peakiness we need to achieve the optimal reliability function. It is straightforward to see that having larger coherence time can help us to reduce the signaling peakiness. However, it is less intuitive that having multiple transmitting antennas is also helpful in terms of reducing the peakiness we need to achieve the optimal performance. To give the intuition behind this phenomena, we give a simple example here: consider a fading channel with $M = 2$ transmit antennas, 1 receive antenna and coherence time $T_c = 1$ and compare it with the SISO fast-fading channel ($M = N = T_c = 1$). We claim that for any channel performance (a certain communication rate and a certain decoding error probability) achieved by the SISO channel, we can achieve the same performance using the 2 transmit antenna channel and only half the signal peakiness. To see this, we just use the same input x for the two transmit antennas, and the channel is

$$y = (h_1 + h_2)x + w,$$

with an average power constraint $\mathcal{E}[|x|^2] = \frac{P}{2}$. Since both h_1 and h_2 are Gaussian with variance 1, $h_1 + h_2$ is also complex Gaussian with variance 2. Hence our channel is also equivalent to the following SISO channel

$$y = \tilde{h}\tilde{x} + w,$$

where $\tilde{h} = \frac{h_1+h_2}{\sqrt{2}}$ is Gaussian with variance 1 and $\tilde{x} = \sqrt{2}x$ has average power p . Hence, we can always get the same performance as a SISO channel by using the same signal at both antennas while reducing the input peakiness by a factor of 2.

3.3 Fading channels with average and peak power constraint

As we have shown in the last section, to achieve the optimal reliability function, we need peaky signaling schemes, in the sense that although the average power goes to 0, we have to keep at least one of the symbols at a certain energy level, which is not going to 0. As r approaches the capacity, the energy for this symbol actually goes to infinity, which converges to an extremely peaky signaling scheme: *flash signaling* [22]. However, in practice, it is not possible for us to use input symbols with infinite energy. In this section, we will study the reliability function of a fading channel (1) with both average and peak power constraint, where the average power constraint goes to 0 and peak power constraint remains at a constant level. Specifically, the peak power constraint is defined as for any input matrix \mathbf{X} , we must have

$$\|\mathbf{X}\|_\infty^2 \leq K_m. \quad (48)$$

Our main result in this section is as follows:

Theorem 5 *For the MIMO model (1) with an additional peak power constraint (48), we have*

$$\overline{E}^{lp}(r) = \sup_{0 \leq \rho \leq 1} -\rho r + \tilde{E}_o(\rho), \quad (49)$$

where $\tilde{E}_o(\rho)$ is defined as follows

$$\tilde{E}_o(\rho) = N(1 + \rho) \sup_{0 < x < MT_c K_m} \frac{1}{x} \ln \frac{1 + \frac{\rho}{1+\rho}x}{(1+x)^{\frac{\rho}{1+\rho}}}. \quad (50)$$

Proof: We can still apply Theorem 2 here and we compute $\tilde{E}_o(\rho)$ as follows:

$$\begin{aligned} \tilde{E}_o(\rho) &= N(1 + \rho) \sup_{\mathbf{X} \neq \mathbf{0}; \|\mathbf{X}\|_\infty^2 \leq K_m} \frac{1}{\|\mathbf{X}\|^2} \ln \frac{\det(\mathbf{I} + \frac{\rho}{1+\rho} \mathbf{X}^\dagger \mathbf{X})}{\det(\mathbf{I} + \mathbf{X}^\dagger \mathbf{X})^{\frac{\rho}{1+\rho}}} \\ &= N(1 + \rho) \sup_{\mathbf{X} \neq \mathbf{0}; \|\mathbf{X}\|_\infty^2 \leq K_m} \frac{\sum_{l=1}^{T_c} \ln \frac{(1 + \frac{\rho}{1+\rho} \lambda_l)}{(1 + \lambda_l)^{\frac{\rho}{1+\rho}}}}{\sum_{l=1}^{T_c} \lambda_l} \\ &\leq N(1 + \rho) \sup_{0 < x \leq MT_c K_m} \frac{1}{x} \ln \frac{(1 + \frac{\rho}{1+\rho}x)}{(1+x)^{\frac{\rho}{1+\rho}}}. \end{aligned} \quad (51)$$

The inequality (51) is true since $\|\mathbf{X}\|_\infty^2 \leq K_m$ indicates that the largest eigenvalue of $\mathbf{X}^\dagger \mathbf{X}$ can be bounded as follows

$$\lambda_l \leq \text{Tr}(\mathbf{X}^\dagger \mathbf{X}) \leq MT_c K_m.$$

Now we have to show that the inequality (51) can be achieved as an equality by choosing an appropriate input matrix \mathbf{X} for any $\rho \in [0, 1]$. Assuming x_ρ is the optimizing x in (51), we choose \mathbf{X} to have identical entries which are all equal to $\frac{x_\rho}{MT_c}$. It is easy to see that this choice of input matrix will achieve the equality in (78). The calculation for $E_{sl}^{lp}(0)$ is very similar to the proof of Theorem 3. \diamond

Comparing (50) with (17), it is straightforward to see the following properties of $E^{lp}(r)$ for average and peak-constrained fading channels

Corollary 3 *The low-SNR reliability function $\overline{E}^{lp}(r)$ (50) with the additional peak power constraint K_m coincides with the reliability function $E^{lp}(r)$ (17) with average power constraint up to a rate r_m , which is the minimum rate satisfying $\tau(r_m) \geq K_m$, and $\tau(r)$ is the minimum signal peakiness with respect to rate r , as defined in (45). For any rate larger than r_m , $\overline{E}^{lp}(r)$ is strictly less than $E^{lp}(r)$ and intersects the line $\overline{E}^{lp}(r) = 0$ at*

$$\overline{C}^{lp} = N \left[1 - \frac{\ln(1 + MT_c K_m)}{MT_c K_m} \right].$$

Thus \overline{C}^{lp} is the low-SNR capacity with a peak constraint K_m . \diamond

Remarks: One important message here is that in the low-SNR regime communications with both average and peak power constraint, increasing the number of transmit antennas can actually increase both the capacity *and* the reliability function for the rate region near capacity, rather than being detrimental to the channel performance. Further, channels with longer coherence time also perform better than channels with shorter coherence time. The role of coherence time is similar to the role of multiple transmit antennas.

Now we look at a few numerical examples. For simplicity, we set $T_c = 1$ and the peak constraint $K_m = 1$. In Figure 3, when $M = 1$ and $M = 2$, it is easy to check that $r_m = 0$. Thus, the whole reliability curve is below the reliability function of channels with only an average power constraint. However, when $M \geq 4$, we have $r_m > 0$ and the phenomena we described in Corollary 3 starts to kick in. In Figure 4, we show how multiple transmit antennas improve the low-SNR capacity of a channel with both average and peak constraint.

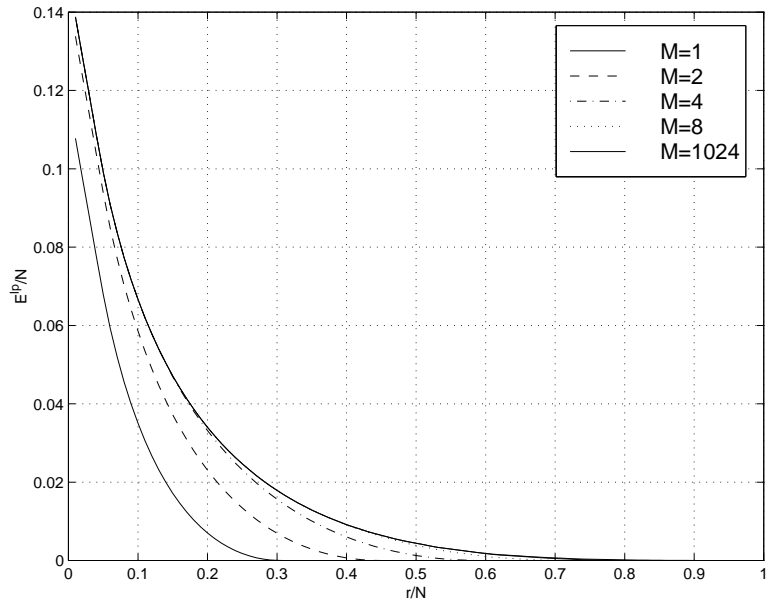


Figure 3: The reliability function for channels with both average and peak power constraint. $T_c = 1$.

