Limiting Performance of Block-Fading Channels with Multiple Antennas

Ezio Biglieri[1] Giuseppe Caire[2] Giorgio Taricco[1]

[1] Politecnico di Torino, Corso Duca degli Abruzzi 24, I-10129 Torino (Italy) • E-mail: `<name>@polito.it`

[2] Institut Eurécom, 2229 Rue des Crêtes, F-06904 Sophia-Antipolis (France)

E-mail: `giuseppe.caire@eurecom.fr` *

February 24, 2000

Abstract

We derive the performance limits of a radio system consisting of a transmitter with t antennas and a receiver with r antennas, a block-fading channel with additive white Gaussian noise, delay and transmit-power constraints, and perfect channel-state information available at both transmitter and receiver. Because of a delay constraint, the transmission of a code word is assumed to span a finite (and typically small) number M of independent channel realizations; therefore, the relevant performance limits are the information outage probability and the "delay-limited" (or "non-ergodic") capacity [11, 16, 35].

We derive the coding scheme that minimizes the information outage probability. This scheme can be interpreted as the concatenation of an optimal code for the AWGN channel without fading to an optimal beamformer. For this optimal scheme we evaluate minimum-outage-probability and delay-limited capacity. Among other results, we prove that, for the fairly general class of regular *fading channels, the asymptotic delay-limited capacity slope, expressed in bit/s/Hz per dB of transmit SNR, is proportional to $\min(t, r)$ and independent of the number of fading blocks M. Since M is a measure of the time diversity (induced by interleaving) or of the frequency diversity of the system, this result shows that, if channel-state information is available also to the transmitter, very high rates with asymptotically small error probabilities are achievable without need of deep interleaving or high frequency diversity. Moreover, for a large number of antennas the delay-limited capacity approaches the ergodic capacity.*

Keywords: Space-time coding, fading channels, beamforming, channel capacity.

*The work of EB and GT was supported by the Italian Space Agency (ASI).

1 Introduction

Recent work (see, e.g., [10, 11, 35]) has explored the ultimate performance limits of multiple-antenna systems in a fading environment. It has been shown that, in a system with t transmit and r receive antennas and a slow fading channel modeled by an $t \times r$ matrix with random i.i.d. complex Gaussian entries (the "independent Rayleigh fading" assumption), the average channel capacity with perfect channel-state information (CSI) at the receiver is about $m \triangleq \min\{t, r\}$ times larger than that of a single-antenna system for the same transmitted power and bandwidth. The capacity increases by about m bit / s / Hz for every 3-dB increase in signal-to-noise ratio (SNR).

In this paper we extend the previous work to a block-fading additive white Gaussian noise (BF-AWGN) channel with transmit-power constraint and perfect CSI available at the transmitter and at the receiver[1] (see [1] for a comprehensive review of information-theoretic issues on fading channels and a rich list of references).

The block-fading model applies to a channel in which several adjacent symbols (referred to in the sequel as a *block*) are affected by the same fading value. For example, this model is applicable to an indoor wireless data network or a personal communication system with mobile terminals moving at walking speed, so that the channel gain, albeit random, varies so slowly with time that it can be assumed as constant along a block (see also [7, 10, 11, 26, 33, 35]). More generally, fading blocks can be thought of as separated in time (e.g., in a time-division system [25]), as separated in frequency (e.g., in a multicarrier system), or as separated both in time and in frequency (e.g., with slow time-frequency hopping [4, 18, 19]). With this model, even though very long code words are transmitted, perfect interleaving cannot be achieved because of delay limitations. In particular, following [5, 6, 25], we assume that a code word spans a number M of fading blocks. As explained in [25], M can be regarded as a measure of the interleaving delay of the system, so that systems subject to a strict delay constraint are characterized by a fixed (and usually small) value of M. On the other hand, a large number N of channel symbols can be transmitted simultaneously from each of the t transmitting antennas during each fading block, so that the assumption $N \to \infty$ is justified.[2]

[1] Hereafter we write CSIT and CSIR to denote the availability of perfect channel-state information at the transmitter and at the receiver, respectively.

[2] $M = 2$ in the IS-54 standard. $M = 4$ in the half-rate GSM standard, and $M = 8$ in its full-rate version. In all

Throughout the paper, we assume that fading blocks are *statistically independent* unless otherwise explicitly stated.

With block fading, the average mutual information over the ensemble of channel realizations cannot characterize the achievable transmission rates: in fact, a BF-AWGN channel spanning a finite number of fading blocks is not *information stable*, and achievable rates should be characterized through the exceedingly general non-ergodic approach of Verdú and Han [38]. Since the *instantaneous mutual information* of the M-block channel is a random variable, we define the *information outage probability* as the probability that this mutual information is lower than the rate of the code used for transmission [25]. This outage probability is closely related to the code word error probability, as averaged over the random coding ensemble and over all channel realizations; hence, it provides useful insight on the performance of a delay-limited coded system [4, 18, 19, 20, 21]. An additional important definition related to outage probability is that of *delay-limited capacity*, sometimes also referred to as *zero-outage capacity*. This is the maximum rate for which the minimum outage probability is zero for a given power constraint [5, 6, 16].[3]

The capacity of block-fading channels is best explained in the framework of *capacity versus outage* presented in [1], where the block-fading channel is modeled as a *compound channel* [8], whose transition probability depends on a random parameter $\theta \in \Theta$, with given probability distribution. Every rate R is associated to the largest set $\Theta_R \subseteq \Theta$ such that, for all $\theta \in \Theta_R$, the capacity C_θ of the channel for given θ satisfies $C_\theta \geq R$. Accordingly, the outage probability corresponding to R is

$$P_{\text{out}}(R) = \mathbb{P}(\theta \notin \Theta_R) = \mathbb{P}(C_\theta < R)$$

and the supremum of the rates, corresponding to zero outage probability, is the *delay-limited capacity*. In our setting, the role of the channel parameter θ is played by a sequence of M random $r \times t$ matrices, describing the fading gains of the multiple antenna channel during the

systems the number of channel symbols in each block exceeds 100 [27].

[3]One should observe that both outage probability and delay-limited capacity are defined for $N \to \infty$. Seemingly, this invalidates our assumption of delay-constrained transmission. However, these quantities should be regarded as useful mathematical abstractions (as well as other classical assumptions like *infinite interleaving*), as they predict accurately the behavior of practical codes for moderately large number of symbols per fading block. For example, for $N \approx 100$ the outage probability predicts surprisingly well the word error probability of good practical codes [18, 19].

M transmission blocks.

As mentioned before, one basic assumption in this paper is that the transmitter knows, before transmitting a code word, the channel gains over all M blocks in which the code word is split. To assure causality, this assumption is valid when the block-fading model is applied to a multicarrier transmission scheme in which the available frequency band (over which the fading is selective) is split into M subbands, as with orthogonal frequency-division multiplexing (OFDM). The subbands are so narrow that fading is frequency-flat in each of them, and they are transmitted simultaneously, via orthogonal subcarriers.[4] When causality prevents to obtain the exact CSIT our results represent an upper bound to achievable performance.

We have already commented upon the attention that multiple-antenna transmission / reception has received of late. Several information-theoretic analyses have been produced: for example, [23] contains an analysis of several practical transmission schemes with multiple antennas, while [35] determines the ergodic capacity of a multiple-antenna channel with CSIR only and independent Rayleigh fading by using the distribution of the eigenvalues of a Wishart matrix [9]. Refs. [10, 11, 12, 35] investigate the outage probability of a single-block BF-AWGN channel with CSIR. Papers [7] and [26] study the performance of a single-block BF-AWGN channel with deterministic frequency-selective fading and when the number of transmit / receive antennas approaches infinity (with and without CSIT), respectively. Both [7] and [26] resort to parallel channel decomposition and standard water-filling in order to maximize the mutual information. Code designs have also been proposed for multiple-antenna transmission (see, e.g., [15, 30, 31, 32, 33, 34]). The effect of correlation among antennas is studied in [2, 28]. Finally, in [24], the ergodic capacity of a block-fading channel with multiple antennas in the absence of CSIT and CSIR is investigated: upper and lower bounds to the capacity are obtained as well as the general form of the capacity achieving signal set.

1.1 In this paper ...

In this paper we solve the problem of minimizing the outage probability at a given *fixed* code rate. Given an M-block BF-AWGN channel with t transmit and r receive antennas (hereafter

[4]From a practical point of view, the transmitter can obtain the CSI either by a dedicated feedback channel (some existing systems already implement a fast power-control feedback channel [27]) or by time-division duplex [40], where the uplink and the downlink time-share the same M subchannels and the fading gains can be estimated from the incoming signal.

referred to as the M-block $t \times r$ BF-AWGN channel), our goal is to find the transmission scheme that minimizes the outage probability under the constraint that the average transmitted power (to be defined properly) shall not exceed a given threshold. The optimal transmission scheme consists of a standard ("Gaussian") code for the AWGN channel, followed by a suitable beamformer (described by a $t \times t$ matrix) which may vary from block to block. These beamforming matrices can be explicitly evaluated once the fading-gain matrices are known. With this approach, the problems of coding and beamforming are decoupled, and no special space-coding design is needed to minimize the outage probability: this stands in contrast to the case of no CSIT, where specific space-code constructions prove to be useful [33]. We prove that, for a fairly general class of fading channels, the asymptotic (in the m) delay-limited capacity grows linearly with m for a fixed transmit SNR independently of the number of fading blocks M. Since M is a measure of the interleaving delay of the system, this result shows that, when CSI is made available to the transmitter, very high rates with asymptotically small error probability are achievable without the need of deep interleaving or of large frequency diversity.

The results presented here complement our previous work [5, 6], where we examined the case of optimal power allocation for a single transmit / receive antenna system.

This paper is organized as follows. Section 2 describes the channel model. Section 3 is a quick *tour d'horizon* on channel-capacity results when there are no delay constraints. Delay constraints are introduced in Section 4. The minimum outage probability is derived in Section 5. Finally, numerical results elucidating the theory are shown and discussed in Section 6, while Section 7 concludes the paper by summarizing our findings.

2 Channel model

We consider the M-block $t \times r$ BF-AWGN channel. We assume a coding scheme whereby every code word contains MNt complex symbols, and is transmitted by dividing it into M blocks of Nt symbols each. The Nt symbols in each block are further grouped into t sub-blocks of N symbols each (see Fig. 1). Finally, the t sub-blocks are simultaneously transmitted by the t different antenna: under the assumption of an ideal Nyquist bandlimited modulation scheme, the spectral efficiency achieved is t symbol / s / Hz. This coding scheme is called a *space-time code* in [33]. Each one of the r receiving antennas observes the superposition of the t symbols

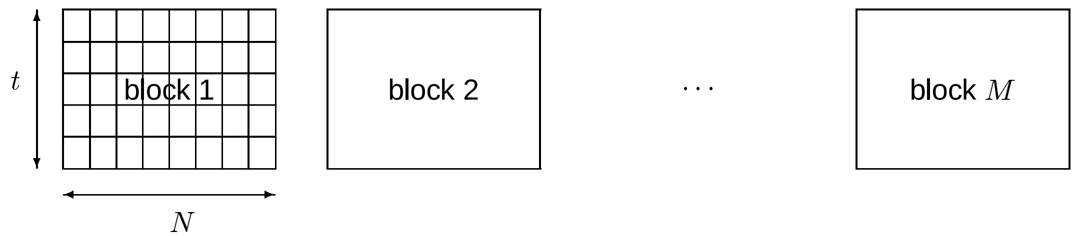

Figure 1: One code word in an M-block fading channel.

transmitted, corrupted by the background noise, the fading (in the form of complex channel gains), and the linear distortion introduced by the propagation channel. The outputs of these r antennas are processed jointly.

Under this BF-AWGN model, the discrete-time baseband equivalent channel can be modeled by

$$\mathbf{y}_k[n] = \mathbf{A}_k \mathbf{x}_k[n] + \mathbf{z}_k[n] \tag{1}$$

for $k = 1, \ldots, M$ (block index) and $n = 1, \ldots, N$ (symbol index along a block). $\mathbf{A}_k \in \mathbb{C}^{r\times t}$ are the *i.i.d.* matrices of complex channel gains (constant along each block under our block-fading assumption).[5] $\mathbf{y}_k[n] \in \mathbb{C}^r$, $\mathbf{x}_k[n] \in \mathbb{C}^t$, $\mathbf{z}_k[n] \in \mathbb{C}^r$ are the received, transmitted, and noise vectors, respectively. The noise is circularly-symmetric complex Gaussian with variance 1 in all of its components: we write $\mathcal{N}_c(\mathbf{0}, \mathbf{I})$ to indicate this, and we observe that, with this normalization of the noise variance, signal power and signal-to-noise ratio coincide. [6]

Input constraints. We supplement the channel model above with input constraints. For the sake of clarity, we illustrate them with reference to an M-subcarrier OFDM system. Here the total transmit power is $(1/M)\sum_{k=1}^{M} \mathcal{E}_k$, where $\mathcal{E}_k$ is the average transmit energy per s/Hz on the kth subcarrier, and the available system bandwidth is normalized to 1. Under our assumption that the noise has unit power spectral density, the SNR in the kth subcarrier is $\mathcal{E}_k$, and the average SNR is $(1/M)\sum_{k=1}^{M} \mathcal{E}_k$. This is proportional to the total transmit power. Therefore, we can consider input constraints in terms of SNR per subcarrier (or, within the most

[5]We assume that the joint cdf of the real and imaginary parts of their components is a continuous function. Thus, the matrix $\mathbf{A}_k\mathbf{A}_k^\dagger$ has rank m and hence exactly m positive eigenvalues with probability 1.

[6]We denote the distribution of a jointly Gaussian complex random column vector $\mathbf{z}$ by $\mathcal{N}_c(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} \triangleq \mathbb{E}[\mathbf{z}]$ and $\boldsymbol{\Sigma} \triangleq \mathbb{E}[(\mathbf{z}-\boldsymbol{\mu})(\mathbf{z}-\boldsymbol{\mu})^\dagger]$ where † denotes Hermitian conjugation.

general channel model, *per block*). In particular, the total transmit energy per s/Hz in the kth subcarrier is obtained as the sum of the energies output by the t transmit antennas, i.e., $\mathcal{E}_k = \mathbb{E}[\mathbf{x}_k^\dagger[n]\mathbf{x}_k[n]] = \mathrm{Tr}(\mathbb{E}[\mathbf{x}_k[n]\mathbf{x}_k^\dagger[n])$. Thus, thanks to our normalization of the noise power spectral density, transmit-energy constraints translate immediately into SNR constraints. In summary, we express our constraints in terms of the average trace of the input covariance matrices, $(1/M)\sum_{k=1}^{M}\mathrm{Tr}(\mathbb{E}[\mathbf{x}_k[n]\mathbf{x}_k^\dagger[n])$.