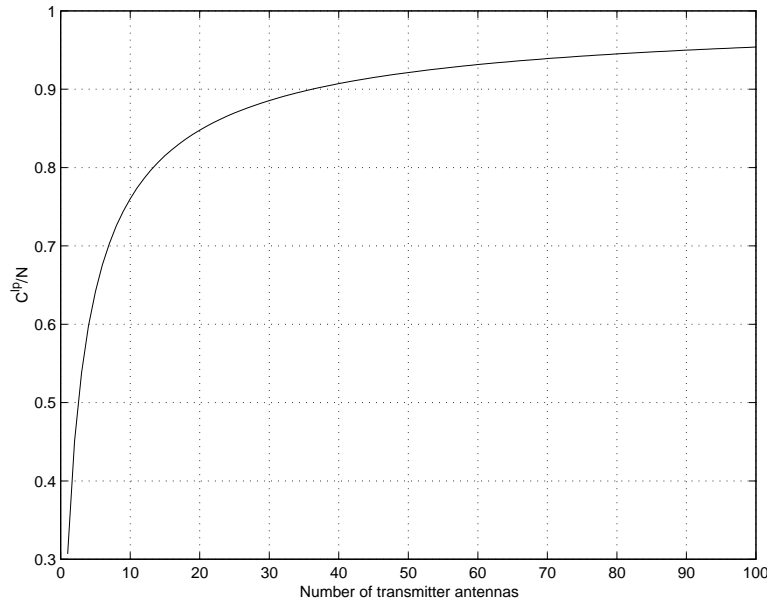


Figure 4: The capacity for channels with both average and peak power constraint. $T_c = 1$.

4 Correlated MIMO fading channels

In the last section, we studied the MIMO channel model (1) with the assumption that each entry of the fading matrix \mathbf{H} is an i.i.d. complex Gaussian random variable. Specifically, we assume that each entry of \mathbf{H} is a symmetric complex Gaussian random variable $\mathcal{CN}(0, 1)$ and we allow an arbitrary correlation between any two entries of \mathbf{H} . This correlation is usually referred to as the *spatial correlation*.

In this case, the transition probability density function $f(\mathbf{Y}|\mathbf{X})$ is a bit more complicated. We write $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{T_c})$, $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{T_c})$ and $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{T_c})$, where $\mathbf{y}_l, \mathbf{w}_l$'s are column vectors with dimension N , and \mathbf{x}_l 's are column vectors with dimension M . We have

$$\mathbf{y}_l = \mathbf{H}\mathbf{x}_l + \mathbf{w}_l, \quad l = 1, 2, \dots, T_c.$$

Define a larger column vector $\hat{\mathbf{y}}^T = (\mathbf{y}_1^T, \mathbf{y}_2^T, \dots, \mathbf{y}_{T_c}^T)$. Given the input matrix \mathbf{X} , $\hat{\mathbf{y}}^T$ is a joint complex Gaussian vector with variance

$$\mathcal{E}[\hat{\mathbf{y}}\hat{\mathbf{y}}^\dagger] = \mathbf{I} + \Xi,$$

where $\Xi = \{\Sigma_{ij}\}; \quad i, j = 1, 2, \dots, T_c$ and Σ_{ij} 's are $N \times N$ matrices such that for any i, j ,

$$\Sigma_{ij} = \mathcal{E}_{\mathbf{H}}[\mathbf{H}\mathbf{x}_i\mathbf{x}_j^\dagger\mathbf{H}^\dagger].$$

Thus, we have the following form of $f(\mathbf{Y}|\mathbf{X})$:

$$f(\mathbf{Y}|\mathbf{X}) = \frac{1}{\pi^{NT_c} \det(\mathbf{I} + \Xi)} \exp\{-\hat{\mathbf{y}}^\dagger(\mathbf{I} + \Xi)^{-1}\hat{\mathbf{y}}\}. \quad (52)$$

For general Gaussian random matrix \mathbf{H} , we cannot always get a closed-form expression for the reliability function as in the capacity case [22]. However, we can get an upper bound on the reliability function.

Theorem 6 *For the non-coherent MIMO model (1) with a general fading matrix \mathbf{H} , we have*

$$E^{lp}(r) = \sup_{0 \leq \rho \leq 1} -\rho r + \tilde{E}_o(\rho), \quad (53)$$

where $\tilde{E}_o(\rho)$ is upper bounded by

$$\tilde{E}_o(\rho) \leq G(1 + \rho) \sup_{x > 0} \frac{1}{x} \ln \frac{1 + \frac{\rho}{1+\rho}x}{(1+x)^{\frac{\rho}{1+\rho}}}, \quad (54)$$

and G is the maximum channel gain [22] and is defined as

$$G = \sup_{\mathbf{X} \neq \mathbf{0}} \frac{\mathcal{E}_{\mathbf{H}}[\|\mathbf{H}\mathbf{X}\|^2]}{\|\mathbf{X}\|^2} = \lambda_{max}(\mathcal{E}_{\mathbf{H}}[\mathbf{H}^\dagger\mathbf{H}]).$$

Here $\lambda_{max}(\mathbf{A})$ denotes the largest eigenvalue of matrix \mathbf{A} .

Proof: It is easy to verify

$$\tilde{E}_o(\rho) = (1 + \rho) \sup_{\mathbf{X} \neq \mathbf{0}} \frac{1}{\|\mathbf{X}\|^2} \ln \frac{\det(\mathbf{I} + \frac{\rho}{1+\rho}\Xi)}{\det(\mathbf{I} + \Xi)^{\frac{\rho}{1+\rho}}} \quad (55)$$

Assuming $\{\lambda_1, \lambda_2, \dots, \lambda_{NT_c}\}$ are the eigenvalues of Ξ , we have

$$\sum_{l=1}^{NT_c} \lambda_l = \text{Tr}(\Xi) = \sum_{i=1}^{T_c} \text{Tr}(\Sigma_{ii}) = \sum_{i=1}^{T_c} \text{Tr}(\mathcal{E}_{\mathbf{H}}[\mathbf{H}\mathbf{x}_i\mathbf{x}_i^\dagger\mathbf{H}^\dagger]) = \mathcal{E}_{\mathbf{H}}[\|\mathbf{H}\mathbf{X}\|^2].$$

We bound $\tilde{E}_o(\rho)$ as follows

$$\begin{aligned} \tilde{E}_o(\rho) &= (1 + \rho) \sup_{\mathbf{X} \neq \mathbf{0}} \frac{\mathcal{E}_{\mathbf{H}}[\|\mathbf{H}\mathbf{X}\|^2]}{\|\mathbf{X}\|^2} \frac{\sum_{l=1}^{NT_c} \ln \frac{1 + \frac{\rho}{1+\rho}\lambda_l}{(1+\lambda_l)^{\frac{\rho}{1+\rho}}}}{\sum_{l=1}^{NT_c} \lambda_l} \\ &\leq (1 + \rho) \sup_{\mathbf{X} \neq \mathbf{0}} \frac{\mathcal{E}_{\mathbf{H}}[\|\mathbf{H}\mathbf{X}\|^2]}{\|\mathbf{X}\|^2} \sup_{x>0} \frac{\ln \frac{1 + \frac{\rho}{1+\rho}x}{(1+x)^{\frac{\rho}{1+\rho}}}}{x} \\ &= G(1 + \rho) \sup_{x>0} \frac{1}{x} \ln \frac{1 + \frac{\rho}{1+\rho}x}{(1+x)^{\frac{\rho}{1+\rho}}}. \end{aligned} \quad (56)$$

The maximum channel gain G for this case can be computed as

$$\begin{aligned} G &= \sup_{\mathbf{X} \neq \mathbf{0}} \frac{\mathcal{E}_{\mathbf{H}}[\|\mathbf{H}\mathbf{X}\|^2]}{\|\mathbf{X}\|^2} \\ &= \sup_{\mathbf{X} \neq \mathbf{0}} \frac{\sum_{l=1}^{T_c} \mathbf{x}_l^\dagger \mathcal{E}_{\mathbf{H}}[\mathbf{H}^\dagger \mathbf{H}] \mathbf{x}_l}{\|\mathbf{X}\|^2} \\ &\leq \sup_{\mathbf{X} \neq \mathbf{0}} \frac{\sum_{l=1}^{T_c} \lambda_{\max}(\mathcal{E}_{\mathbf{H}}[\mathbf{H}^\dagger \mathbf{H}]) \|\mathbf{x}_l\|^2}{\|\mathbf{X}\|^2} \\ &= \lambda_{\max}(\mathcal{E}_{\mathbf{H}}[\mathbf{H}^\dagger \mathbf{H}]) \end{aligned} \quad (57)$$

It is easy to verify the low rate condition (15) is true using similar bounding techniques as above and thus the theorem is proved. \diamond

Remarks: Now we discuss the conditions under which this upper bound can be achieved. It is easy to see that for the equality in (54) to hold, we need both inequalities (56) and (57) to be achieved as equalities. We know (56) can be made an equality if the eigenvalues of Ξ are either zero or x_ρ . Equation (57) is an equality when for any l , \mathbf{x}_l is in the eigenspace corresponding to $\lambda_{\max}(\mathcal{E}_{\mathbf{H}}[\mathbf{H}^\dagger \mathbf{H}])$. Unfortunately, sometimes it is not possible for both conditions to be true. As a simple example, consider the case when $M = 1$. In this case, the equality in (57) is trivially true for any \mathbf{X} . However, we will lose control of the eigenvalues of Ξ , since in this case,

$$\Sigma_{ij} = \mathcal{E}_{\mathbf{H}}[\mathbf{H}\mathbf{H}^\dagger] \mathbf{x}_i \mathbf{x}_j^\dagger.$$

We claim that for any eigenvalue λ of $\mathcal{E}_{\mathbf{H}}[\mathbf{H}\mathbf{H}^\dagger]$, there is a corresponding eigenvalue $\lambda' = \lambda\|\mathbf{X}\|^2$ for $\Xi = \{\Sigma_{ij}\}$ in the case when $M = 1$. To see this, assume λ is an eigenvalue of $\mathcal{E}_{\mathbf{H}}[\mathbf{H}\mathbf{H}^\dagger]$ and $\mathbf{v} \in \mathcal{C}^{N \times 1}$ is the corresponding eigenvector. We must have

$$\mathcal{E}_{\mathbf{H}}[\mathbf{H}\mathbf{H}^\dagger]\mathbf{v} = \lambda\mathbf{v}.$$

Let $\mathbf{B} \in \mathcal{C}^{T_c \times T_c}$ be a matrix whose (i, j) entry is $\mathbf{x}_i \mathbf{x}_j^\dagger$. Denote $\alpha = \mathbf{B}\mathbf{1}$, where $\mathbf{1}$ is the all-1 vector of dimension T_c . Construct a new column vector $\tilde{\mathbf{v}}$ of dimension NT_c as follows

$$\tilde{\mathbf{v}}^t = (\alpha_1 \mathbf{v}^t, \alpha_2 \mathbf{v}^t, \dots, \alpha_{T_c} \mathbf{v}^t)$$

It is easy to check that

$$\Xi \tilde{\mathbf{v}} = \lambda\|\mathbf{X}\|^2 \tilde{\mathbf{v}},$$

and thus $\lambda\|\mathbf{X}\|^2$ must be an eigenvalue for Ξ . In the case when $\mathcal{E}_{\mathbf{H}}[\mathbf{H}\mathbf{H}^\dagger]$ has more than one non-zero eigenvalues which are unequal, we can not make (56) an equality, i.e., make all the scaled eigenvalues to be either 0 or x_ρ . On the other hand, for the opposite extreme case when $N = 1$, we can check that the upper bound in (54) can be achieved.

Corollary 4 *When we have M transmit antennas and one receive antenna, the low-SNR reliability function is*

$$E^{lp}(r) = \sup_{0 \leq \rho \leq 1} -\rho r + \lambda_{max}(\mathcal{E}_{\mathbf{H}}[\mathbf{H}^\dagger \mathbf{H}]) (1 + \rho) \sup_{x > 0} \frac{1}{x} \ln \frac{1 + \frac{\rho}{1+\rho} x}{(1+x)^{\frac{\rho}{1+\rho}}} \quad (58)$$

and thus has a performance gain of $\lambda_{max}(\mathcal{E}_{\mathbf{H}}[\mathbf{H}^\dagger \mathbf{H}])$.

Proof: We choose \mathbf{X} such that for any $l = 1, 2, \dots, T_c$, $\mathbf{x}_l = \mathbf{x}$. When $N = 1$, $\Sigma_{ij} = \mathcal{E}_{\mathbf{H}}[\mathbf{H}\mathbf{x}\mathbf{x}^\dagger \mathbf{H}^\dagger] = \mathbf{x}^\dagger \mathcal{E}_{\mathbf{H}}[\mathbf{H}^\dagger \mathbf{H}]\mathbf{x} = \sigma$ is a scalar and Ξ has only one eigenvalue σT_c . Choose \mathbf{x} such that it lies in the eigenspace corresponding to $\lambda_{max}(\mathcal{E}_{\mathbf{H}}[\mathbf{H}^\dagger \mathbf{H}])$ and $\|\mathbf{x}\|^2 \lambda_{max} T_c = x_\rho$ and it is easy to see that both (56) and (57) are satisfied. \diamond

Now we study the quantity $\lambda_{max}(\mathcal{E}_{\mathbf{H}}[\mathbf{H}^\dagger \mathbf{H}])$. We claim

$$\lambda_{max}(\mathcal{E}_{\mathbf{H}}[\mathbf{H}^\dagger \mathbf{H}]) \geq N, \quad (59)$$

which is the channel gain under independently faded matrix \mathbf{H} . To see this, consider

$$\sum_{l=1}^M \lambda_l = Tr(\mathcal{E}_{\mathbf{H}}[\mathbf{H}^\dagger \mathbf{H}]) = MN \leq M\lambda_{max}.$$

Thus for any fading matrix \mathbf{H} , (59) must be true, which means the reliability function for a correlated \mathbf{H} may be larger than (16). For the case when $N = 1$, (58) tells us that any fading matrix \mathbf{H} will do better than the

independent fading case. The independent fading channel model is widely studied because it can provide better *diversity gain* and *degree of freedom gain* in the high SNR regime, as compared with the correlated fading matrix. To get approximately independent fading paths, we have to put the antennas physically far apart (more than half the transmission wavelength), which limits the possibility of having multiple antennas on small communication devices. However, in the low-SNR regime, the *diversity gain* and *degree of freedom gain* are not more important than the *power gain*. Thus, in this regime, the design goal of utilizing the MIMO system has changed and it might not be a good idea to *make* the fading paths independent at all.

Next we assume that we have control over the fading matrix \mathbf{H} and explore the best fading matrix \mathbf{H} in the sense that we can obtain the best reliability function under this fading matrix. Specifically, we need to maximize the reliability function (53) over \mathbf{H} under the constraint that \mathbf{H} is complex Gaussian and each entry has variance 1.

Theorem 7 *A fully correlated fading channel matrix \mathbf{H} ($h_{ij} = h; \forall i, j$) provides the maximum performance gain and the reliability function for this fading model is*

$$E^{lp}(r) = \sup_{0 \leq \rho \leq 1} -\rho r + \tilde{E}_o(\rho), \quad (60)$$

where $\tilde{E}_o(\rho)$ is upper bounded by

$$\tilde{E}_o(\rho) = MN(1 + \rho) \sup_{x > 0} \frac{1}{x} \ln \frac{1 + \frac{\rho}{1+\rho}x}{(1+x)^{\frac{\rho}{1+\rho}}}. \quad (61)$$

Proof: This model is a special case of the general fading model we discussed above. Given the proof of Theorem 6, it suffices to show here that for any \mathbf{H} ,

$$\lambda_{max}(\mathcal{E}_{\mathbf{H}}[\mathbf{H}^\dagger \mathbf{H}]) \leq MN, \quad (62)$$

where the equality can be achieved by this choice of \mathbf{H} , and the two inequalities (56), (57) can be made equalities for this special fading matrix by choosing an appropriate \mathbf{X} .