An equivalent channel model. Following a standard approach, we generate an equivalent, more convenient model for (1) by using the singular-value decomposition [17]

$$\mathbf{A}_k = \mathbf{U}_k\mathbf{S}_k\mathbf{V}_k^\dagger$$

where $\mathbf{U}_k \in \mathbb{C}^{r\times r}$ and $\mathbf{V}_k \in \mathbb{C}^{t\times t}$ are unitary, and $\mathbf{S}_k \in \mathbb{C}^{r\times t}$ is a diagonal matrix whose main-diagonal elements are the "singular values" $\lambda_{k,1}^{1/2} \geq \ldots \geq \lambda_{k,m}^{1/2}$, with $\lambda_{k,i}$ the i-th largest eigenvalue of the non-negative definite Hermitian matrix $\mathbf{A}_k\mathbf{A}_k^\dagger$. By defining $\widetilde{\mathbf{y}}[n] \triangleq \mathbf{U}_k^\dagger\mathbf{y}_k[n]$, $\widetilde{\mathbf{z}}_k[n] \triangleq \mathbf{U}_k^\dagger\mathbf{z}_k[n]$, and $\widetilde{\mathbf{x}}_k[n] \triangleq \mathbf{V}_k^\dagger\mathbf{x}_k[n]$, we can rewrite the channel input-output relation (1) in the form

$$\widetilde{\mathbf{y}}_k[n] = \mathbf{S}_k\widetilde{\mathbf{x}}_k[n] + \widetilde{\mathbf{z}}_k[n] \tag{2}$$

This is equivalent to (1): in fact, $\widetilde{\mathbf{z}}_k[n] \sim \mathcal{N}_c(\mathbf{0},\mathbf{I})$ and $\mathrm{Tr}(\mathbb{E}[\widetilde{\mathbf{x}}_k[n]\widetilde{\mathbf{x}}_k^\dagger[n]]) = \mathrm{Tr}(\mathbb{E}[\mathbf{x}_k[n]\mathbf{x}_k^\dagger[n]])$[7]. Thus, any input constraint on $\widetilde{\mathbf{x}}_k[n]$ directly translates into a constraint on the original channel input $\mathbf{x}_k[n]$.

3 No delay constraints: Ergodic capacity

We examine first the situation in which there are no delay constraints, so that M is allowed to increase without bound. The channel here is BF-AWGN with block length $N < \infty$. Since the matrix process $\{\mathbf{A}_k\}_{k=1}^{M}$ describing the multiple antenna channel is (blockwise) i.i.d., the channel is information stable [38] as $M \to \infty$ and capacity concides with the maximum average mutual information. Actually, the assumption of i.i.d. matrices is not necessary for information stability, and the results of Proposition 1 below holds also for more general ergodic but correlated matrix processes. With CSIR, the capacity of this channel does not depend on the value of N [22], and the following result holds [35]:

[7] $\mathrm{Tr}(\mathbf{A}) \triangleq \sum_i A_{ii}$ denotes the *trace* of the matrix $\mathbf{A}$

Proposition 1. *Under the input-power constraint Tr*$(\mathbb{E}[\mathbf{x}[n]\mathbf{x}^\dagger[n]]) \leq \gamma$, *the capacity of the* $t \times r$ *BF-AWGN channel is given by:*

1. *With CSIR and no CSIT,*
$$C_{\mathrm{CSIR}}(\gamma) = \sum_{i=1}^{m} \mathbb{E}[\log_2(1+\gamma\lambda_i/t)] \tag{3}$$
2. *With CSIR and CSIT,*
$$C_{\mathrm{CSIR,CSIT}}(\gamma) = \sum_{i=1}^{m} \mathbb{E}[[\log_2(\xi\lambda_i)]_+] \tag{4}$$
where $\xi = \xi(\gamma)$ *is the solution to*
$$\mathbb{E}\left[\sum_{i=1}^{m}(\xi - 1/\lambda_i)_+\right] = \gamma, \tag{5}$$
$m \triangleq \min\{r,t\}$, $[a]_+ \triangleq \max\{0,a\}$, $\lambda_1,\ldots,\lambda_m$ *are the nonzero eigenvalues of* $\mathbf{A}\mathbf{A}^\dagger$, *and* $\mathbf{A}$ *is a random matrix distributed as any one of the* $\mathbf{A}_k$*'s.*

♦

Remark 1. For all block lengths $N = 1, 2, \ldots$, capacities (3) and (4) are achieved by sequences of codes with length MNt with $M \to \infty$. Capacity (3) can be achieved by random codes whose symbols are independent and have a circularly-symmetric complex Gaussian distribution $\mathcal{N}_c(0,\gamma/t)$. This implies that all antennas transmit the same average energy per symbol. Capacity (4) can be achieved by generating random codes with i.i.d. components $\sim \mathcal{N}_c(0,1)$ and having each code word split into M blocks of N vectors $\widetilde{\mathbf{x}}_k[n]$ with t components each. For block k, the optimal linear transformation
$$\mathbf{W}_k = \mathbf{V}_k \mathrm{diag}(\sqrt{\gamma_{k,1}},\ldots,\sqrt{\gamma_{k,m}},\underbrace{0,\ldots,0}_{t-m}) \tag{6}$$
is computed, where $\gamma_{k,i} \triangleq [\xi - 1/\lambda_{k,i}]_+$. Finally, the vectors $\mathbf{x}_k[n] = \mathbf{W}_k\widetilde{\mathbf{x}}_k[n]$ are transmitted from the t antennas. This optimal scheme can be seen as the concatenation of an optimal encoder for the unfaded AWGN channel, followed by an optimal beamformer with weighting matrix $\mathbf{W}_k$ varying from block to block. This is a rather attractive scheme for coding, because capacity can be approached by constructing a single "Gaussian" code book, which depends on both the fading statistics and on the number of transmit/receive antennas only through its coding rate. In this sense, we may say that optimal beamforming "achieves capacity." ♦

4 Performance under delay constraints

We now assume a finite value for M, which translates into a delay constraint, and perfect CSI. We also assume transmission at a constant rate R. For any given M, we generate a sequence of BF-AWGN channels, indexed by their block length N, $N = 1, 2, \ldots$, and we study their limiting performance as $N \to \infty$. The coding theorem proved by the authors in [6, Proposition 2] applies verbatim to this vector channel.

Define the row vector $\boldsymbol{\lambda}_k \triangleq (\lambda_{k,1}, \ldots, \lambda_{k,m})$ and the $m \times M$ matrix $\boldsymbol{\Lambda} \triangleq (\boldsymbol{\lambda}_1^T, \ldots, \boldsymbol{\lambda}_M^T)$. Further, let $\boldsymbol{\gamma}_k \triangleq (\gamma_{k,1}, \ldots, \gamma_{k,m})$ be the row vector of the elements along the main diagonal of $\mathbb{E}[\widetilde{\mathbf{x}}_k[n]\widetilde{\mathbf{x}}_k^\dagger[n]]$, assumed to be independent of n,[8] and define the $t \times M$ matrix $\boldsymbol{\Gamma} \triangleq (\boldsymbol{\gamma}_1^T, \ldots, \boldsymbol{\gamma}_M^T)$ of individual transmit SNRs for the t antennas over the M blocks. We introduce the following:

Definition 1 (Instantaneous mutual information). *The maximum instantaneous mutual information $I_M(\boldsymbol{\Lambda}, \boldsymbol{\Gamma})$ of the M-block BF-AWGN vector channel with eigenvalues $\boldsymbol{\Lambda}$ and transmit SNRs $\boldsymbol{\Gamma}$ is defined as*

$$I_M(\boldsymbol{\Lambda}, \boldsymbol{\Gamma}) \triangleq \frac{1}{M} \sum_{k=1}^{M} \sum_{i=1}^{m} \log_2(1 + \lambda_{k,i}\gamma_{k,i}) \tag{7}$$

♦

The mutual information (7) is achieved by independent Gaussian-distributed input symbols $\mathbf{x}_k[n] \sim \mathcal{N}_c(\mathbf{0}, \mathbf{W}_k\mathbf{W}_k^\dagger)$, with $\mathbf{W}_k$ given by (6). Observe also that with CSIT the instantaneous SNRs $\boldsymbol{\Gamma}$ depend on $\boldsymbol{\Lambda}$ (to stress this, occasionally we shall write $\boldsymbol{\Gamma}(\boldsymbol{\Lambda})$).

We introduce a *short-term input constraint* by requiring the "instantaneous" SNR per block not to exceed a threshold γ. This is expressed by

$$P_M(\boldsymbol{\Gamma}) \triangleq \frac{1}{M} \sum_{k=1}^{M} \sum_{i=1}^{m} \gamma_{k,i} \leq \gamma \tag{8}$$

When we transmit a long sequence of code words, the transmitter may choose to allocate more power to the code words sent when the channel is bad, and less power to the codewords sent

[8] This does not entail any loss of generality. In fact, if the SNR depends on n, and hence we write $\gamma_{k,i}[n]$, we have, from Jensen's inequality,

$$\frac{1}{MN} \sum_{k=1}^{M} \sum_{i=1}^{m} \sum_{n=1}^{N} \log_2(1 + \lambda_{k,i}\gamma_{k,i}[n]) \leq \frac{1}{M} \sum_{k=1}^{M} \sum_{i=1}^{m} \log_2 \left(1 + \lambda_{k,i} \frac{1}{N} \sum_{n=1}^{N} \gamma_{k,i}[n]\right)$$

which shows that the optimum SNR distribution is uniform over the code word position within the same block and transmit-antenna signal.

when the channel is good. If this is the case, it makes sense to consider a *long-term input constraint*, as a constraint on the *average* SNR per block over a long sequence of code words. Assuming ergodicity, we can express the long-term input constraint as [6]

$$\mathbb{E}[P_M(\mathbf{\Gamma})] \leq \gamma \tag{9}$$

(For later use, notice that the long-term constraint is weaker than its short-term counterpart: in fact, (8) implies (9).)

For a given function $\mathbf{\Gamma}(\mathbf{\Lambda})$ satisfying (8) or (9), $I_M(\mathbf{\Lambda}, \mathbf{\Gamma})$ is a random variable. We have the following (see [1, 6, 16, 25] and references therein):

Definition 2 (Information outage probability). Let $\mathbf{\Gamma}$ satisfy (8) or (9). The information outage probability when the code rate is R is defined as:

$$P_{\text{out}}(R, \gamma) \triangleq \mathbb{P}(I_M(\mathbf{\Lambda}, \mathbf{\Gamma}) < R) \tag{10}$$

♦

In the following, we are especially concerned with the minimization of $P_{\text{out}}(R, \gamma)$ with respect to the choice of the SNR allocation function $\mathbf{\Gamma}$ with CSIR and CSIT[9]. This minimum outage probability is intimately related to the capacity of the M-block BF-AWGN vector channel: in order to distinguish the channel capacity without delay constraints from the capacity of the M-block channel, the latter is referred to as the "delay-limited" capacity [16]. We have:

Definition 3 (Delay-limited capacity). *The delay-limited capacity of the M-block BF-AWGN vector channel, subject to a short-term (resp., long-term) input constraint, is given by*

$$C_{\text{delay}}(\gamma) \triangleq \sup_{\mathbf{\Gamma}(\mathbf{\Lambda})} \inf_{\mathbf{\Lambda}} I_M(\mathbf{\Lambda}, \mathbf{\Gamma}) \tag{11}$$

where the supremum is over all $\mathbf{\Gamma}(\mathbf{\Lambda})$ satisfying (8) (resp., (9)), and the infimum is over all the non-negative $\mathbf{\Lambda}$. ♦

The next proposition, which follows immediately from [6, Proposition 2], shows a definition of delay-limited capacity equivalent to the above. Its proof (that we skip for brevity's sake) is based on the observation that the equivalent vector channel (2) can be seen as a scalar channel with Mm blocks.

[9] A conjecture on the outage probability minimization for perfect CSIR and no CSIT is made in [35].

Proposition 2. *Define the maximum ϵ-achievable rate of the M-block BF-AWGN vector channel subject to a short-term (resp., long-term) input constraint as*

$$C_\epsilon(\gamma) \triangleq \sup_{\mathbf{\Gamma}(\mathbf{\Lambda})} \sup\{R : P_{\text{out}}(R, \gamma) \leq \epsilon\} \tag{12}$$

where the supremum is over all $\mathbf{\Gamma}(\mathbf{\Lambda})$ satisfying (8) (resp., (9)). Then the delay-limited capacity (11) is given by

$$C_{\text{delay}}(\gamma) = \lim_{\epsilon \downarrow 0} C_\epsilon(\gamma)$$

♦

Remark 2. As in the ergodic capacity case, minimum outage probability and delay-limited capacity can be achieved by concatenating "Gaussian" codes independent of the fading statistics with an optimal beamformer varying from block to block. We hasten to observe that the optimal beamforming matrices $\mathbf{W}_k$ are generally not the same for the ergodic and for the delay-limited case, since they correspond to different SNRs $\gamma_{k,i}$. ♦

5 Minimum outage probability

Since the vector channel (2) is equivalent to a scalar channel with Mm blocks and fading power gains $\lambda_{k,i}$, Propositions 3 and 4 of [6] apply almost verbatim. For the sake of completeness, we restate them without proof in the notation pertaining to the vector BF-AWGN channel.

Short-term problem. The short-term constrained minimization problem can be formulated as

$$\begin{cases} \text{Minimize} & \mathbb{P}(I_M(\mathbf{\Lambda}, \mathbf{\Gamma}) < R) \\ \text{Subject to} & P_M(\mathbf{\Gamma}) \leq \gamma \end{cases} \tag{13}$$

It should be rather intuitive that the solution to the above problem must maximize $I_M(\mathbf{\Lambda}, \mathbf{\Gamma})$ for each $\mathbf{\Lambda}$ (or at least, for each $\mathbf{\Lambda}$ in a certain subset of $\mathbb{R}_+^{Mm}$), so that (13) can be reduced to a mutual-information maximization problem, whose solution is well-known [13]. Formally, we have the following:

Proposition 3. *Problem (13) is solved by*

$$\widehat{\mathbf{\Gamma}}(\mathbf{\Lambda}) = \begin{cases} \mathbf{\Gamma}^{\rm st}(\mathbf{\Lambda},\gamma) & \text{if } \mathbf{\Lambda} \in \mathcal{R}_{\rm on}(R,\gamma) \\ \mathbf{G}(\mathbf{\Lambda}) & \text{if } \mathbf{\Lambda} \in \mathcal{R}_{\rm off}(R,\gamma) \end{cases} \tag{14}$$

where

1. *The* (k,i)*-th component of* $\mathbf{\Gamma}^{\rm st}(\mathbf{\Lambda},\gamma)$ *is given by*

   $$\gamma_{k,i}^{\rm st} = \left[\xi^{\rm st}(\mathbf{\Lambda},\gamma) - \frac{1}{\lambda_{k,i}}\right]_+ \tag{15}$$

   where

   $$\xi^{\rm st}(\mathbf{\Lambda},\gamma) = \frac{1}{|\mathcal{M}(\gamma)|}\sum_{(k,i)\in\mathcal{M}(\gamma)}\frac{1}{\lambda_{k,i}} + \frac{M}{|\mathcal{M}(\gamma)|}\gamma \tag{16}$$

   and $\mathcal{M}(\gamma)$ *is the unique set of indexes* (k,i) *such that* $1/\lambda_{k,i} \le \xi^{\rm st}(\mathbf{\Lambda},\gamma)$ *for all* $(k,i) \in \mathcal{M}(\gamma)$ *and* $1/\lambda_{k,i} > \xi^{\rm st}(\mathbf{\Lambda},\gamma)$ *for all* $(k,i) \notin \mathcal{M}(\gamma)$.