It is easy to see that (62) is true since

$$\lambda_{max}(\mathcal{E}_{\mathbf{H}}[\mathbf{H}^\dagger \mathbf{H}]) \leq Tr(\mathcal{E}_{\mathbf{H}}[\mathbf{H}^\dagger \mathbf{H}]) = MN.$$

When \mathbf{H} has identical entries with variance 1, $\mathcal{E}_{\mathbf{H}}[\mathbf{H}^\dagger \mathbf{H}]$ is a matrix of dimension $M \times M$ with all entries equal to N . Thus, $\lambda_{max}(\mathcal{E}_{\mathbf{H}}[\mathbf{H}^\dagger \mathbf{H}]) = MN$ for fully correlated fading.

Next we choose a scalar x and repeat it at all entries of the input matrix \mathbf{X} . First we check that (57) is achieved as an equality. In this case, $\|\mathbf{X}\|^2 = MT_c|x|^2$ and

$$\mathcal{E}_{\mathbf{H}}[\|\mathbf{H}\mathbf{X}\|^2] = M^2NT_c|x|^2.$$

Thus,

$$\frac{\mathcal{E}_{\mathbf{H}}[\|\mathbf{H}\mathbf{X}\|^2]}{\|\mathbf{X}\|^2} = MN.$$

For (56), it is also easy to see that Ξ has only one non-zero eigenvalue which is equal to $\mathcal{E}_{\mathbf{H}}[\|\mathbf{H}\mathbf{X}\|^2] = M^2NT_c|x|^2$. If we choose $|x|^2 = \frac{x\rho}{M^2NT_c}$, we are done. \diamond

Next, we find the minimum signal peakiness $\tau(r)$ for this fully correlated MIMO channel. With the input matrix we constructed in the proof of Theorem 7, it is straightforward to have the following theorem:

Corollary 5 *For the fully correlated MIMO channel, the minimum signaling peakiness is*

$$\tau(r) = \frac{\kappa'(r)}{M^2NT_c}. \quad (63)$$

where $\kappa'(r)$ is defined as

$$\kappa'(r) = \bar{\tau}\left(\frac{r}{MN}\right) = \arg \max_{x>0} \sup_{0 \leq \rho \leq 1} -\rho \frac{r}{MN} + (1 + \rho) \frac{1}{x} \ln \frac{(1 + \frac{\rho}{1+\rho}x)}{(1+x)^{\frac{\rho}{1+\rho}}}.$$

In other words, when we have a fully correlated MIMO fading channel, we can achieve a performance gain of MN and a peakiness gain of M^2NT_c , as compared to a SISO fast-fading channel. \diamond

It is not necessary to require that \mathbf{H} has identical entries in order to achieve the optimal performance gain in a MIMO channel. Next we consider the case that any two entries of \mathbf{H} differ by a phase shift:

$$\mathcal{E}[h_{nm}h_{n'm'}] = e^{j\theta_{nmn'm'}}, \quad (64)$$

for any $n, n' = 1, 2, \dots, N$ and $m, m' = 1, 2, \dots, M$. Further, we assume the correlation between any pair of transmitter and receive antenna is the product of the *transmit correlation* and *receiver correlation* [1], i.e.,

$$\mathcal{E}[h_{nm}h_{n'm'}] = R_{mm'}^r R_{nn'}^t,$$

where $\mathbf{R}^r = \{R_{mm'}^r\}$ and $\mathbf{R}^t = \{R_{nn'}^t\}$ are the transmit correlation matrix and the receive correlation matrix, respectively. In our case, where the only difference between any two entries of \mathbf{H} is a phase shift as in (64), it is easy to see that this is only possible when all entries of \mathbf{R}^r or \mathbf{R}^t have magnitude 1. Denote $\mathbf{u} = (1, e^{ju_2}, \dots, e^{ju_M})^T$ and $\mathbf{v} = (1, e^{jv_2}, \dots, e^{jv_N})^T$. Assume $\mathbf{R}^t = \mathbf{u}\mathbf{u}^\dagger$ and $\mathbf{R}^r = \mathbf{v}\mathbf{v}^\dagger$. Physically, this form of correlation matrix

means that the phase shift between the m -th and m' -th transmit antenna is $u_m - u_{m'}$ and the phase shift between the n -th and n' -th receive antenna is $v_n - v_{n'}$ ($u_1 = v_1 = 0$). Thus,

$$\mathcal{E}[h_{nm}h_{n'm'}] = e^{j(u_m - u_{m'} + v_n - v_{n'})}. \quad (65)$$

Let \mathbf{H}_h denote the fading matrix with identical entries, as the one we assumed in Theorem 7. The correlated fading matrix with correlation (65) can then be written as

$$\mathbf{H} = \mathbf{V}\mathbf{H}_h\mathbf{U}, \quad (66)$$

where $\mathbf{U} = \text{diag}(\mathbf{u})$ and $\mathbf{V} = \text{diag}(\mathbf{v})$.

Corollary 6 *For a MIMO fading channel with fading matrix \mathbf{H} satisfying (66), we can also achieve a performance gain of MN and peakiness gain M^2NT_c .*

Proof: For \mathbf{H} determined by (66), we can write the channel as

$$\mathbf{Y} = \mathbf{V}\mathbf{H}_h\mathbf{U}\mathbf{X} + \mathbf{W}.$$

Consider $\tilde{\mathbf{X}} = \mathbf{U}\mathbf{X}$, $\tilde{\mathbf{Y}} = \mathbf{V}^{-1}\mathbf{Y}$ and $\tilde{\mathbf{W}} = \mathbf{V}^{-1}\mathbf{W}$. We have an equivalent channel

$$\tilde{\mathbf{Y}} = \mathbf{H}_h\tilde{\mathbf{X}} + \tilde{\mathbf{W}}. \quad (67)$$

It is easy to check that both \mathbf{U} and \mathbf{V} are unitary matrices. Thus $\tilde{\mathbf{W}}$ in (67) has the same distribution as \mathbf{W} . Further, the power constraint for the input for (67) is

$$\mathcal{E}[\|\tilde{\mathbf{X}}\|^2] = \mathcal{E}[\text{Tr}(\tilde{\mathbf{X}}\tilde{\mathbf{X}}^\dagger)] = \mathcal{E}[\text{Tr}(\mathbf{U}\mathbf{X}\mathbf{X}^\dagger\mathbf{U}^\dagger)] = \mathcal{E}[\text{Tr}(\mathbf{X}\mathbf{X}^\dagger)] = \mathcal{E}[\|\mathbf{X}\|^2] \leq pT_c.$$

This channel is equivalent to the channel considered in Theorem 7 and thus we can also achieve the same performance gain of MN by choosing $\tilde{\mathbf{X}} = \mathbf{X}^*$ (equivalently, $\mathbf{X} = \mathbf{U}^{-1}\mathbf{X}^*$), where \mathbf{X}^* is the optimal input matrix we constructed in the proof of Theorem 7. Moreover, it is easy to see that the same peakiness gain is achieved by the same input matrix. \diamond

Now we look at the physical conditions on the antenna spacing under which these two fully correlated fading models are appropriate, using the angular domain representation of the MIMO multi-path fading channel [13, 20]. Let Δ_t and Δ_r denote the spacing distances, normalized by the transmission wavelength λ_c , between adjacent transmit or receive antennas and let L_t and L_r denote the normalized transmit and receive antenna apertures (size

of the antenna array). Thus, we must have $L_t = M\Delta_t$ and $L_r = N\Delta_r$. The angular basis matrices for both the transmit and receive sides are defined as

$$\begin{aligned}\mathbf{U}_r &= (\mathbf{e}_r(0), \mathbf{e}_r(\frac{1}{L_r}), \dots, \mathbf{e}_r(\frac{N-1}{L_r})); \\ \mathbf{U}_t &= (\mathbf{e}_t(0), \mathbf{e}_t(\frac{1}{L_t}), \dots, \mathbf{e}_t(\frac{M-1}{L_t})),\end{aligned}$$

where $\mathbf{e}_t(\Omega)$ and $\mathbf{e}_r(\Omega)$ are the *spatial signature* vectors and are defined as [20]

$$\begin{aligned}\mathbf{e}_t(\Omega) &= \frac{1}{\sqrt{M}} \begin{bmatrix} 1 \\ \exp(j2\pi\Delta_t\Omega) \\ \vdots \\ \exp(j2\pi(M-1)\Delta_t\Omega) \end{bmatrix}; \\ \mathbf{e}_r(\Omega) &= \frac{1}{\sqrt{N}} \begin{bmatrix} 1 \\ \exp(j2\pi\Delta_r\Omega) \\ \vdots \\ \exp(j2\pi(N-1)\Delta_r\Omega) \end{bmatrix}.\end{aligned}$$

The physical paths whose transmit solid angles $\Omega_t = \cos \phi_t$, where ϕ_t denotes the transmit angle, have differences within $\frac{1}{L_t}$ and receive solid angles have differences within $\frac{1}{L_r}$ are highly correlated and thus are called *space unresolvable*. Based on this observation, we can divide paths into different angular windows which correspond to different solid angle intervals of length $\frac{1}{L_t}$ and $\frac{1}{L_r}$ and further, the fading matrix \mathbf{H} can be connected to an angular domain fading matrix \mathbf{H}^a by

$$\mathbf{H} = \mathbf{U}_r \mathbf{H}^a \mathbf{U}_t^\dagger, \quad (68)$$

where each entry h_{ij}^a of \mathbf{H}^a denotes the aggregated path gain for all paths starting from the j -th transmit angular window to the i -th receive angular window.

In a richly-scattering environment, when the antenna apertures at both the transmit and receive sides are less than half the wavelength, all the physical paths become space unresolvable. Thus, in the angular domain, the corresponding fading matrix \mathbf{H}^a only has only one non-zero entry, i.e., $h_{11}^a = h$, where h can be assumed to be complex Gaussian due to large number of paths. Consequently, from (68), it is easy to see that the fading matrix \mathbf{H} will have identical entries. Thus, if we put both the transmit and receive antennas within half the transmission wavelength, we can get a fading matrix \mathbf{H} with identical entries.

For the other fading model where we assume any two entries of \mathbf{H} differ by a phase shift satisfying (65), we consider the following scenario where the physical paths are not evenly distributed across different transmit

and receive angles. Instead, there is one dominate angle such that all the non-negligible paths are included in a neighborhood of this angle. When the space sample resolutions $\frac{1}{L_r}$ and $\frac{1}{L_t}$ are low enough, or equivalently, when L_t and L_r are small enough, (*not necessarily less than $\frac{\lambda_c}{2}$,*) these paths will fit into the same transmit-receive angular window pair. Thus, only one space-resolvable path exists and the corresponding \mathbf{H}^a has only one non-zero entry, i.e., $h_{ij}^a = h$, for some i, j (not necessarily $i = j = 1$). From (68), it is easy to check that in this case the fading matrix \mathbf{H} satisfies (66).

Remarks: Theorem 7, Corollary 5 and Corollary 6 tell us the potential performance gain and peakiness gain we can have with MIMO channels in the low-SNR regime, if the entries of the fading matrix \mathbf{H} are fully correlated. The first observation here is that it is important to make the fading correlated in the low-SNR regime. We can potentially have a reliability function which is M times better than the result with independent fading matrix. For this model, both the low-SNR capacity and reliability function increases in proportion to the *product* of the number of the transmit antennas and the receive antennas. Further, the minimum signal peakiness reduces by a factor of $\frac{1}{MN}$ as compared to the independent fading case, which makes it much easier to achieve the reliability function. Thus, in the low-SNR regime with multiple antennas, we should try to make the channel fading more correlated, (for example, by putting the antennas close to each other physically) to fully utilize the advantage of having multiple antennas.

The reason that spatial correlation is beneficial in this regime is that in the low-SNR regime, the power gain directly determines the performance gain. Correlated channel paths between transmit and receive antenna pairs help amplify the transmitted signal and thus achieve a better power gain.

Another observation here is that multiple transmit antennas can be beneficial. For the independent fading case, we have shown that multiple receive antennas give us a performance gain while multiple transmit antennas only provide a peakiness gain. However, with fully correlated fading, we proved that multiple transmit antennas contribute a peakiness gain of M^2 and a performance gain of M , while multiple receive antennas give us a performance gain of N and a peakiness gain of N . Thus, transmit antennas are very important for low-SNR communications. To get more insight into this phenomena, we consider a simple example with $M = 2$ and $N = T_c = 1$ and compare it to a SISO channel. Further, assume we have fully correlated fading over the two paths. The channel can be represented by

$$y = h(x_1 + x_2) + w,$$

with an average power constraint $\mathcal{E}[|x_1|^2 + |x_2|^2] = p$. Consider the naive transmitting scheme when we always

transmit $x_1 = x_2 = x$. Thus, under this signaling scheme, the channel becomes

$$y = h(2x) + w,$$

with an average power constraint $\mathcal{E}[|x|^2] = \frac{p}{2}$. For the performance gain, we consider capacity for simplicity.

Let C_M denotes the capacity with M transmit antennas. The quantity of interest in the low-SNR regime is $\dot{C}_2(0) = \lim_{p \rightarrow 0} \frac{C_2(p)}{p}$. It is easy to see $C_2(p) = C_1(2p)$ and thus

$$\dot{C}_2(0) = \lim_{p \rightarrow 0} \frac{C_1(2p)}{p} = 2\dot{C}_1(0),$$

where we find our performance gain. For the peakiness gain, assume κ is the minimum peakiness for the SISO channel, which means an on-off signaling scheme with $\sqrt{\kappa}$ as the non-zero symbol is optimal for a certain point in the reliability curve. With the same fading, for the two-transmitter channel, the same on-off signaling scheme is still optimal except now we only need $2x = \sqrt{\kappa}$ and thus $|x|^2 = \frac{\kappa}{4}$, which gives us the peakiness gain.

5 Conclusions

In this paper, we investigated the tradeoff between communication rate and average probability of decoding error for a non-coherent multiple-antenna fading channel in a low-SNR regime, using the framework of error-exponent theory. We started with the assumption that the fading matrix \mathbf{H} has i.i.d. entries. In this regime, we showed that using M transmit antennas and N receive antennas allow us to realized a performance gain of N and peakiness gain M . Further, if the channel is constant every T_c symbols, we can see a further peakiness gain of T_c . However, neither increasing M or T_c can improve the asymptotic communication rate or the reliability function of the channel. Further, when both the average and peak power are constrained, having larger M or T_c can improve both the channel capacity and the low-SNR reliability function.

In the low-SNR regime, channel correlation can actually improve the channel performance. In the extreme case where the fading is fully correlated, in the sense that the entries of the fading matrix \mathbf{H} are either identical or differ by a phase shift, we can achieve a performance gain of MN and a peakiness gain of M^2NT_c . Thus, the advantage of having multiple antennas is best realized when we have fully correlated fading channels. This suggests that the antennas should be placed close together in the low-SNR regime.

A Proof of Theorem 2

We prove Theorem 2 in the following three subsections.