2. *The* outage *or* power-off region $\mathcal{R}_{\rm off}(R,\gamma)$ *is given by*

   $$\mathcal{R}_{\rm off}(R,\gamma) \triangleq \left\{\mathbf{\Lambda} : I_M(\mathbf{\Lambda},\mathbf{\Gamma}^{\rm st}(\mathbf{\Lambda},\gamma)) < R\right\} \tag{17}$$

   and the power-on region $\mathcal{R}_{\rm on}$ *is the complement of* $\mathcal{R}_{\rm off}$ *in* $\mathbb{R}_+^{Mm}$.

3. $\mathbf{G}(\mathbf{\Lambda})$ *is an arbitrary function* $\mathbb{R}_+^{Mm} \to \mathbb{R}_+^{Mt}$ *satisfying the short-term constraint* $P_M(\mathbf{G}) \le \gamma$.

♦

Remark 3. $\mathbf{\Gamma}^{\rm st}(\mathbf{\Lambda},\gamma)$ is the solution to the maximization problem

$$\begin{cases} \text{Maximize} & I_M(\mathbf{\Lambda},\mathbf{\Gamma}) \\ \text{Subject to} & P_M(\mathbf{\Gamma}) \le \gamma \end{cases} \tag{18}$$

Its solution is readily found by using Lagrange multipliers and the Kuhn-Tucker theorem [13]. Note that, since $\mathbf{G}(\mathbf{\Lambda})$ is arbitrary, the solution to (13) is in general not unique. However, since whenever $\mathbf{\Lambda} \in \mathcal{R}_{\rm off}(R,\gamma)$ transmission gives rise to an outage event, it is wise to set $\mathbf{G}(\mathbf{\Lambda}) = \mathbf{0}$.

Note that the definitions of $\xi^{\rm st}(\mathbf{\Lambda},\gamma)$ and $\mathcal{M}(\gamma)$ depend on each other. An algorithm for finding $\mathcal{M}(\gamma)$ is provided in [6]. ♦

Long-term problem. The long-term constrained minimization problem can be stated as

$$\begin{cases} \text{Minimize} & \mathbb{P}(I_M(\mathbf{\Lambda},\mathbf{\Gamma}) < R) \\ \text{Subject to} & \mathbb{E}[P_M(\mathbf{\Gamma})] \leq \gamma \end{cases} \tag{19}$$

The above problem is less standard because of the presence of the expectation in the constraint equation. The solution in the most general case for the joint probability distribution of $\mathbf{\Lambda}$ is given in [6] in terms of a random function, where randomization is required if the distribution of $\mathbf{\Lambda}$ has point masses. For simplicity, here we restrict our treatment to the case of continuous joint cumulative distribution functions $F(\mathbf{\Lambda})$. In this case we have:

Proposition 4. *Problem (19) is solved by*

$$\widehat{\mathbf{\Gamma}}(\mathbf{\Lambda}) = \begin{cases} \mathbf{\Gamma}^{\mathrm{lt}}(\mathbf{\Lambda}, R) & \textit{if } \mathbf{\Lambda} \in \mathcal{R}^*_{\mathrm{on}}(R,\gamma^*) \\ \mathbf{0} & \textit{if } \mathbf{\Lambda} \in \mathcal{R}^*_{\mathrm{off}}(R,\gamma^*) \end{cases} \tag{20}$$

where

1. *The* (k,i)*-th component of* $\mathbf{\Gamma}^{\mathrm{lt}}(\mathbf{\Lambda}, R)$ *is given by*

$$\gamma^{\mathrm{lt}}_{k,i} = \left[\xi^{\mathrm{lt}}(\mathbf{\Lambda}, R) - \frac{1}{\lambda_{k,i}}\right]_+ \tag{21}$$

where

$$\xi^{\mathrm{lt}}(\mathbf{\Lambda}, R) = \left(\frac{2^{MR}}{\prod_{(k,i)\in\mathcal{M}^*(R)} \lambda_{k,i}}\right)^{1/|\mathcal{M}^*(R)|} \tag{22}$$

and $\mathcal{M}^*(R)$ *is the unique set of indexes* (k,i) *such that* $1/\lambda_{k,i} \leq \xi^{\mathrm{lt}}(\mathbf{\Lambda}, R)$ *for all* $(k,i) \in \mathcal{M}^*(R)$ *and* $1/\lambda_{k,i} > \xi^{\mathrm{lt}}(\mathbf{\Lambda}, R)$ *for all* $(k,i) \notin \mathcal{M}^*(R)$*.*

2. *The outage region* $\mathcal{R}^*_{\mathrm{off}}(R,\gamma^*)$ *is given by*

$$\mathcal{R}^*_{\mathrm{off}}(R,\gamma^*) \triangleq \left\{\mathbf{\Lambda} : P_M(\mathbf{\Gamma}^{\mathrm{lt}}(\mathbf{\Lambda}, R)) > \gamma^*\right\} \tag{23}$$

The power-on region $\mathcal{R}^*_{\mathrm{on}}(R,\gamma^*)$ *is its complement in* $\mathbb{R}_+^{Mm}$*.*

3. *The* threshold $\gamma^* > 0$ *is set in order to satisfy the long-term contraint (9) with equality, i.e., it is the solution of*

$$\mathbb{E}[P_M(\mathbf{\Gamma}^{\mathrm{lt}}(\mathbf{\Lambda}, R))\, 1\{\mathbf{\Lambda} \in \mathcal{R}_{\mathrm{on}}(R,\gamma^*)\}] = \gamma$$

where $1\{\mathcal{A}\} \triangleq 1$ *if* $\mathcal{A}$ *is true and 0 otherwise.*

♦

Remark 4. Let us fix an ordering of $\mathbf{\Lambda}$, specified by the index set $(k[\ell], i[\ell])_{\ell=1}^{Mm}$ such that $(\lambda_{k[\ell],i[\ell]})_{\ell=1}^{Mm}$ is a nonincreasing sequence. Then, the index sets $\mathcal{M}(\gamma)$ and $\mathcal{M}^*(R)$ are determined by their own cardinality because their definition is based on a threshold. If $|\mathcal{M}(\gamma)| = \mu$ (or $|\mathcal{M}^*(R)| = \mu$) for some $\mu = 0, \ldots, Mm$, then $\mathcal{M} = \{(k[\ell], i[\ell])\}_{\ell=1}^{\mu}$ satisfies the definition of Proposition 3. ♦

Remark 5. $\mathbf{\Gamma}^{\rm lt}(\mathbf{\Lambda}, R)$ is the solution to the minimization problem, dual of (18)

$$\begin{cases} \text{Minimize} & P_M(\mathbf{\Gamma}) \\ \text{Subject to} & I_M(\mathbf{\Lambda}, \mathbf{\Gamma}) \geq R \end{cases} \tag{24}$$

Again, $\mathbf{\Gamma}^{\rm lt}(\mathbf{\Lambda}, R)$ is obtained by using Lagrange multipliers and the Kuhn-Tucker theorem [13].

Note that the definitions of $\xi^{\rm lt}(\mathbf{\Lambda}, R)$ and $\mathcal{M}^*(R)$ depend on each other. An algorithm for finding $\mathcal{M}^*(R)$ is provided in [6]. ♦

Remark 6. The optimum SNR allocation $\widehat{\mathbf{\Gamma}}(\mathbf{\Lambda})$ obtained from Proposition 4 corresponds to setting a threshold γ^* such that, if the SNR per block necessary to prevent outage exceeds γ^*, transmission is turned off, while if it is below γ^* transmission is turned on and the SNR is allocated according to $\mathbf{\Gamma}^{\rm lt}$. This is the allocation requiring the minimum SNR per block in order to avoid an outage event (see (24)). The threshold γ^* is chosen so that the long-term SNR per block is actually equal to γ. A remarkable fact is that the optimal power allocation rule of Proposition 4 depends on the fading statistics only through this threshold value γ^*. For unknown fading statistics, a given constraint γ and a given target rate R, γ^* can be estimated adaptively [6]. ♦

Remark 7. Although apparently different, the outage regions $\mathcal{R}_{\rm off}(R, \gamma)$ and $\mathcal{R}^*_{\rm off}(R, \gamma^*)$ defined in Propositions 3 and 4 are equivalent for the same values of the arguments, i.e., $\mathcal{R}^*_{\rm off}(R, s) = \mathcal{R}_{\rm off}(R, s)$ (as shown in Appendix A) so that the notation $\mathcal{R}^*$ is hereafter replaced by $\mathcal{R}$. Obviously, $\gamma^* \geq \gamma$, since, as observed before, the short-term constraint implies the long-term constraint. This leads to the important observation that choosing a long-term input constraint is tantamount to requiring a larger γ^* in lieu of γ, while turning off the transmission when an outage cannot be avoided. For both long-term and short-term constraints we have, formally,

$$P_{\rm out}(R, \gamma) = \mathbb{P}(\mathbf{\Lambda} \in \mathcal{R}_{\rm off}(R, s)) \tag{25}$$

for $s = \gamma$ (short-term) of $s = \gamma^*$ (long-term). In the *capacity versus outage* setting described in [1] based on the notion of compound channel of [8], $\mathbf{\Lambda}$ plays the role of the channel parameter θ and $\mathcal{R}_{\text{on}}(R, s)$ of the set Θ_R for which the rate R is below the conditional mutual information for given $\mathbf{\Lambda}$. ♦

Remark 8. It is not difficult to see [6, App. B.2] that $\mathcal{R}_{\text{off}}(R, s) \subseteq \mathbb{R}_+^{Mm}$ is a connected (although not necessarily convex) region containing the origin, and that it shrinks as s increases, so that

$$\lim_{s\to\infty} P_{\text{out}}(R, s) = \lim_{s\to\infty} \mathbb{P}(\mathbf{\Lambda} \in \mathcal{R}_{\text{off}}(R, s)) = 0$$

for all random $\mathbf{\Lambda}$ with continuous cdf. ♦

See Appendix A for an example of outage region corresponding to $m = 1$ and $M = 2$.

5.1 Delay-limited capacities: Their calculation

Most physical-channel models have a continuous joint distribution $F(\mathbf{\Lambda})$ whose support is the whole orthant $\mathbb{R}_+^{Mm}$. If this is true, under a short-term constraint $P_{\text{out}}(R, \gamma)$ is positive for all finite γ. This implies that under a short-term constraint $C_{\text{delay}}(\gamma)$ is identically zero. On the contrary, depending on the fading statistics, it may occur that under a long-term constraint $P_{\text{out}}(R, \gamma) = 0$ for all $\gamma \geq \gamma(R)$, the minimum average SNR per block given by

$$\gamma(R) = \mathbb{E}\left[P_M(\mathbf{\Gamma}^{\text{lt}}(\mathbf{\Lambda}, R))\right] \tag{26}$$

In this case, $R = C_{\text{delay}}(\gamma)$ is obtained by inverting the relation $\gamma = \gamma(R)$.

As for the calculation of delay-limited capacities, it is convenient to confine ourselves to the consideration of a "well-behaved" class of fading channels that we call *regular* according to the following definition:

Definition 4 (Regular fading). *An M-block vector BF-AWGN channel is said to be regular if the fading distribution is continuous, and*

$$\mathbb{E}[1/\bar{\lambda}_M] < \infty \tag{27}$$

where $\bar{\lambda}_M$ is the geometric mean of the $\lambda_{k,i}$, i.e., $\bar{\lambda}_M \triangleq \prod_{k,i} \lambda_{k,i}^{1/(Mm)}$. ♦

Based on the above definition, we have the following result.

Proposition 5. *The minimum SNR per block function $\gamma(R)$ of a* regular *BF-AWGN channel subject to a long-term constraint has the following properties:*

1. $\gamma(0) = 0$;
2. $\gamma(R)$ *is continuous and monotonically increasing;*
3. $\gamma(R)$ *satisfies the inequality*

$$\gamma(R) \leq m 2^{R/m} \mathbb{E}[1/\bar{\lambda}_M] < \infty$$

♦

Proof. Deferred to Appendix B.

The following Corollary derives from the above Proposition.

Corollary 1. *The delay-limited capacity of a* regular *BF-AWGN channel subject to a long-term constraint is positive for every $\gamma > 0$.* ♦

Proof. The delay-limited capacity is the solution of the equation $\gamma(R) = \gamma$. Therefore, from Proposition 5, this equation has a single positive solution whenever $\gamma > 0$. ■

Remark 9. The Rayleigh fading channel with $m = M = 1$ is not regular and its delay-limited capacity is null. In fact, in this case, $\mathbb{E}[1/\bar{\lambda}_1] = \int_0^\infty (e^{-x}/x)dx$ diverges. ♦

Remark 10. The $t \times r$ independent Rayleigh M-BF-AWGN channel is regular and, from Proposition 5, it has nonzero delay-limited capacity whenever $Mm > 1$ since $\mathbb{E}[1/\bar{\lambda}_M] < \infty$ in all cases. In fact:

- If $M > 1$ and $m = 1$, we have

$$\mathbb{E}[1/\bar{\lambda}_M] = \left\{\mathbb{E}\left[\lambda_1^{-1/M}\right]\right\}^M = \Gamma(1 - 1/M)^M < \infty$$

  where $\Gamma(x) \triangleq \int_0^\infty u^{x-1} e^{-u} du$.

- If $M \geq 1$ and $m > 1$, from [35], we have

$$\mathbb{E}[1/\bar{\lambda}_M] = \left\{ \int_0^\infty d\lambda_1 \int_0^{\lambda_1} d\lambda_2 \cdots \int_0^{\lambda_{m-1}} d\lambda_m \left[K_{m,n} \prod_{i=1}^{m} e^{-\lambda_i} \lambda_i^{n-m-1/(Mm)} \prod_{i<j} (\lambda_i - \lambda_j)^2 \right] d\boldsymbol{\lambda} \right\}^M$$

where $n \triangleq \max\{r, t\}$ and $K_{m,n}$ are constant defined in [35]. The above expectation is finite because the order of infinity of the integrand is $1/(Mm) - (n-m) \leq 1/(Mm) \leq 1/2$ in the λ_i's when $\lambda_i \downarrow 0$.