A.1 Sphere-packing bound

First, we consider the sphere-packing bound (7) and show

$$E^{lp}(r) \leq E_{sp}^{lp}(r).$$

We start with the expression for $E_o(p, \rho)$ in (8):

$$\begin{aligned} & \int \left(\int q(x) e^{\beta(b(x)-p)} f(y|x)^{\frac{1}{1+\rho}} dx \right)^{1+\rho} dy \\ &= \int f(y|0) \left(\int q(x) e^{\beta(b(x)-p)} \left[\frac{f(y|x)}{f(y|0)} \right]^{\frac{1}{1+\rho}} dx \right)^{1+\rho} dy \\ &\geq \left(\int f(y|0) \int q(x) e^{\beta(b(x)-p)} \left[\frac{f(y|x)}{f(y|0)} \right]^{\frac{1}{1+\rho}} dx dy \right)^{1+\rho} \end{aligned} \quad (69)$$

$$\begin{aligned} &= \left(\int q(x) e^{\beta(b(x)-p)} \int f(y|0)^{\frac{\rho}{1+\rho}} f(y|x)^{\frac{1}{1+\rho}} dy dx \right)^{1+\rho} \\ &= \left(\int q(x) e^{\beta(b(x)-p)} e^{\ln \int f(y|0)^{\frac{\rho}{1+\rho}} f(y|x)^{\frac{1}{1+\rho}} dy} dx \right)^{1+\rho} \\ &\geq e^{(1+\rho)\mathcal{E}[\ln \int f(y|0)^{\frac{\rho}{1+\rho}} f(y|x)^{\frac{1}{1+\rho}} dy]} \end{aligned} \quad (70)$$

The inequalities (69) and (70) are two applications of Jensen's inequality, due to the convexity of the two functions $f(t) = t^{1+\rho}$ and $g(t) = e^t$. From here, it is easy to see that

$$\begin{aligned} E_o(p, \rho) &\leq \sup_{q \in \mathcal{D}(p)} \sup_{\beta \geq 0} -(1+\rho)\mathcal{E}[\ln \int f(y|0)^{\frac{\rho}{1+\rho}} f(y|x)^{\frac{1}{1+\rho}} dy] \\ &= \sup_{q \in \mathcal{D}(p)} -(1+\rho) \sum_{k=0}^K q(x_k) \ln \int f(y|0)^{\frac{\rho}{1+\rho}} f(y|x_k)^{\frac{1}{1+\rho}} dy \end{aligned} \quad (71)$$

Note the sup in (71) is over all possible choices of the finite discrete alphabet x_0, x_1, \dots, x_K , and all possible probability assignments $q(x_0), q(x_1), \dots, q(x_K)$ under the constraint that $\sum_{k=0}^K q(x_k) b(x_k) = p$. However, if for some k , $x_k = 0$, it is easy to check that

$$\ln \int f(y|0)^{\frac{\rho}{1+\rho}} f(y|x_k)^{\frac{1}{1+\rho}} dy = 0.$$

Thus, we can remove the zero-cost symbol from the summation in (71) and we have

$$\begin{aligned} E_o(p, \rho) &\leq \sup_{q \in \mathcal{D}(p)} -(1+\rho) \sum_{k: x_k \neq 0} q(x_k) \ln \int f(y|0)^{\frac{\rho}{1+\rho}} f(y|x_k)^{\frac{1}{1+\rho}} dy \\ &= \sup_{q \in \mathcal{D}(p)} \sum_{k: x_k \neq 0} q(x_k) \frac{-(1+\rho) \ln \int f(y|0)^{\frac{\rho}{1+\rho}} f(y|x_k)^{\frac{1}{1+\rho}} dy}{b(x_k)} \end{aligned}$$

$$\begin{aligned}
&\leq \sup_{q \in \mathcal{D}(p)} \sum_{k: x_k \neq 0} q(x_k) b(x_k) \sup_{x \in A'} \frac{-(1+\rho) \ln \int f(y|0)^{\frac{\rho}{1+\rho}} f(y|x)^{\frac{1}{1+\rho}} dy}{b(x)} \\
&= p \sup_{x \in A'} \frac{-(1+\rho) \ln \int f(y|0)^{\frac{\rho}{1+\rho}} f(y|x)^{\frac{1}{1+\rho}} dy}{b(x)} \\
&= p \tilde{E}_o(\rho)
\end{aligned} \tag{72}$$

From the definition of $E^{lp}(r)$ and applying the sphere-packing bound here, we have

$$E^{lp}(r) \leq \limsup_{p \rightarrow 0} \frac{E_{sp}(rp, p)}{p} \leq \sup_{\rho \geq 0} -\rho r + \tilde{E}_o(\rho).$$

Remarks: The main modification we have here, as compared to Gallager's proof, is that we apply Jensen's inequality in (70), rather than using Gallager's arguments based on the fact that all alphabet symbols except 0 are larger than p , which is not true here since we allow the alphabet symbols to change with SNR. The equality will hold in (70) if and only if the expression $e^{\beta b(x)} \int f(y|0)^{\frac{\rho}{1+\rho}} f(y|x)^{\frac{1}{1+\rho}} dy$ takes the same value for all symbols which occur with non-zero probability. As pointed out in [3], the bound in (70) will be tight if we choose on-off signaling. With this modification, we next apply (72), which is motivated by the simple inequality (15) in [21], to get the sphere-packing bound. The other inequality (69) in the above proof is essentially a procedure of *linearization* at low SNR. Using a key inequality established by Gallager in [3], we will show in next section that the non-linear terms actually does not affect this first-order calculation. Thus, all inequalities used to achieve this upper bound on the reliability function can be made equalities by on-off signaling. This gives us some intuition about why this seemingly-arbitrary bound is tight.

A.2 Random-coding bound

Next we show the random-coding bound

$$E^{lp}(r) \geq E_r^{lp}(r).$$

Consider an on-off signaling scheme as follows,

$$q(x) = \begin{cases} x_o & w.p. \frac{p}{b(x_o)}; \\ 0 & w.p. 1 - \frac{p}{b(x_o)}, \end{cases} \tag{73}$$

where x_o is an arbitrary non-zero constant not changing with p . Again, we start with the expression for $E_o(p, \rho)$ in (8). With this particular choice of q , we have a lower bound for $E_o(p, \rho)$,

$$E_o(p, \rho) \geq \sup_{\beta \geq 0} -\ln \int \left(\int q(x) e^{\beta(b(x)-p)} f(y|x)^{\frac{1}{1+\rho}} dx \right)^{1+\rho} dy$$

$$\begin{aligned}
&= \sup_{\beta \geq 0} -\ln \int \left(\left(1 - \frac{p}{b(x_o)}\right) e^{-\beta p} f(y|0)^{\frac{1}{1+\rho}} + \frac{p}{b(x_o)} e^{\beta(b(x_o)-p)} f(y|x_o)^{\frac{1}{1+\rho}} \right)^{1+\rho} dy \\
&= \sup_{\beta \geq 0} (1+\rho)\beta p - \ln \int f(y|0) \left(1 - \frac{p}{b(x_o)} + \frac{p}{b(x_o)} e^{\beta b(x_o)} \left(\frac{f(y|x_o)}{f(y|0)} \right)^{\frac{1}{1+\rho}} \right)^{1+\rho} dy \quad (74)
\end{aligned}$$

To obtain a further bound, we make use of the following bound (Lemma 1 of [3]):

Lemma 3 For any $\rho \in [0, 1]$, any $x \geq 0$, and any $z \in [0, 1]$,

$$(1+x-z)^{1+\rho} \leq 1 + (1+\rho-\rho z)(x-z) + x^{1+\rho}. \quad (75)$$

◇

By treating $\frac{p}{b(x_o)}$ as z and $\frac{p}{b(x_o)} e^{\beta b(x_o)} \left(\frac{f(y|x_o)}{f(y|0)} \right)^{\frac{1}{1+\rho}}$ as x in (74), we can carry out the integral as follows

$$\begin{aligned}
&\int f(y|0) \left(1 - \frac{p}{b(x_o)} + \frac{p}{b(x_o)} e^{\beta b(x_o)} \left(\frac{f(y|x_o)}{f(y|0)} \right)^{\frac{1}{1+\rho}} \right)^{1+\rho} dy \\
&\leq \int f(y|0) \left\{ 1 + (1+\rho-\rho \frac{p}{b(x_o)}) \left[\frac{p e^{\beta b(x_o)}}{b(x_o)} \left(\frac{f(y|x_o)}{f(y|0)} \right)^{\frac{1}{1+\rho}} - \frac{p}{b(x_o)} \right] + \left[\frac{p e^{\beta b(x_o)}}{b(x_o)} \left(\frac{f(y|x_o)}{f(y|0)} \right)^{\frac{1}{1+\rho}} \right]^{1+\rho} \right\} dy \\
&= 1 + (1+\rho-\rho \frac{p}{b(x_o)}) \left[\frac{p}{b(x_o)} e^{\beta b(x_o)} \int f(y|0)^{\frac{\rho}{1+\rho}} f(y|x_o)^{\frac{1}{1+\rho}} dy - \frac{p}{b(x_o)} \right] + p^{1+\rho} \frac{e^{(1+\rho)\beta b(x_o)}}{b(x_o)^{1+\rho}} \quad (76)
\end{aligned}$$

It is easy to see that equation (74) can be lower bounded by choosing $\beta = \beta_o = -\frac{1}{b(x_o)} \ln \int f(y|0)^{\frac{\rho}{1+\rho}} f(y|x_o)^{\frac{1}{1+\rho}} dy$.