♦

5.2 Delay-limited capacities: Asymptotics

In this subsection we study the asymptotic behavior of the delay-limited capacity in the case of regular channel and large SNR or large number of antennas.

5.2.1 Large SNR

Proposition 6. *Under a long-term power constraint and with optimal power control (Proposition 4), the delay-limited capacity of a $t \times r$ BF-AWGN regular channel is given by*

$$C_{\text{delay}}(\gamma) \sim m \, \log_2[\gamma/(m \, \mathbb{E}[1/\bar{\lambda}_M])] \tag{28}$$

asymptotically as $\gamma \to \infty$. ♦

Proof. Deferred to Appendix C.

Remark 11. Considering the case of regular fading, there is a simple (albeit suboptimal) power allocation scheme (with a long-term constraint) $\bar{\boldsymbol{\Gamma}}$ that will be referred to as *channel inversion*. It is defined by

$$\bar{\gamma}_{k,i} = \frac{2^{R/m}}{\bar{\lambda}_M} \qquad \text{for } k = 1 \ldots, M, i = 1, \ldots, m$$

According to this rule, the same power is allocated to all the Mm channel degrees of freedom, without doing waterfilling. The resulting mutual information is not constant for all $\boldsymbol{\Lambda}$ but is always larger than R. In fact,

$$\begin{aligned} I_M(\mathbf{\Lambda},\mathbf{\Gamma}) &= \frac{1}{M}\sum_{k=1}^{M}\sum_{i=1}^{m}\log_2\left(1+\lambda_{k,i}2^{R/m}/\bar{\lambda}_M\right) \\ &\geq \frac{1}{M}\sum_{k=1}^{M}\sum_{i=1}^{m}\log_2\left(\lambda_{k,i}2^{R/m}/\bar{\lambda}_M\right) \\ &= R \end{aligned}$$

Obviously, as shown in the proof of Proposition 5, the long-term average transmitted power $\bar{\gamma}(R) = m2^{R/m}\mathbb{E}[1/\bar{\lambda}_M]$ resulting from $\bar{\mathbf{\Gamma}}$ is larger than $\gamma(R)$, resulting from the optimal power allocation that keeps the instantaneous mutual information fixed to R. However, for regular fading, $\bar{\gamma}(R)$ is finite and, given the simplicity of the scheme, that allocates the same power to all coding dimensions, it makes sense to consider it as a suboptimal solution. A coding scheme with rate R and power allocation $\bar{\mathbf{\Gamma}}$ still achieves zero outage probability. Interestingly, from Proposition 6 we get that this suboptimal scheme is asymptotically optimal. In fact, $\bar{\gamma}(R) \sim \gamma(R)$ for large R. ♦

5.2.2 Large number of antennas

We now assume that the entries of the channel matrices $\mathbf{A}_k$ are i.i.d. with mean zero and variance 1. We study the behavior of the normalized delay-limited capacity (expressed in bit per antenna degree of freedom), $c(\gamma) = C_{\text{delay}}(\gamma)/m$, as $m \to \infty$ and $\max\{t,r\}/m \to \alpha > 0$. We show that the limiting $c(\gamma)$ does not depend on the number of blocks M and coincides with the limiting normalized ergodic capacity. This should not come as a surprise, because as $m \to \infty$ and $t/r \to \alpha$, the empirical distribution of the elements of $\mathbf{\Lambda}$ converges almost surely to a deterministic distribution with known density. Therefore, transmitting over many blocks (subcarriers) has no effect on capacity.

We make use of the following result (see, e.g., [39, 3] and references therein):

Proposition 7. **[3]** *Let $\mathbf{A} \in \mathbb{C}^{r\times t}$ be a random matrix with i.i.d. elements with zero mean and unit variance. Consider the following empirical cdf*

$$F_m(\theta) = \frac{1}{m}\sum_{i=1}^{m} 1\{\theta_i \leq \theta\}$$

where $\theta_1, \dots, \theta_m$ are the non-zero eigenvalues of $\frac{1}{m}\mathbf{A}\mathbf{A}^\dagger$. Then, as $m \to \infty$ and $\max\{t, r\}/m \to \alpha$, $F_m(\theta) \to F(\theta)$ almost surely, where $F(\theta)$ has probability density function

$$f(\theta) = \frac{1}{2\pi\theta}\sqrt{[\theta - a(\alpha)]_+[b(\alpha) - \theta]_+} \tag{29}$$

with $a(\alpha) \triangleq (\sqrt{\alpha} - 1)^2$ and $b(\alpha) \triangleq (\sqrt{\alpha} + 1)^2$. ♦

We define $\mathbf{\Theta} = \mathbf{\Lambda}/m$ so that $\theta_{k,i} = \lambda_{k,i}/m$. Since $\theta_{k,i}\gamma^{\rm lt}_{k,i}(\mathbf{\Theta}) = \lambda_{k,i}\gamma^{\rm lt}_{k,i}(\mathbf{\Lambda})$, we can consider an equivalent channel with eigenvalues $\theta_{k,i}$, SNR assignment $\gamma^{\rm lt}_{k,i}(\mathbf{\Theta})$, and input constraint

$$\mathbb{E}\left[\frac{1}{Mm}\sum_{k=1}^{M}\sum_{i=1}^{m}\gamma_{k,i}(\mathbf{\Theta})\right] = \gamma$$

Then, we have the following:

Proposition 8. *Under the above assumptions, $c(\gamma) \to C_{\rm CSIR,CSIT}(\gamma)/m$ as $m \to \infty$, irrespectively of M, where $C_{\rm CSIR,CSIT}(\gamma)$ is defined in (4).* ♦

Proof. Deferred to Appendix D.

Remark 12. Reference [7] shows that, for $\gamma \to \infty$, $c(\gamma) \sim \log_2\gamma$ as confirmed by our numerical results (Figures 5 and 6). ♦

6 Numerical results

In this section we illustrate the theory outlined above through some numerical examples. In all cases we consider a BF-AWGN channel where the entries of the channel gain matrices $\mathbf{A}_k$ are statistically independent from each other.

6.1 One transmit antenna

With a single transmit antenna and r receive antennas, the matrices $\mathbf{A}_k$ are length-r column vectors, so that $|\mathbf{A}_k|^2$ is the only non-zero eigenvalue of $\mathbf{A}_k\mathbf{A}_k^\dagger$, $\mathbf{V}_k = 1$ (a scalar), and the optimal beamforming matrices $\mathbf{W}_k$ reduce to the scalars $\sqrt{\gamma_{1,k}}$. The instantaneous mutual information is

$$I_M(\mathbf{\Lambda}, \mathbf{\Gamma}) = \frac{1}{M}\sum_{k=1}^{M}\log_2(1 + |\mathbf{A}_k|^2\gamma_{1,k}) \tag{30}$$

If the fading has a Rayleigh distribution, $|\mathbf{A}_k|^2$ is chi-square distributed with $2r$ degrees of freedom.

6.2 One receive antenna

With t transmit antennas and a single receive antenna, the matrices $\mathbf{A}_k$ are length-t row vectors, so that $|\mathbf{A}_k|^2$ is the only non-zero eigenvalue of $\mathbf{A}_k^\dagger \mathbf{A}_k$ and $\mathbf{V}_k$ is a unitary matrix whose first column is equal to $\mathbf{A}_k^\dagger/|\mathbf{A}_k|$. The instantaneous mutual information is given again by (30), and for an independent Rayleigh BF-AWGN channel $|\mathbf{A}_k|^2$ is chi-square distributed with $2t$ degrees of freedom. As we can see, we have full reciprocity between transmit-only and receive-only diversity. As observed in the Introduction, this reciprocity, which holds for the AWGN channel, does not hold with the fading channel without CSIT [35]: reciprocity here is due to the availability of such CSIT.

For $M = 1$, the delay-limited capacity with optimal power allocation under a long-term constraint (corresponding to $\gamma_{1,1} = (2^R - 1)/|\mathbf{A}_1|^2$) is

$$C_{\text{delay}} = \log_2 \left(1 + \frac{\gamma}{\mathbb{E}[|\mathbf{A}_1|^{-2}]} \right) \tag{31}$$

On an independent Rayleigh BF-AWGN channel, $\mathbb{E}[|\mathbf{A}_1|^{-2}] = 1/(t-1)$, and hence:

$$C_{\text{delay}} = \log_2[1 + (t-1)\gamma] \tag{32}$$

6.3 More than one transmit/receive antennas

Generalizing the results above to the case of more than one transmit / receive antennas, we consider the independent Rayleigh BF-AWGN channel (i.e., we assume that al the entries of each $\mathbf{A}_k$ are statistically independent and distributed as $\mathcal{N}_c(\mathbf{0}, \mathbf{I})$) with $M = 1$ (i.e., every code word is affected by a constant fading value — no interleaving). Although in principle it is possible to calculate the delay-limited capacity with the optimal power-allocation scheme, analytical complexity suggests to resort to Monte-Carlo integration. Figs. 2 and 3 show the delay-limited capacity versus SNR (γ) for the $t \times t$ ($t = 2, 4, 8$, and 16) and the $t \times 2$ (or $2 \times t$ for reciprocity), $t = 2$ to 8 BF-AWGN channel. Moreover, Fig. 2 shows, for comparison, the achievable rate with channel inversion (see Remark 11). We note that the delay-limited capacity with optimal power allocation exceeds the capacity of the $t \times t$ AWGN channel ($C_{\text{AWGN}} = \log_2(1 + t^2\gamma)$,

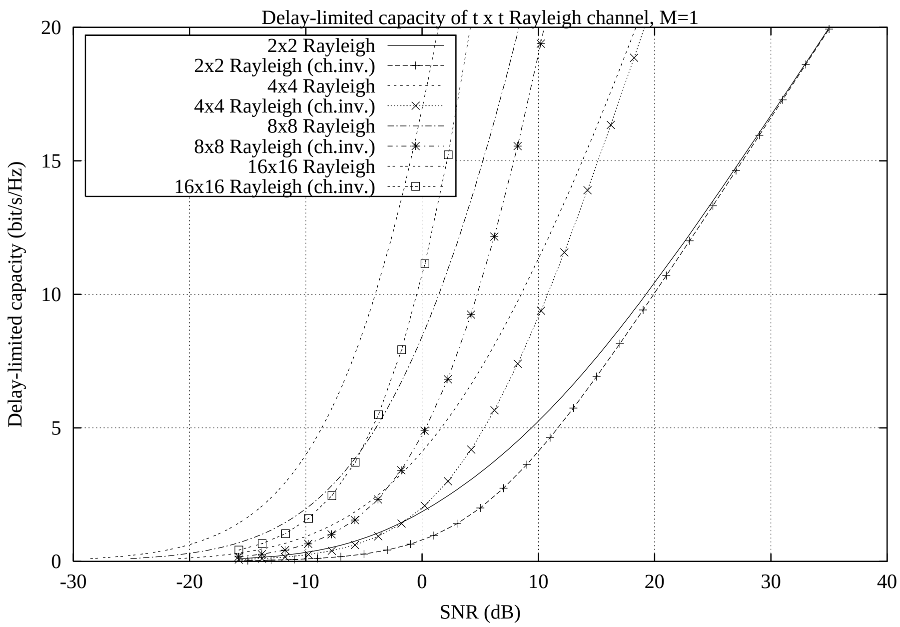

Figure 2: Delay-limited capacity (bit / s / Hz) for the independent Rayleigh $t \times t$ BF-AWGN for $M = 1$ and $t = 2, 4, 8$, and 16 obtained by Monte-Carlo integration. The achievable rate with channel inversion is also shown for comparison.

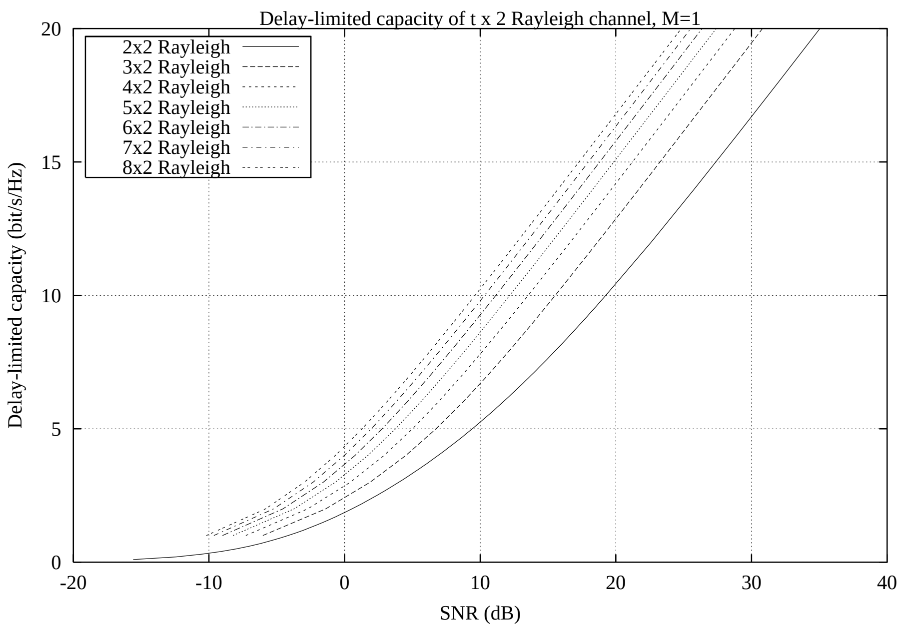

Figure 3: Delay-limited capacity (bit/s/Hz) for the independent Rayleigh $t \times 2$ or $2 \times t$ BF-AWGN for $M = 1$ and $t = 2$ to 8 obtained by Monte-Carlo integration.

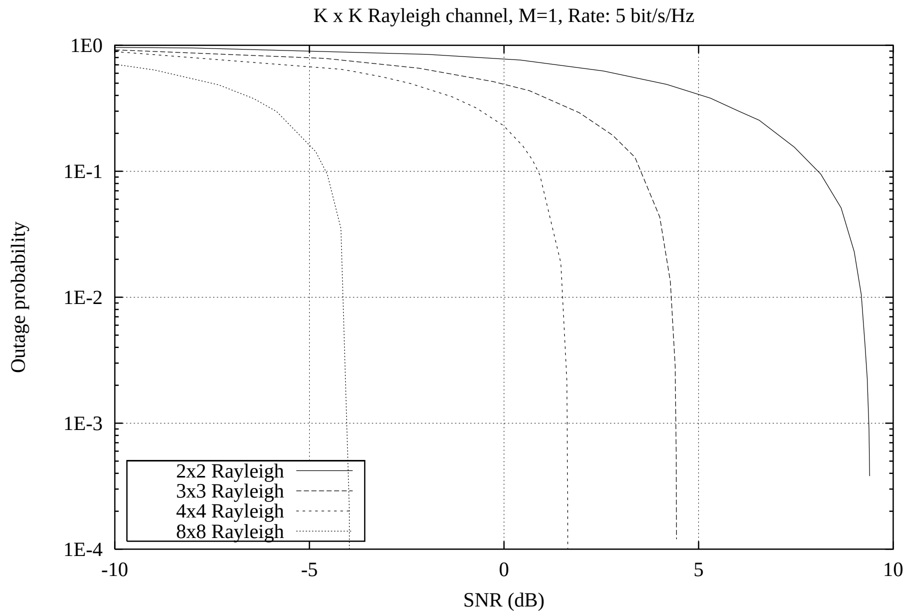

Figure 4: Outage probability for the independent Rayleigh $t \times t$ BF-AWGN for $M = 1$ and $t = 2, 3, 4, 8$ obtained by Monte-Carlo integration.

see [35]) for sufficiently high SNR γ. This is a consequence of the fact that delay-limited capacity with optimal power allocation behaves asymptotically as $m \log_2 \gamma$, while for the AWGN channel it behaves as $\log_2 \gamma$.

6.4 Outage Probability

Fig. 4 shows the outage probability corresponding to the $t \times t$ Rayleigh BF-AWGN channel with $t = 2, 3, 4$, and 8 and $M = 1$ versus the SNR γ. The required transmission rate is $R = 5$ bit/s/Hz. It can be noticed that an outage probability level of 10^{-2} requires an SNR very close to that corresponding to a delay-limited capacity equal to 5 bit/s/Hz.

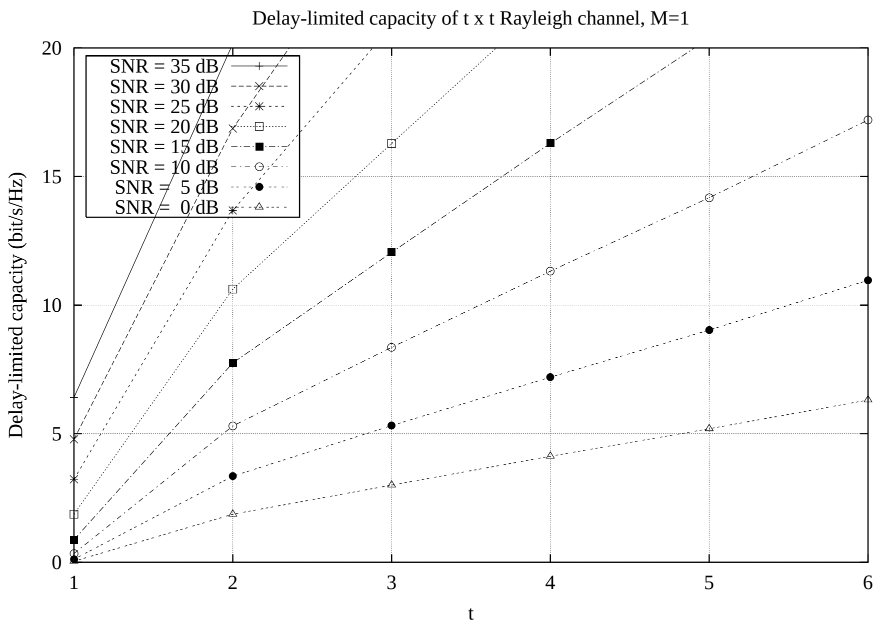

Figure 5: Delay-limited capacity (bit / s / Hz) versus number of antennas (t) for the independent Rayleigh $t \times t$ BF-AWGN for $M = 1$ (results obtained by Monte-Carlo integration).

6.5 Delay-limited capacity versus number of antennas

Figs. 5 and 6 show the delay-limited capacity versus the number t of antennas for the independent Rayleigh $t \times t$ BF-AWGN channel for $M = 1$ and 4 (results obtained by Monte-Carlo integration). Besides the very high values of capacity achievable, it is interesting to note the linear growth rate of the capacity with t, experienced even for small t.

Moreover, comparing Figs. 5 and 6 we note that they are almost independent of M. Since M reflects the amount of interleaving allowed, this fact suggests that antenna diversity can be traded for interleaving (and hence interleaving delay), as observed in a different context in [37]. If the M blocks are transmitted in parallel over separate carriers, as suggested in the Introduction, then antenna diversity is traded off for frequency diversity.

The results obtained for the independent Rayleigh $t \times 2$ (or $2 \times t$, by reciprocity) BF-AWGN channel with $M = 1$ are shown in Fig. 7. Again, increasing M yields little improvement in the

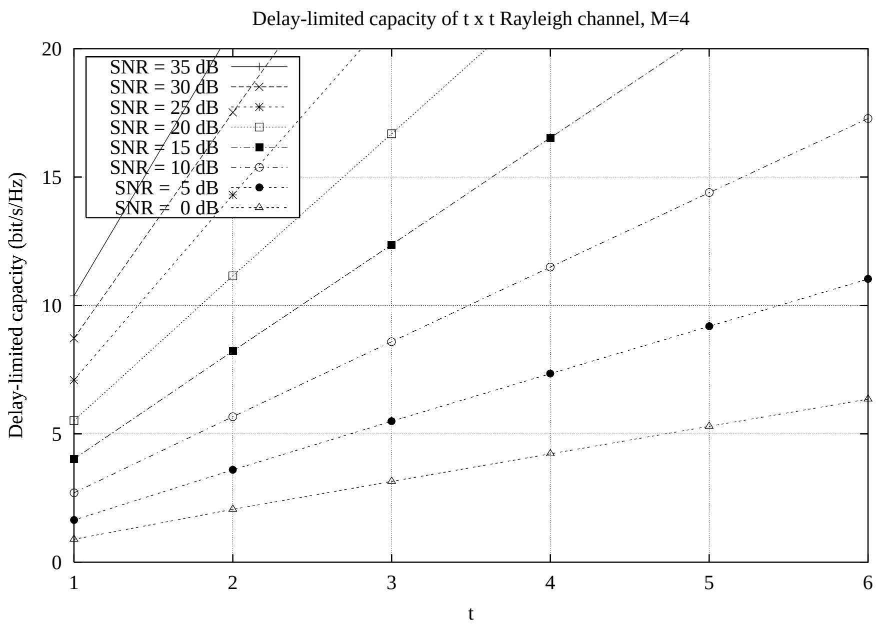

Figure 6: Delay-limited capacity (bit / s / Hz) versus number of antennas (t) for the independent Rayleigh $t \times t$ BF-AWGN for $M = 4$ (results obtained by Monte-Carlo integration).

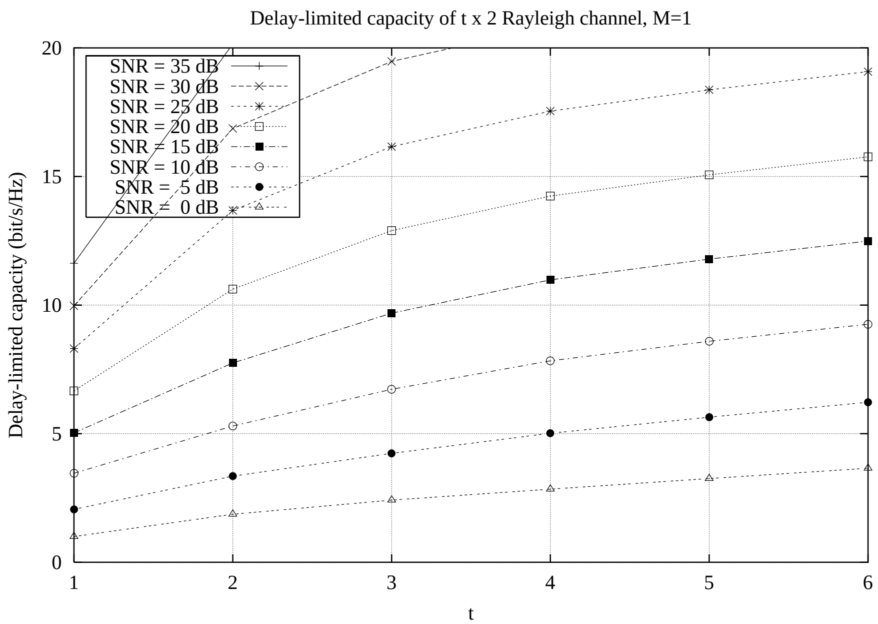

Figure 7: Delay-limited capacity (bit / s / Hz) versus number of antennas (t) for the independent Rayleigh $t \times 2$ (or $2 \times t$) BF-AWGN for $M = 1$ (results obtained by Monte-Carlo integration).

delay-limited capacity.

7 Conclusions

To understand the ultimate performance limits of a radio system consisting of a transmitter with t antennas and a receiver with r antennas, we have evaluated the minimum outage probability and the delay-limited capacity of a channel with transmit-power constraint, independent flat fading between the transmit and receive antennas, Gaussian noise added independently at each receiver antenna, and CSI available at the transmitter and at the receiver. Optimum coding for this channel is achieved by concatenating a code that is optimum for the unfaded AWGN channel with a suitable beamformer, which changes from block to block. Considering the long-term power constraint, we have shown that:

1. The delay-limited capacity is strictly positive for a wide family of fading channels (in particular, for Rayleigh fading channels with $Mm > 1$).

2. The optimal power allocation scheme can be given an explicit form, and a suboptimal (but asymptotically optimal) simpler power allocation scheme can be exhibited (see Remark 11).

3. The delay-limited capacity increases linearly with the signal-to-noise ratio in dB, and is almost independent of M (so that antenna diversity can substitute for time diversity, as generated by interleaving, or for frequency diversity). Thus, antenna diversity can be a substitute for time diversity (as introduced by interleaving) or frequency diversity.

4. For large m, the delay-limited capacity is asymptotically equal to the ergodic capacity (in a sense, by expanding the spatial dimension we get back the ergodicity that would otherwise be lost due to the time or frequency constraints).

5. The availability of CSIT makes transmit-antenna diversity equivalent, in terms of capacity improvement, to receive-antenna diversity.

These results are based on perfect CSIR and CSIT. The latter may be sometimes unavailable *exactly* due to the system causality so that, in such case, the results obtained represent an upper bound to the achievable performance.

Acknowledgments

We are most grateful to the Associate Editor and the anonymous reviewers, whose constructive comments allowed us to improve the technical quality as well as the legibility of the paper.

Appendices

A Equivalence of short-term and long-term outage regions

In this appendix we show the equivalence of the outage regions $\mathcal{R}_{\text{off}}(R,\gamma)$ and $\mathcal{R}^*_{\text{off}}(R,\gamma^*)$ (and of their complements) provided that $\gamma = \gamma^*$. First, we illustrate this result in a simple two-dimensional case. Then, we proceed to show the general result.

A.1 Example

Consider the case of a single transmitting / receiving antenna ($m = 1$) and an $M = 2$ BF-AWGN channel. We derive analytically the outage regions in order to show that they are equivalent in the case of short-term and long-term power constraints. Assume that $\gamma = \gamma^* = s > 0$ is a given parameter. Since $m = 1$, for easier notation we drop in the following the index i of $\lambda_{k,i}$ and so on. We restrict our attention to the region $\mathcal{S}_2 = \{\mathbf{\Lambda} : \lambda_1 \geq \lambda_2 \geq 0\}$. The results can be extended to $\mathbb{R}^2_+$ by a symmetric argument.

Short-term constraint. The outage or power-off region is determined by using the results of Proposition 3. Given $\xi = \xi^{\text{st}}(\mathbf{\Lambda},\gamma)$, two instances may occur:

1. $1/\lambda_2 \geq \xi > 1/\lambda_1$ and $\mathcal{M}(s) = \{1\}$.
2. $\xi > 1/\lambda_2 > 1/\lambda_1$ and $\mathcal{M}(s) = \{1,2\}$.

In the first case, power is transmitted only over the "stronger" block 1 ($\gamma_1^{\text{st}} > \gamma_2^{\text{st}} = 0$). In the second case, power is transmitted over both blocks ($\gamma_1^{\text{st}} > \gamma_2^{\text{st}} > 0$). Given $\mathbf{\Lambda} = (\lambda_1, \lambda_2)$ and s, we define the regions $\mathcal{R}_\mu(s) \triangleq \{\mathbf{\Lambda} \in \mathcal{S}_2, |\mathcal{M}(s)| = \mu\}$ for $\mu = 1, 2$ and find their analytic expression.

From Proposition 3, we have

$$\xi = \xi^{\text{st}}(\mathbf{\Lambda}, s) = \begin{cases} 2s + 1/\lambda_1 & |\mathcal{M}(s)| = 1 \\ s + \frac{1}{2}(1/\lambda_1 + 1/\lambda_2) & |\mathcal{M}(s)| = 2 \end{cases}$$

Hence, substituting these values of ξ in the inequalities defining the occurrence of $|\mathcal{M}(s)| = 1$ or 2, we have:

$$\mathcal{R}_1(s) = \left\{\mathbf{\Lambda} :\ s \leq \frac{1}{2}\left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1}\right), \lambda_1 \geq \lambda_2 > 0\right\} \tag{33}$$

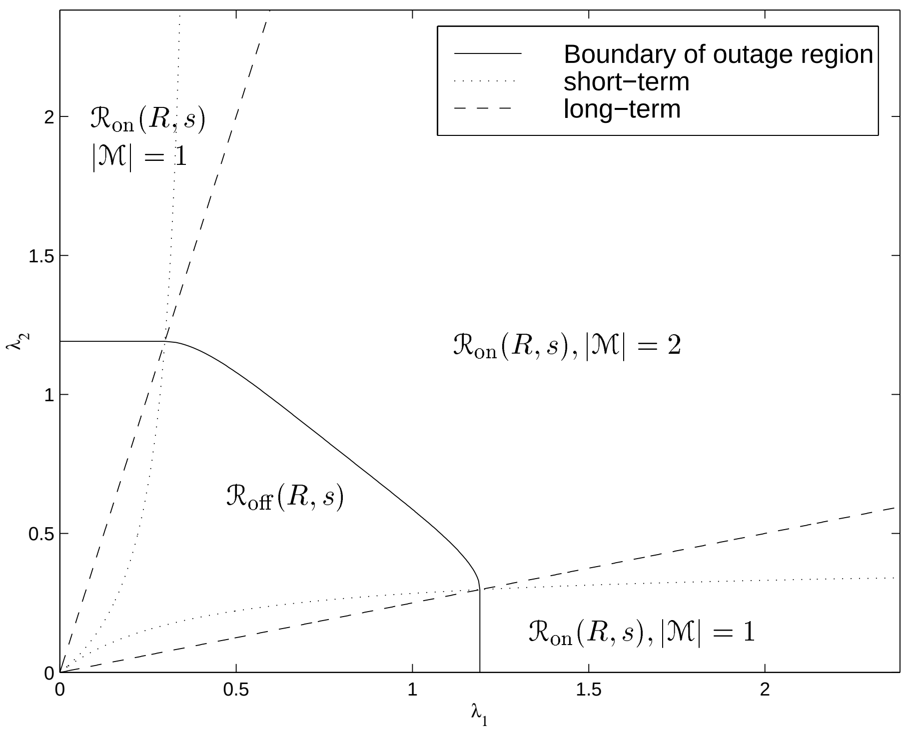

Figure 8: Outage region $\mathcal{R}_{\text{off}}(R,s)$ for a system with $t = r = 1$, $M = 2$, $R = 1$ bit/s/Hz, and $s = 1$ dB. The boundaries of the regions of constant $|\mathcal{M}(s)|$, corresponding to the short-term power constraint (resp. $|\mathcal{M}^*(R)|$, long-term) are also indicated as dotted (resp. dashed) lines.

and

$$\mathcal{R}_2(s) = \left\{ \mathbf{\Lambda} : \ s > \frac{1}{2}\left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1}\right), \lambda_1 \geq \lambda_2 > 0 \right\} \tag{34}$$

Note that $\mathcal{R}_1(s)$ and $\mathcal{R}_2(s)$ are disjoint and their union is $\mathcal{S}_2$.