After doing this, the second term in (76) will vanish and we get the following lower bound for $E_o(p, \rho)$:

$$\begin{aligned}
E_o(p, \rho) &\geq (1+\rho)\beta_o p - \ln \left(1 + p^{1+\rho} \frac{e^{(1+\rho)\beta_o b(x_o)}}{b(x_o)^{1+\rho}} \right) \\
&\geq (1+\rho)\beta_o p - \frac{e^{(1+\rho)\beta_o b(x_o)}}{b(x_o)^{1+\rho}} p^{1+\rho}
\end{aligned}$$

Thus, we have

$$\begin{aligned}
E^{lp}(r) &\geq \liminf_{p \rightarrow 0} \frac{E_r^{lp}(rp)}{p} \\
&\geq \liminf_{p \rightarrow 0} \frac{\sup_{0 \leq \rho \leq 1} -\rho r p + (1+\rho)\beta_o p - \frac{e^{(1+\rho)\beta_o b(x_o)}}{b(x_o)^{1+\rho}} p^{1+\rho}}{p} \\
&= \liminf_{p \rightarrow 0} \sup_{0 \leq \rho \leq 1} -\rho r + (1+\rho)\beta_o - \frac{e^{(1+\rho)\beta_o b(x_o)}}{b(x_o)^{1+\rho}} p^\rho \quad (77)
\end{aligned}$$

Now we claim that we always have

$$E^{lp}(r) \geq \sup_{0 \leq \rho \leq 1} -\rho r + (1+\rho)\beta_o. \quad (78)$$

To see this, consider the following two situations. For the first case, assume the optimizing ρ in (78) is $\rho^* > 0$. The corresponding β_o equals to β_o^* . In this case, we utilize (77) and we have

$$\begin{aligned} E^{lp}(r) &\geq \liminf_{p \rightarrow 0} -\rho^* r + (1 + \rho^*)\beta_o^* - \frac{e^{(1+\rho^*)\beta_o^* b(x_o)}}{b(x_o)^{1+\rho^*}} p^{\rho^*} \\ &= -\rho^* r + (1 + \rho^*)\beta_o^* \\ &= \sup_{0 \leq \rho \leq 1} -\rho r + (1 + \rho)\beta_o. \end{aligned}$$

For the second case, consider the optimizing ρ in (78) is $\rho^* = 0$. For this case, it is easy to check that the corresponding $\beta_o^* = 0$. Thus, the right side of (78) is actually zero. Since we always have $E(r) \geq 0$, we again have (78).

So far, we have proved that for any $x_o > 0$, we always have

$$E^{lp}(r) \geq \sup_{0 \leq \rho \leq 1} -\rho r - (1 + \rho) \frac{\ln \int f(y|0)^{\frac{\rho}{1+\rho}} f(y|x_o)^{\frac{1}{1+\rho}} dy}{b(x_o)}.$$

However, since $x_o \neq 0$ is arbitrary, we have

$$\begin{aligned} E^{lp}(r) &\geq \sup_{x_o \in A'} \sup_{0 \leq \rho \leq 1} -\rho r - (1 + \rho) \frac{\ln \int f(y|0)^{\frac{\rho}{1+\rho}} f(y|x_o)^{\frac{1}{1+\rho}} dy}{b(x_o)} \\ &= \sup_{0 \leq \rho \leq 1} -\rho r + \sup_{x_o \in A'} -(1 + \rho) \frac{\ln \int f(y|0)^{\frac{\rho}{1+\rho}} f(y|x_o)^{\frac{1}{1+\rho}} dy}{b(x_o)} \\ &= \sup_{0 \leq \rho \leq 1} -\rho r + \tilde{E}_o(\rho). \end{aligned}$$

This finishes the proof of the random-coding bound.

Remarks: In this part of the proof, as compared with [3], the modification we made is first to choose an arbitrary $x_o \in A'$ to be the symbol used in the on-off signaling scheme, instead of choosing an optimal symbol as in [3], where the alphabet is fixed as p shrinks. Then we use a similar procedure, especially the key inequality (75), as in [3] to achieve the bound for this arbitrary on-off signaling. Then in the final part, we use the fact that x_o can be arbitrarily chosen from A' and get the same lower bound as the sphere-packing bound.

A.3 Straight-line bound for low-rate region

Next we show the straight-line bound. We start with the following expression for the straight-line bound, which is equivalent to the definition in Theorem 1:

$$E_{sl}(R, p) = \inf_{R' \geq R} \left(1 - \frac{R}{R'}\right) E(0, p) + \frac{R}{R'} E_{sp}(R', p). \quad (79)$$

What we need to show here is the following

$$E^{lp}(r) \leq \inf_{r' \geq r} (1 - \frac{r}{r'}) E_{sl}^{lp}(0) + \frac{r}{r'} E_{sp}^{lp}(r'), \quad (80)$$

where $E_{sl}^{lp}(0)$ is defined by (14).

To see this, we first state the following lemma which is a consequence of Theorem 5 in [3]:

Lemma 4 *If the zero-error capacity is zero, then*

$$\limsup_{p \rightarrow 0} \frac{E(0, p)}{p} \leq \sup_q \frac{\mathcal{E}[-\ln \int f(y|x_1)^{\frac{1}{2}} f(y|x_2)^{\frac{1}{2}} dy]}{\mathcal{E}[b(x)]}, \quad (81)$$

and the sup is over all possible probability distributions with a discrete and finite alphabet set.

Proof: To prove this lemma, we just need to go through the proof of Theorem 5 in [3], keeping in mind that we are not fixing the alphabet. The achievability argument using the expurgated bound is not needed here since we only state our Lemma as an upper bound. For a given SNR level p , for any input distribution $q \in \mathcal{D}(p)$, it is easy to see that inequality (78) in [3] is always valid, where the sup is over probability distributions which have the same alphabet as q . We can further upper bound (78) by relaxing the alphabet constraint of the sup to be any discrete and finite alphabet as in (81). From here, (81) follows. \diamond

Note we define the right side of (81) to be $E_{sl}^{lp}(0)$ as in (14). Assuming for a given r , r^* is the optimizing $r' \geq r$ in the right side of (80), we have

$$\begin{aligned} E^{lp}(r) &\leq \limsup_{p \rightarrow 0} \frac{E_{sl}(rp, p)}{p} \\ &= \limsup_{p \rightarrow 0} \frac{\inf_{r' \geq r} (1 - \frac{r}{r'}) E(0, p) + \frac{r}{r'} E_{sp}(r'p, p)}{p} \\ &\leq \limsup_{p \rightarrow 0} \frac{(1 - \frac{r}{r^*}) E(0, p) + \frac{r}{r^*} E_{sp}(r^*p, p)}{p} \\ &\leq (1 - \frac{r}{r^*}) E_{sl}^{lp}(0) + \frac{r}{r^*} E_{sp}^{lp}(r^*) \\ &= \inf_{r' \geq r} (1 - \frac{r}{r'}) E_{sl}^{lp}(0) + \frac{r}{r'} E_{sp}^{lp}(r'), \end{aligned}$$

which completes the proof of Theorem 2.