Now, we define the conditioned outage regions $\mathcal{R}_{\text{off},\mu}(R,s) \triangleq \mathcal{R}_{\text{off}}(R,s) \cap \mathcal{R}_\mu(s)$. We have:

$$\begin{aligned} \mathcal{R}_{\text{off},1}(R,s) &\triangleq \{\mathbf{\Lambda} : I_2(\mathbf{\Lambda}, \mathbf{\Gamma}) < R\} \cap \mathcal{R}_1(s) \\ &= \left\{\mathbf{\Lambda} : \frac{1}{2}\log_2(1 + 2\lambda_1 s) < R\right\} \cap \mathcal{R}_1(s) \\ &= \{\mathbf{\Lambda} : \lambda_1 < (2^{2R} - 1)/(2s)\} \cap \mathcal{R}_1(s) \\ &= \bar{\mathcal{R}}_{\text{off},1}(R,s) \cap \mathcal{R}_1(s) \end{aligned} \tag{35}$$

with

$$\bar{\mathcal{R}}_{\text{off},1}(R,s) \triangleq \left\{\mathbf{\Lambda} : \lambda_1 < (2^{2R} - 1)/(2s)\right\}$$

and

$$\begin{aligned} \mathcal{R}_{\text{off},2}(R,s) &\triangleq \left\{\mathbf{\Lambda} : \frac{1}{2}\log_2(\xi\lambda_1) + \frac{1}{2}\log_2(\xi\lambda_2) < R\right\} \cap \mathcal{R}_2(s) \\ &= \left\{\mathbf{\Lambda} : \log_2\left[s + \frac{1}{2}(1/\lambda_1 + 1/\lambda_2)\right] + \frac{1}{2}\log_2(\lambda_1\lambda_2) < R\right\} \cap \mathcal{R}_2(s) \\ &= \left\{\mathbf{\Lambda} : \frac{1}{2}\log_2\left[\frac{(\lambda_1 + \lambda_2 + 2\lambda_1\lambda_2 s)^2}{4\lambda_1\lambda_2}\right] < R\right\} \cap \mathcal{R}_2(s) \\ &= \left\{\mathbf{\Lambda} : \frac{(\lambda_1 + \lambda_2 + 2\lambda_1\lambda_2 s)^2}{4\lambda_1\lambda_2} < 2^{2R}\right\} \cap \mathcal{R}_2(s) \\ &= \bar{\mathcal{R}}_{\text{off},2}(R,s) \cap \mathcal{R}_2(s) \end{aligned} \tag{36}$$

with

$$\bar{\mathcal{R}}_{\text{off},2}(R,s) \triangleq \left\{\mathbf{\Lambda} : \frac{(\lambda_1 + \lambda_2 + 2\lambda_1\lambda_2 s)^2}{4\lambda_1\lambda_2} < 2^{2R}\right\}$$

The overall outage region is the union of the two disjoint (because obtained by intersection with disjoint regions) conditioned outage regions:

$$\mathcal{R}_{\text{off}}(R,s) = \mathcal{R}_{\text{off},1}(R,s) \cup \mathcal{R}_{\text{off},2}(R,s)$$

These regions are shown in Figure 8: both the inner outage region $\mathcal{R}_{\text{off}}(R,s)$ and the outer power-on region $\mathcal{R}_{\text{on}}(R,s)$ are partitioned in two sub-regions according to the value taken by $|\mathcal{M}(s)| = 1$ or 2. In the figure, the boundaries of $\mathcal{R}_1(s)$ and $\mathcal{R}_2(s)$ are solid lines drawn according to eqs. (33) and (34).

Long-term constraint. The outage or power-off region is determined by using the results of Proposition 4. Given $\xi = \xi^{\text{lt}}(\mathbf{\Lambda}, \gamma)$, two instances may occur:

1. $1/\lambda_2 \geq \xi > 1/\lambda_1$ and $\mathcal{M}^*(R) = \{1\}$.
2. $\xi > 1/\lambda_2 > 1/\lambda_1$ and $\mathcal{M}^*(R) = \{1, 2\}$.

Given $\mathbf{\Lambda} = (\lambda_1, \lambda_2)$ and R, we define the regions $\mathcal{R}^*_\mu(R) \triangleq \{\mathbf{\Lambda} \in \mathcal{S}_2, |\mathcal{M}^*(R)| = \mu\}$ for $\mu = 1, 2$ and find their analytic expression.

From Proposition 4, we have

$$\xi = \xi^{\text{lt}}(\mathbf{\Lambda}, R) = \begin{cases} 2^{2R}/\lambda_1 & |\mathcal{M}^*(R)| = 1 \\ 2^R/\sqrt{\lambda_1\lambda_2} & |\mathcal{M}^*(R)| = 2 \end{cases}$$

Hence:

$$\mathcal{R}^*_1(R) = \{\mathbf{\Lambda} : \ \lambda_1 \geq 2^{2R}\lambda_2, \lambda_1 \geq \lambda_2 > 0\} \tag{37}$$

and

$$\mathcal{R}^*_2(R) = \{\mathbf{\Lambda} : \ \lambda_1 < 2^{2R}\lambda_2, \lambda_1 \geq \lambda_2 > 0\} \tag{38}$$

Then, the conditioned outage regions $\mathcal{R}^*_{\text{off},\mu}(R, s) \triangleq \mathcal{R}^*_{\text{off}}(R, s) \cap \mathcal{R}^*_\mu(R)$ are:

$$\begin{aligned} \mathcal{R}^*_{\text{off},1}(R, s) &\triangleq \{\mathbf{\Lambda} : P_2(\mathbf{\Gamma}) > s\} \cap \mathcal{R}^*_1(R) \\ &= \left\{\mathbf{\Lambda} : \frac{1}{2}(2^{2R}/\lambda_1 - 1/\lambda_1) > s\right\} \cap \mathcal{R}^*_1(R) \\ &= \{\mathbf{\Lambda} : \lambda_1 < (2^{2R} - 1)/(2s)\} \cap \mathcal{R}^*_1(R) \\ &= \bar{\mathcal{R}}_{\text{off},1}(R, s) \cap \mathcal{R}^*_1(R) \end{aligned} \tag{39}$$

and

$$\begin{aligned} \mathcal{R}^*_{\text{off},2}(R, s) &\triangleq \left\{\mathbf{\Lambda} : \frac{1}{2}(2^R/\sqrt{\lambda_1\lambda_2} - 1/\lambda_1) + \frac{1}{2}(2^R/\sqrt{\lambda_1\lambda_2} - 1/\lambda_2) > s\right\} \cap \mathcal{R}^*_2(R) \\ &= \left\{\mathbf{\Lambda} : \frac{2^R}{\sqrt{\lambda_1\lambda_2}} - \frac{\lambda_1 + \lambda_2}{2\lambda_1\lambda_2} > s\right\} \cap \mathcal{R}^*_2(R) \\ &= \left\{\mathbf{\Lambda} : \frac{(\lambda_1 + \lambda_2 + 2\lambda_1\lambda_2 s)^2}{4\lambda_1\lambda_2} < 2^{2R}\right\} \cap \mathcal{R}^*_2(R) \\ &= \bar{\mathcal{R}}_{\text{off},2}(R, s) \cap \mathcal{R}^*_2(R) \end{aligned} \tag{40}$$

Again, the overall outage region is the union of the two disjoint conditioned outage regions:

$$\mathcal{R}^*_{\text{off}}(R, s) = \mathcal{R}^*_{\text{off},1}(R, s) \cup \mathcal{R}^*_{\text{off},2}(R, s)$$

Again, all regions are shown in Figure 8 where now the boundaries of $\mathcal{R}_1^*(s)$ and $\mathcal{R}_2^*(s)$ are the dashed lines drawn according to eqs. (37) and (38).

Comparison. Comparing eqs. (35) and (36) to eqs. (39) and (40), we note that each conditioned outage region is obtained by intersecting a common region $\bar{\mathcal{R}}_{\text{off},\mu}(R,s)$ with $\mathcal{R}_\mu(s)$ (short-term constraint) and $\mathcal{R}_\mu^*(R)$ (long-term constraint). The results are different, as can be seen from Figure 8 but the union of the conditioned outage regions (namely, the outage region itself) is the same in both cases because the boundaries of $\mathcal{R}_{\text{off},\mu}(R,s)$ and $\mathcal{R}_{\text{off},\mu}^*(R,s)$ coincide. In fact, the common boundary between $\mathcal{R}_{\text{off},1}(R,s)$ and $\mathcal{R}_{\text{on},1}(R,s)$, $\mathcal{R}_{\text{off},2}^*(R,s)$ and $\mathcal{R}_{\text{on},2}^*(R,s)$ in $\mathcal{S}_2$ is the vertical segment $\lambda_1 = (2^{2R}-1)/(2s)$ from $\lambda_2 = 0$ to $\lambda_2 = (1-2^{-2R})/(2s)$. The common boundary between $\mathcal{R}_{\text{off},2}(R,s)$ and $\mathcal{R}_{\text{on},2}(R,s)$, $\mathcal{R}_{\text{off},2}^*(R,s)$ and $\mathcal{R}_{\text{on},2}^*(R,s)$ is the line of equation

$$\frac{(\lambda_1+\lambda_2+2\lambda_1\lambda_2 s)^2}{4\lambda_1\lambda_2} = 2^{2R}$$

joining the points of coordinates $((2^{2R}-1)/(2s),(1-2^{-2R})/(2s))$ and $((2^R-1)/s,(2^R-1)/s)$.

The boundary line between $\mathcal{R}_1(s)$ and $\mathcal{R}_2(s)$ is defined by equation:

$$\frac{1}{\lambda_2} - \frac{1}{\lambda_1} = 2s$$

and the boundary line between $\mathcal{R}_1^*(R)$ and $\mathcal{R}_2^*(R)$ is defined by equation:

$$\lambda_1 = 2^{2R}\lambda_2$$

The curves intersect in $(0,0)$ and $((2^{2R}-1)/(2s),(1-2^{-2R})/(2s))$ (in Fig. 8: $(1.192, 0.298)$). Hence, the boundaries of $\mathcal{R}_{\text{off},\mu}(R,s)$ and $\mathcal{R}_{\text{off},\mu}^*(R,s)$ coincide. In the following section we extend this result to the general case.

A.2 General result

Proposition 9. *The outage regions $\mathcal{R}_{\text{off}}(R,s)$ and $\mathcal{R}_{\text{off}}^*(R,s)$ defined in Propositions 3 and 4 are equivalent.* ♦

Proof. From the results of Propositions 3 and 4 we have the following expressions of the outage regions:

$$\mathcal{R}_{\text{off}}(R,s) = \left\{\mathbf{\Lambda} : I_M(\mathbf{\Lambda},\mathbf{\Gamma}^{\text{st}}(\mathbf{\Lambda},s)) < R, P_M(\mathbf{\Gamma}^{\text{st}}(\mathbf{\Lambda},s)) = s\right\}$$

for the short-term problem and

$$\mathcal{R}^*_{\rm off}(R,s) = \Big\{\mathbf{\Lambda} : I_M(\mathbf{\Lambda}, \mathbf{\Gamma}^{\rm lt}(\mathbf{\Lambda}, R)) = R, P_M(\mathbf{\Gamma}^{\rm lt}(\mathbf{\Lambda}, R)) > s\Big\}$$

for the long-term problem.

By definition, the orthant $\mathbb{R}_+^{Mm}$ is divided into two sub-regions, namely, the outage and the power-on region and this division is determined by the inner boundary. The boundaries are given by

$$\begin{aligned}
\mathcal{B}_{\rm off}(R,s) &= \Big\{\mathbf{\Lambda} : I_M(\mathbf{\Lambda}, \mathbf{\Gamma}^{\rm st}(\mathbf{\Lambda}, s)) = R, P_M(\mathbf{\Gamma}^{\rm st}(\mathbf{\Lambda}, s)) = s\Big\} \\
\mathcal{B}^*_{\rm off}(R,s) &= \Big\{\mathbf{\Lambda} : I_M(\mathbf{\Lambda}, \mathbf{\Gamma}^{\rm lt}(\mathbf{\Lambda}, R)) = R, P_M(\mathbf{\Gamma}^{\rm lt}(\mathbf{\Lambda}, R)) = s\Big\}
\end{aligned}$$

The assert is proven if we show that the boundary regions coincide, namely: $\mathcal{B}_{\rm off}(R,s) = \mathcal{B}^*_{\rm off}(R,s)$. To this purpose, we define the conditioned outage regions

$$\begin{aligned}
\mathcal{B}_{{\rm off},\mu}(R,s) &= \Big\{\mathbf{\Lambda} : I_M(\mathbf{\Lambda}, \mathbf{\Gamma}^{\rm st}(\mathbf{\Lambda}, s)) = R, P_M(\mathbf{\Gamma}^{\rm st}(\mathbf{\Lambda}, s)) = s, |\mathcal{M}(s)| = \mu\Big\} \\
&= \mathcal{B}_{\rm off}(R,s) \cap \mathcal{R}_\mu(s) \\
\mathcal{B}^*_{{\rm off},\mu}(R,s) &= \Big\{\mathbf{\Lambda} : I_M(\mathbf{\Lambda}, \mathbf{\Gamma}^{\rm lt}(\mathbf{\Lambda}, R)) = R, P_M(\mathbf{\Gamma}^{\rm lt}(\mathbf{\Lambda}, R)) = s, |\mathcal{M}^*(R)| = \mu\Big\} \\
&= \mathcal{B}_{\rm off}(R,s) \cap \mathcal{R}_\mu(R)
\end{aligned}$$

where $\mathcal{R}_\mu(s) \triangleq \{\mathbf{\Lambda} : |\mathcal{M}(s)| = \mu\}$ and $\mathcal{R}^*_\mu(R) \triangleq \{\mathbf{\Lambda} : |\mathcal{M}^*(R)| = \mu\}$. Now, we show that $\mathcal{B}_{{\rm off},\mu}(R,s) = \mathcal{B}^*_{{\rm off},\mu}(R,s)$. From the results of Propositions 3 and 4 (and from Remark 4, which allows us to write $\mathcal{M}$ for $\mathcal{M}(s)$ and $\mathcal{M}^*(R)$ since $|\mathcal{M}| = \mu$ in both cases *for a fixed* $\mathbf{\Lambda}$) we have:

$$\begin{aligned}
\mathcal{B}^*_{{\rm off},\mu}(R,s) &= \left\{\mathbf{\Lambda} : \frac{1}{M}\sum_{(k,i)\in\mathcal{M}}\left[\xi^{\rm lt}(\mathbf{\Lambda}, R) - \frac{1}{\lambda_{k,i}}\right] = s, |\mathcal{M}| = \mu\right\} \\
&= \left\{\mathbf{\Lambda} : \frac{1}{M}\sum_{(k,i)\in\mathcal{M}}\left[\frac{2^{MR/\mu}}{\prod_{(k',i')\in\mathcal{M}}\lambda_{k',i'}^{1/\mu}} - \frac{1}{\lambda_{k,i}}\right] = s, |\mathcal{M}| = \mu\right\} \\
&= \left\{\mathbf{\Lambda} : \frac{\mu}{M}\frac{2^{MR/\mu}}{\lambda'} - \frac{1}{M}\sum_{(k,i)\in\mathcal{M}}\frac{1}{\lambda_{k,i}} = s, |\mathcal{M}| = \mu\right\} \\
&= \left\{\mathbf{\Lambda} : \frac{2^{MR/\mu}}{\lambda'} = \frac{1}{\mu}\sum_{(k,i)\in\mathcal{M}}\frac{1}{\lambda_{k,i}} + \frac{Ms}{\mu} = \xi^{\rm st}(\mathbf{\Lambda}, s), |\mathcal{M}| = \mu\right\} \\
&= \left\{\mathbf{\Lambda} : \frac{MR}{\mu} - \frac{1}{\mu}\sum_{(k,i)\in\mathcal{M}}\log_2\lambda_{k,i} = \log_2\xi^{\rm st}(\mathbf{\Lambda}, s), |\mathcal{M}| = \mu\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \left\{\mathbf{\Lambda} : \frac{1}{\mu}\sum_{(k,i)\in\mathcal{M}} \log_2(\lambda_{k,i}\xi^{\mathrm{st}}(\mathbf{\Lambda},s)) = R, |\mathcal{M}| = \mu\right\} \\
&= \mathcal{B}_{\mathrm{off}}(R,s)
\end{aligned}
$$

where $\lambda' \triangleq \prod_{(k,i)\in\mathcal{M}(s)} \lambda_{k,i}^{1/|\mathcal{M}(s)|}$.

Finally, since by definition $\mathcal{B}_{\mathrm{off}}(R,s) = \bigcup_{\mu=1}^{Mm} \mathcal{B}_{\mathrm{off},\mu}(R,s)$ and $\mathcal{B}^*_{\mathrm{off}}(R,s) = \bigcup_{\mu=1}^{Mm} \mathcal{B}^*_{\mathrm{off},\mu}(R,s)$, the equality $\mathcal{B}_{\mathrm{off},\mu}(R,s) = \mathcal{B}^*_{\mathrm{off},\mu}(R,s)$ proves the assert. □

B Proof of Proposition 5

Proof. Our proof is based on the following lemma:

Lemma 1. *For every positive ξ and $x_1,\dots,x_M$, the following inequality holds:*

$$
\frac{1}{M}\sum_{i=1}^{M}\log_2[x_i + (\xi - x_i)_+] \geq \log_2\left[\frac{1}{M}\sum_{i=1}^{M}(\xi - x_i)_+\right] \tag{41}
$$

♦

Proof. Assume, w.l.o.g., that $x_1 \geq \ldots \geq x_\mu \geq \xi \geq x_{\mu+1} \geq \ldots \geq x_M$ for $0 \leq \mu < M$ where $\mu = 0$ corresponds to the case $\xi \geq x_1 \geq \ldots \geq x_M$. Since $x_i \geq \xi$ for $i = 1,\dots,\mu$, we have

$$
\frac{1}{M}\sum_{i=1}^{\mu}\log_2(x_i/\xi) \geq 0 \geq \log_2\left[\frac{M-\mu}{M}\right]
$$

Adding $\log_2 \xi$ to the first and last members, we obtain:

$$
\frac{1}{M}\sum_{i=1}^{\mu}\log_2 x_i + \frac{M-\mu}{M}\log_2\xi \geq \log_2\left[\frac{M-\mu}{M}\xi\right]
$$

Finally, the assert follows from

$$
\frac{1}{M}\sum_{i=1}^{\mu}\log_2 x_i + \frac{M-\mu}{M}\log_2\xi = \frac{1}{M}\sum_{i=1}^{M}\log_2[x_i + (\xi - x_i)_+]
$$

and

$$
\frac{1}{M}\sum_{i=1}^{M}(\xi - x_i)_+ = \frac{1}{M}\sum_{i=\mu+1}^{M}(\xi - x_i) \leq \frac{M-\mu}{M}\xi
$$

■

The function $\gamma(R)$ defined in equation (26) derives its properties from the properties of $P_M(\mathbf{\Gamma}^{\mathrm{lt}}(\mathbf{\Lambda},R))$, when $\mathbf{\Lambda}$ is fixed, which are reviewed in Remark 5.

1. $\gamma(0) = 0$ since setting $R = 0$ in (24) is equivalent to remove the lower bound constraint on the mutual information. Hence, $P_M(\mathbf{\Gamma}^{\text{lt}}(\mathbf{\Lambda}, 0))$ can be minimized down to zero.

2. The continuity and increasing properties derive from the continuity and increasing properties of $P_M(\mathbf{\Gamma}^{\text{lt}}(\mathbf{\Lambda}, R))$. Continuity derives from the continuity of the fading distribution and $P_M(\mathbf{\Gamma}^{\text{lt}}(\mathbf{\Lambda}, R))$ increases with R for a fixed $\mathbf{\Lambda}$ because increasing R in (24) is equivalent to tightening the constraint and therefore increasing the minimum attained.

3. The inequality $\gamma(R) \leq m2^{R/m}\mathbb{E}[1/\bar{\lambda}_M] < \infty$ derives from the following upper bound. Letting $\gamma_{k,i}^{\text{lt}}$ be as defined in Proposition 4, we have

$$\begin{aligned} R &= \frac{1}{M}\sum_{k=1}^{M}\sum_{i=1}^{m}\log_2(1+\gamma_{k,i}^{\text{lt}}\lambda_{k,i}) \\ &= \frac{1}{M}\sum_{k=1}^{M}\sum_{i=1}^{m}\log_2(1/\lambda_{k,i}+\gamma_{k,i}^{\text{lt}}) + m\log_2\bar{\lambda}_M \\ &\overset{(a)}{\geq} m\log_2\left(\frac{1}{Mm}\sum_{k=1}^{M}\sum_{i=1}^{m}\gamma_{k,i}^{\text{lt}}\right) + m\log_2\bar{\lambda}_M \\ &= m\log_2(P_M(\mathbf{\Gamma}^{\text{lt}}(\mathbf{\Lambda}, R))/m) + m\log_2\bar{\lambda}_M \end{aligned} \tag{42}$$

where (a) derives from Lemma 1. Thus, (42) is equivalent to

$$P_M(\mathbf{\Gamma}^{\text{lt}}(\mathbf{\Lambda}, R)) \leq m2^{R/m}/\bar{\lambda}_M$$

By taking expectation of both sides and by using the assumption of regular fading, we can prove the assertion.

■

C Proof of Proposition 6

Proof. Our proof is base on the following lemma:

Lemma 2. *Let $\lambda_1, \ldots, \lambda_n$ be a set of n positive random variables with continuous joint cdf. Denoting $\bar{\lambda} \triangleq \prod_{i=1}^{n}\lambda_i^{1/n}$ their geometric mean and $\lambda_{\min} \triangleq \min\{\lambda_1, \ldots, \lambda_n\}$ the minimum, we have*

$$\lim_{\varepsilon\to 0}\varepsilon\,\mathbb{E}[\lambda_{\min}^{-1}|\lambda_{\min} > \varepsilon\,\bar{\lambda}] = \lim_{\varepsilon\to 0}\varepsilon\,\mathbb{E}[\lambda_{\min}^{-1}1\{\lambda_{\min} > \varepsilon\,\bar{\lambda}\}] = 0$$

provided that $\mathbb{E}[\bar{\lambda}^{-1}] < \infty$ (even if $\mathbb{E}[\lambda_{\min}^{-1}]$ is not finite). ♦

Proof. First, we note that, by continuity of the joint cdf, we have

$$\lim_{\varepsilon\to 0}\mathbb{P}(\lambda_{\min} > \varepsilon\,\bar{\lambda}) = \mathbb{P}(\lambda_{\min} > 0) = 1$$

Thus, it is sufficient to show that

$$\lim_{\varepsilon\to 0}\varepsilon\,\mathbb{E}[\lambda_{\min}^{-1}1\{\lambda_{\min} > \varepsilon\,\bar{\lambda}\}] = 0 \tag{43}$$

If $\mathbb{E}[\lambda_{\min}^{-1}] < \infty$ the result is trivial, so that we assume throughout the proof that $\mathbb{E}[\lambda_{\min}^{-1}]$ is unbounded. Let us upper bound the expectation in (43) as follows:

$$\begin{aligned}
\varepsilon\,\mathbb{E}[\lambda_{\min}^{-1}1\{\lambda_{\min} > \varepsilon\,\bar{\lambda}\}] &\overset{a}{=} \sum_{i=1}^{n}\varepsilon\,\mathbb{E}[\lambda_i^{-1}1\{\lambda_i > \varepsilon\,\bar{\lambda}\}1\{\lambda_i < \lambda_{j\neq i}\}] \\
&\overset{b}{\leq} \sum_{i=1}^{n}\varepsilon\,\mathbb{E}[\lambda_i^{-1}1\{\lambda_i > \varepsilon\,\bar{\lambda}\}] \\
&\overset{c}{=} \sum_{i=1}^{n}\varepsilon\,\mathbb{E}[\lambda_i^{-1}1\{\lambda_i > \varepsilon^{n/(n-1)}\,\bar{\lambda}_i\}] \\
&\overset{d}{\leq} \sum_{i=1}^{n}\varepsilon\,\mathbb{E}[\lambda_i^{-1}1\{\lambda_i > \varepsilon^{n/(n-1)}\,\bar{\lambda}_i\}[1\{\lambda_i < \varepsilon^{\rho}\} + 1\{\lambda_i \geq \varepsilon^{\rho}\}]] \\
&\overset{e}{\leq} \sum_{i=1}^{n}\varepsilon\,\mathbb{E}[\lambda_i^{-1}1\{\lambda_i > \varepsilon^{n/(n-1)}\,\bar{\lambda}_i\}1\{\lambda_i < \varepsilon^{\rho}\}] + \sum_{i=1}^{n}\varepsilon^{1-\rho} \qquad (44)
\end{aligned}$$

where (a) derives from the definition of the minimum; (b) by upper bounding $1\{\lambda_i < \lambda_{j\neq i}\}$ (short-hand notation for $\prod_{j\neq i}1\{\lambda_i < \lambda_j\}$) by 1; (c) by defining $\bar{\lambda}_i \triangleq \prod_{j\neq i}\lambda_j^{1/(n-1)}$; (d) from the equality $1 = 1\{\lambda_i < \varepsilon^{\rho}\} + 1\{\lambda_i \geq \varepsilon^{\rho}\}$; (e) from $\lambda_i^{-1}1\{\lambda_i \geq \varepsilon^{\rho}\} \leq \varepsilon^{-\rho}$. Then, for every $0 < \rho < 1$, the second sum in (44) approaches 0 as $\varepsilon \to 0$ so that (43) is proved if we show that, for every i such that $1 \leq i \leq n$ and for some $0 < \rho < 1$,

$$\lim_{\varepsilon\to 0}\varepsilon\,\mathbb{E}[\lambda_i^{-1}1\{\lambda_i > \varepsilon^{n/(n-1)}\,\bar{\lambda}_i\}1\{\lambda_i < \varepsilon^{\rho}\}] = 0 \tag{45}$$

We restrict to the case of unbounded $\mathbb{E}[\lambda_i^{-1}]$ (otherwise, the result is trivially proved). Denoting $f(\lambda_i, \bar{\lambda}_i)$ the joint distribution function of the random variables λ_i and $\bar{\lambda}_i$, we can write the above expression as follows:

$$\begin{aligned}
\varepsilon\,\mathbb{E}[\lambda_i^{-1}1\{\lambda_i > \varepsilon^{n/(n-1)}\,\bar{\lambda}_i\}1\{\lambda_i < \varepsilon^{\rho}\}] &= \varepsilon\int_0^{\varepsilon^{\rho-n/(n-1)}}\int_{\varepsilon^{n/(n-1)}\bar{\lambda}_i}^{\varepsilon^{\rho}}\lambda_i^{-1}f(\lambda_i,\bar{\lambda}_i)\,d\lambda_i\,d\bar{\lambda}_i \\
&= \varepsilon\int_0^{\varepsilon^{\rho}}\int_{\varepsilon^{n/(n-1)}\bar{\lambda}_i}^{\varepsilon^{\rho}}\lambda_i^{-1}f(\lambda_i,\bar{\lambda}_i)\,d\lambda_i\,d\bar{\lambda}_i \qquad (46) \\
&+ \varepsilon\int_{\varepsilon^{\rho}}^{\varepsilon^{\rho-n/(n-1)}}\int_{\varepsilon^{n/(n-1)}\bar{\lambda}_i}^{\varepsilon^{\rho}}\lambda_i^{-1}f(\lambda_i,\bar{\lambda}_i)\,d\lambda_i\,d\bar{\lambda}_i \qquad (47)
\end{aligned}$$