B Proof of Lemma 1

We first show

$$\frac{\det(\mathbf{I} + \mathbf{A} + \mathbf{B})}{\det(\mathbf{I} + \mathbf{A}) \det(\mathbf{I} + \mathbf{B})} = \frac{\det(\mathbf{I} + (\mathbf{I} + \mathbf{A})^{-1} \mathbf{B})}{\det(\mathbf{I} + \mathbf{B})}$$

$$\begin{aligned}
&= \frac{\det(\mathbf{I} + \mathbf{B} - (\mathbf{I} - (\mathbf{I} + \mathbf{A})^{-1})\mathbf{B})}{\det(\mathbf{I} + \mathbf{B})} \\
&= \det(\mathbf{I} - (\mathbf{I} - (\mathbf{I} + \mathbf{A})^{-1})\mathbf{B}(\mathbf{I} + \mathbf{B})^{-1}) \\
&= \det(\mathbf{I} - \mathbf{A}(\mathbf{I} + \mathbf{A})^{-1}\mathbf{B}(\mathbf{I} + \mathbf{B})^{-1}).
\end{aligned}$$

Now it suffices to show that $\sigma(\mathbf{A}(\mathbf{I} + \mathbf{A})^{-1}\mathbf{B}(\mathbf{I} + \mathbf{B})^{-1}) \subset [0, 1]$, i.e., all the eigenvalues of the matrix $\mathbf{A}(\mathbf{I} + \mathbf{A})^{-1}\mathbf{B}(\mathbf{I} + \mathbf{B})^{-1}$ lie in the interval $[0, 1]$. To show this, we have to use the following property of the eigenvalues of a product matrix, quoted from [6] (Corollary 1.7.7, Page 67), which we state as a theorem here.

Theorem 8 *Let $\mathbf{C}, \mathbf{D} \in \mathcal{C}^{M \times M}$. If \mathbf{C} is positive semi definite, then*

$$\sigma(\mathbf{CD}) \subset F(\mathbf{C})F(\mathbf{D}), \quad (82)$$

where $F(\mathbf{C})$ is the field of values of \mathbf{C} and is defined as follows

$$F(\mathbf{C}) = \{\mathbf{x}^\dagger \mathbf{C} \mathbf{x} : \mathbf{x} \in \mathcal{C}^{M \times 1}, \|\mathbf{x}\|^2 = 1\}. \quad (83)$$

◇

Consider $\mathbf{C} = \mathbf{A}(\mathbf{I} + \mathbf{A})^{-1}$ and $\mathbf{D} = \mathbf{B}(\mathbf{I} + \mathbf{B})^{-1}$. Now we claim that both \mathbf{C} and \mathbf{D} are positive semi definite matrices. To see this, we start with the singular value decomposition [5] of $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger$ and we write \mathbf{C} as

$$\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger(\mathbf{I} + \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger)^{-1} = \mathbf{U}\mathbf{\Lambda}(\mathbf{I} + \mathbf{\Lambda})^{-1}\mathbf{U}^\dagger.$$

Assume $\sigma(\mathbf{A}) = \{\lambda_1, \lambda_2, \dots, \lambda_M\}$. It is easy to see that

$$\sigma(\mathbf{C}) = \left\{ \frac{\lambda_1}{1 + \lambda_1}, \frac{\lambda_2}{1 + \lambda_2}, \dots, \frac{\lambda_M}{1 + \lambda_M} \right\}.$$

Since $\lambda_l \geq 0$ for any l , we have $\mathbf{C} \geq 0$ and further $\sigma(\mathbf{C}) \subset [0, 1]$. Similarly, we can show $\mathbf{D} \geq 0$ and $\sigma(\mathbf{D}) \subset [0, 1]$.

Thus, we can apply Theorem 8 here and we have

$$\sigma(\mathbf{CD}) \subset F(\mathbf{C})F(\mathbf{D}).$$

Further, it is easy to see that $F(\mathbf{C}) \subset [0, 1]$ since

$$0 \leq \mathbf{x}^\dagger \mathbf{C} \mathbf{x} \leq \lambda_{max}(\mathbf{C})\|\mathbf{x}\|^2 \leq 1.$$

Similarly, $F(\mathbf{D}) \subset [0, 1]$. Hence, we must have

$$\sigma(\mathbf{CD}) = \sigma(\mathbf{A}(\mathbf{I} + \mathbf{A})^{-1}\mathbf{B}(\mathbf{I} + \mathbf{B})^{-1}) \subset [0, 1],$$

which completes the proof of Lemma 1.

References

- [1] C. Chuah, D. N. Tse, J. M. Kahn and R. A. Valenzuela. Capacity scaling in MIMO wireless systems under correlated fading. *IEEE Trans. Inform. Theory*, vol. 48, pp. 637–650, March 2003.
- [2] R. G. Gallager. *Information Theory and Reliable Communication*. John Wiley and Sons, New York, NY, 1968.
- [3] R. G. Gallager. Energy limited channels: Coding, multiaccess, and spread spectrum, November 1987. Tech. Report LIDS-P-1714, LIDS, MIT, Cambridge, Mass. A later version appeared as a book chapter in *Codes, Graphs, and Systems: A Celebration of the Life and Career of G. David Forney, Jr. on the Occasion of His Sixtieth Birthday (Kluwer International Series in Engineering and Computer Science, 670)*.
- [4] B. Hajek and V. G. Subramanian. Capacity and reliability function for small peak signal constraints. *IEEE Transactions on Information Theory*, 48(4):828–839, April 2002.
- [5] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press 1985.
- [6] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press 1991.
- [7] R. S. Kennedy. *Fading Dispersive Communication Channels*. New York: Wiley-Interscience, 1969.
- [8] A. Lapidoth and S.M. Moser. Capacity bounds via duality with applications to multi-antenna systems on flat fading channels. *IEEE Transactions on Information Theory*, 49(10):2426–2467, Oct. 2003.
- [9] A. Lozano, A. M. Tulino and S. Verdu. Multiple-antenna capacity in the low-power regime. *IEEE Transactions on Information Theory*, 49(10):2527–2544, Oct. 2003.
- [10] M. Medard and R. G. Gallager. Bandwidth scaling for fading multipath channels. *IEEE Transactions on Information Theory*, 48(4):840–853, April 2002.
- [11] F. Neeser and J. L. Massey. Proper complex random processes with applications to information theory. *IEEE Trans. Inform. Theory*, vol. 39, pp. 1293–1302, July 1993.
- [12] S. Ray, M. Mdard and L. Zheng. On Error Probability for Wideband MIMO Channels. *Proceedings of Conference on Information Sciences and Systems (CISS)*, Baltimore, MD, Mar. 16-18, 2005.
- [13] A. M. Sayeed. Deconstructing Multi-Antenna Fading Channels. *IEEE Transactions on Signal Processing (special issue on MIMO systems)*, pp. 2563-2579, Oct. 2002.

- [14] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp. Lower bounds to error probability for coding on discrete memoryless channels. II. *Information and Control*, 10:522–552, 1967.
- [15] V. G. Subramanian. *Broadband Fading Channels: Signal Burstiness and Capacity*, Ph.D. thesis, University of Illinois, 1999.
- [16] V. G. Subramanian. and B. Hajek. Broad-band fading channels: Signal burstiness and capacity. *IEEE Transactions on Information Theory*, 48(6):809–827, Apr. 2002.
- [17] V. Tarokh, H. Jafarkhani, and A. R. Calderbank. Space-time block code from orthogonal designs. *IEEE Trans. Info. Theory*, vol. 45, pp. 1456–67, July 1999.
- [18] I. E. Telatar. Capacity of multi-antenna Gaussian channels. *European Trans. On Telecommunications*, vol. 10, pp. 585-95, Nov/Dec 1999.
- [19] E. Telatar and D. Tse. Capacity and mutual information of wideband multipath fading channels. *IEEE Transaction on Information Theory*, 46(7):1384-1400, 2000.
- [20] D. N. Tse and P. Viswanath. *Fundamentals of Wireless Communications*. Preprint.
- [21] S. Verdú. On channel capacity per unit cost. *IEEE Transactions on Information Theory*, 36(9):1019–1030, Sept. 1990.
- [22] S. Verdú. Spectral efficiency in the wideband regime. *IEEE Transactions on Information Theory*, 48(6):1319–1343, June 2002.
- [23] X. Wu. *Wireless Communications in the Energy-limited Regime*. Ph.D. Thesis, University of Illinois at Urbana-Champaign, Dec. 2004
- [24] X. Wu and R. Srikant. Asymptotic behavior of error exponents in the wideband regime. To appear in the *IEEE Transactions on Information Theory*, 2007.
- [25] L. Zheng and D. N. Tse. Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels. *IEEE Transactions on Information Theory*, 49(5):1073–96 May, 2003.
- [26] L. Zheng and D. N. Tse. The diversity-multiplexing tradeoff for non-coherent multiple antenna channels. *Proc. Allerton Conference*, 2002.