Now, since

$$\mathbb{E}[\bar{\lambda}^{-1}] = \mathbb{E}[\lambda_i^{-1/n}\bar{\lambda}_i^{-(n-1)/n}] < \infty$$

the asymptotic behavior of the joint distribution function $f(\lambda_i, \bar{\lambda}_i)$ as $\lambda_i, \bar{\lambda}_i \to 0$ is given by $O(\lambda_i^{-\rho'}\bar{\lambda}_i^{-\rho''})$ for some $0 < \rho' < 1 - 1/n$ and $\rho'' < 1/n$. Therefore, for sufficiently small ε and a suitable constant K_0, we have the following inequality:

$$f(\lambda_i, \bar{\lambda}_i) < K_0 \lambda_i^{-\rho'}\bar{\lambda}_i^{-\rho''}$$

Applying this inequality to (46) we obtain:

$$\begin{aligned}
\varepsilon \int_0^{\varepsilon^\rho} \int_{\varepsilon^{n/(n-1)}\bar{\lambda}_i}^{\varepsilon^\rho} \lambda_i^{-1} f(\lambda_i, \bar{\lambda}_i)\, d\lambda_i\, d\bar{\lambda}_i &\leq K_0 \varepsilon \int_0^{\varepsilon^\rho} \int_{\varepsilon^{n/(n-1)}\bar{\lambda}_i}^{\varepsilon^\rho} \lambda_i^{-1-\rho'}\bar{\lambda}_i^{-\rho''}\, d\lambda_i\, d\bar{\lambda}_i \\
&= \frac{K_0}{\rho'} \int_0^{\varepsilon^\rho} \bar{\lambda}_i^{-\rho''} \left[\varepsilon^{1-\rho' n/(n-1)}\bar{\lambda}_i^{-\rho'} - \varepsilon^{1-\rho\rho'}\right] d\bar{\lambda}_i \\
&\leq \frac{K_0}{\rho'} \int_0^{\varepsilon^\rho} \bar{\lambda}_i^{-\rho'-\rho''} \varepsilon^{1-\rho' n/(n-1)}\, d\bar{\lambda}_i \\
&= \frac{K_0}{\rho'(1-\rho'-\rho'')} \varepsilon^{1-\rho' n/(n-1)+\rho(1-\rho'-\rho'')} \qquad (48)
\end{aligned}$$

This terms approaches 0 as $\varepsilon \to 0$ since $1 - \rho' n/(n-1) + \rho(1 - \rho' - \rho'') > 0$ for every positive ρ.

Now we consider the asymptotic behavior of the joint distribution function $f(\lambda_i, \bar{\lambda}_i)$ as $\lambda_i \to 0$ and $\bar{\lambda}_i \to \infty$ with the assumption $\mathbb{E}[\bar{\lambda}^{-1}] < \infty$. Again, for sufficiently small ε and a suitable constant K_1, we have the following inequality:

$$f(\lambda_i, \bar{\lambda}_i) < K_1 \lambda_i^{-\rho'}\bar{\lambda}_i^{-\rho''}$$

where, in this case, we still have $0 < \rho' < 1 - 1/n$ but $\rho'' > 2 - 1/n$. Applying this inequality to (47) we obtain:

$$\begin{aligned}
\varepsilon \int_{\varepsilon^\rho}^{\varepsilon^{\rho - n/(n-1)}} \int_{\varepsilon^{n/(n-1)}\bar{\lambda}_i}^{\varepsilon^\rho} \lambda_i^{-1} f(\lambda_i, \bar{\lambda}_i)\, d\lambda_i\, d\bar{\lambda}_i &\leq \varepsilon \int_{\varepsilon^\rho}^{\infty} \int_{\varepsilon^{n/(n-1)}\bar{\lambda}_i}^{\varepsilon^\rho} \lambda_i^{-1} f(\lambda_i, \bar{\lambda}_i)\, d\lambda_i\, d\bar{\lambda}_i \qquad (49) \\
&\leq K_1 \varepsilon \int_{\varepsilon^\rho}^{\infty} \int_{\varepsilon^{n/(n-1)}\bar{\lambda}_i}^{\varepsilon^\rho} \lambda_i^{-1-\rho'}\bar{\lambda}_i^{-\rho''}\, d\lambda_i\, d\bar{\lambda}_i \\
&= \frac{K_1}{\rho'} \int_{\varepsilon^\rho}^{\infty} \bar{\lambda}_i^{-\rho''} \left[\varepsilon^{1-\rho' n/(n-1)}\bar{\lambda}_i^{-\rho'} - \varepsilon^{1-\rho\rho'}\right] d\bar{\lambda}_i \\
&\leq \frac{K_1}{\rho'} \int_{\varepsilon^\rho}^{\infty} \bar{\lambda}_i^{-\rho'-\rho''} \varepsilon^{1-\rho' n/(n-1)}\, d\bar{\lambda}_i \\
&= \frac{K_1}{\rho'(1-\rho'-\rho'')} \varepsilon^{1-\rho' n/(n-1)-\rho(\rho'+\rho''-1)} \qquad (50)
\end{aligned}$$

This terms approaches 0 as $\varepsilon \to 0$ since $1-\rho' n/(n-1) - \rho(\rho'+\rho''-1) > 0$ for $\rho < (1-\rho' n/(n-1))/(\rho'+\rho''-1)$ (note that $1-\rho' n/(n-1) > 0$ and $\rho'+\rho''-1 > 0$). ■

Now we turn to the proof of Proposition 6. The proof is divided in two steps:

1. Recalling that $C_{\text{delay}}(\gamma)$ is the solution of the equation $\gamma(R) = \gamma$, we obtain from Proposition 5:
$$\gamma(R) \le m 2^{R/m} \mathbb{E}[1/\bar{\lambda}_M] \Longrightarrow C_{\text{delay}}(\gamma) \ge m \log_2 \frac{\gamma}{m\mathbb{E}[1/\bar{\lambda}_M]}$$

2. Setting $\lambda_{\min} \triangleq \min_{k,i}\{\lambda_{k,i}\}$, we consider the region $\{\mathbf{\Lambda} : \lambda_{\min} > \bar{\lambda}_M 2^{-R/m}\}$. In this region, $|\mathcal{M}^*(R)| = Mm$ and $\gamma_{k,i}^{\text{lt}} > 0$ for all k, i. Then,
$$\begin{aligned} P_M(\mathbf{\Gamma}^{\text{lt}}(\mathbf{\Lambda}, R)) &\ge \frac{1}{M}\sum_{k=1}^{M}\sum_{i=1}^{m}\left(\frac{2^{R/m}}{\bar{\lambda}_M} - \frac{1}{\lambda_{k,i}}\right)1\{\lambda_{\min} > \bar{\lambda}_M 2^{-R/m}\} \\ &\ge m 2^{R/m}\left(\frac{1}{\bar{\lambda}_M} - \frac{2^{-R/m}}{\lambda_{\min}}\right)1\{\lambda_{\min} > \bar{\lambda}_M 2^{-R/m}\} \end{aligned} \tag{51}$$
By taking expectation of both sides of (51), we obtain
$$\gamma(R) \ge m 2^{R/m}\mathbb{E}\left[\frac{1}{\bar{\lambda}_M} 1\left\{\lambda_{\min} > \bar{\lambda}_M 2^{-R/m}\right\}\right] - 2^{-R/m}\mathbb{E}\left[\frac{1}{\lambda_{\min}} 1\left\{\lambda_{\min} > \bar{\lambda}_M 2^{-R/m}\right\}\right]$$
Now, we apply Lemma 2 to show that $2^{-R/m}\mathbb{E}[1/\lambda_{\min} 1\{\lambda_{\min} > \bar{\lambda} 2^{R/m}] \to 0$ as $R \to \infty$ and obtain the following asymptotic bound:
$$\gamma(R) \ge m 2^{R/m} \mathbb{E}[1/\bar{\lambda}_M] \Longrightarrow C_{\text{delay}}(\gamma) \le m \log_2 \frac{\gamma}{m\mathbb{E}[1/\bar{\lambda}_M]}$$

Combining the above results, we obtain the assertion. ■

D Proof of Proposition 8

Proof. Let us assume M fixed. Since $\theta_{k,i} = \lambda_{k,i}/m$, we have
$$\mathcal{M} = \{(k,i) : \xi^{\text{lt}}(\mathbf{\Lambda}) \ge \frac{1}{\lambda_{k,i}}\} = \{(k,i) : \lambda_{k,i} \ge \frac{1}{\xi^{\text{lt}}(\mathbf{\Lambda})}\} = \{(k,i) : \theta_{k,i} \ge \frac{1}{m\xi^{\text{lt}}(\mathbf{\Lambda})}\}$$
Let us define
$$\theta' \triangleq \max_{(k,i)\notin\mathcal{M}} \theta_{k,i} = \max\{\theta_{k,i} : \theta_{k,i} < \frac{1}{m\xi^{\text{lt}}(\mathbf{\Lambda})}\}$$

and

$$\theta'' \triangleq \min_{(k,i)\in\mathcal{M}} \theta_{k,i} = \min\{\theta_{k,i} : \theta_{k,i} \geq \frac{1}{m\xi^{\rm lt}(\mathbf{\Lambda})}\}$$

Since $m\xi^{\rm lt}(\mathbf{\Lambda}) = 2^{MR/|\mathcal{M}|}\prod_{(k,i)\in\mathcal{M}}\theta_{k,i}^{-1/|\mathcal{M}|}$, we have

$$\log_2\theta_{k,i} \overset{\geq}{<} -\frac{MR}{|\mathcal{M}|} + \frac{1}{|\mathcal{M}|}\sum_{(k,i)\in\mathcal{M}}\log_2\theta_{k,i} \iff (k,i) \overset{\in}{\notin} \mathcal{M}$$

Therefore, we have

$$\log_2\theta' < -\frac{MR}{|\mathcal{M}|} + \frac{1}{|\mathcal{M}|}\sum_{(k,i)\in\mathcal{M}}\log_2\theta_{k,i} \leq \log_2\theta''$$

Moreover, defining the empirical cumulative distribution function

$$F_m(\theta) \triangleq \frac{1}{Mm}\sum_{k=1}^{M}\sum_{i=1}^{m} 1\{\theta_{k,i} \leq \theta\}$$

we have $F_m(\theta) = 1 - |\mathcal{M}|/(Mm)$ for all $\theta \in (\theta', \theta'')$ and in particular for θ_0 such that

$$\log_2\theta' < \log_2\theta_0 = -\frac{MR}{|\mathcal{M}|} + \frac{1}{|\mathcal{M}|}\sum_{(k,i)\in\mathcal{M}}\log_2\theta_{k,i} = -\log_2(m\xi^{\rm lt}(\mathbf{\Lambda})) \leq \log_2\theta'' \qquad (52)$$

Substituting $|\mathcal{M}| = Mm(1 - F_m(\theta_0))$ in (52), we obtain

$$\log_2\theta_0 = -\frac{R/m}{1 - F_m(\theta_0)} + \frac{1}{|\mathcal{M}|}\sum_{(k,i)\in\mathcal{M}}\log_2\theta_{k,i}$$

with $\mathcal{M} = \{(k,i) : \theta_{k,i} \geq \theta_0\}$. From Fact 1, $F_m(\theta) \to F(\theta)$ as $m \to \infty$, independently of M. Therefore, letting $m \to \infty$ removes dependency from M and yields the following equation:

$$\log_2\theta_0 = -\lim_{m\to\infty}\frac{R/m}{1 - F_m(\theta_0)} + \int_{\theta_0}^{\infty}\log_2\theta\frac{dF(\theta)}{1 - F(\theta_0)}$$

where the second integral denotes the limiting expectation of $\log_2\theta_{k,i}$ conditioned on $\{\theta_{k,i} \geq \theta_0\}$. Thus, observing that R is the delay-limited capacity, we obtain

$$c(\gamma) = \lim_{m\to\infty}\frac{C_{\rm delay}(\gamma)}{m} = \int_{\theta_0}^{\infty}\log_2(\theta/\theta_0)dF(\theta) = \int_{0}^{\infty}(\log_2(\theta/\theta_0))_+dF(\theta) \qquad (53)$$

Finally, since

$$\mathbb{E}\left[\frac{1}{M}\sum_{k,i\in\mathcal{M}}\left(\xi^{\rm lt}(\mathbf{\Lambda}) - \frac{1}{\lambda_{k,i}}\right)\right] = \gamma$$

we have from the definition of θ_0, (52):

$$\lim_{m\to\infty}\mathbb{E}\left[\frac{1}{Mm}\sum_{k,i\in\mathcal{M}}\left(\frac{1}{\theta_0} - \frac{1}{\theta_{k,i}}\right)\right] = \int_{\theta_0}^{\infty}\left(\frac{1}{\theta_0} - \frac{1}{\theta}\right)dF(\theta) = \int_{0}^{\infty}\left(\frac{1}{\theta_0} - \frac{1}{\theta}\right)_+dF(\theta) = \gamma \qquad (54)$$

Comparing (53) and (54) to (4) and (5) as $m \to \infty$, the proof follows. ∎

References

[1] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communication aspects," *IEEE Trans. Inform. Theory*, Vol. 44, No. 6, October 1998.

[2] F. Boixadera Espax and J. J. Boutros, "Capacity considerations for wireless multiple-input multiple-output channels," *1999 Workshop on Multiaccess, Mobility and Teletraffic for Wireless Communications (MMT'99)*, Venezia, Italy, October 6–8, 1999.

[3] Z.D. Bai and Y.Q. Yin, "Limit of the smallest eigenvalue of a large dimensional sample covariance matrix," *Ann. Probab.*, vol. 21, pp. 12751294, 1993.

[4] G. Caire, R. Knopp and P. Humblet, "System capacity of F-TDMA cellular systems," *IEEE Trans. on Commun.*, Vol. 46, No. 12, pp. 1649–1661, August 1998.

[5] G. Caire, G. Taricco and E. Biglieri, "Optimal power control for minimum outage rate in wireless communications," *Proc. ICC '98*, Atlanta, GA, June 1998.

[6] G. Caire, G. Taricco and E. Biglieri, "Optimal power control for the fading channel," *IEEE Trans. on Inform. Theory*, Vol. 45, No. 5, pp. 1468–1489, July 1999.

[7] C.-N. Chuah, D. Tse, and J. M. Kahn, "Capacity of multi-antenna array systems in indoor wireless environment," *GLOBECOM'98*, Sidney, Australia, Nov. 8–12, 1998.

[8] I. Csiszár and J. Körner, *Information Theory*. New York: Academic Press, 1981.

[9] A. Edelman, *Eigenvalues and Condition Numbers of Random Matrices*. PhD Thesis, Dept. of Mathematics, Massachusetts Institute of Technology, Cambdridge, MA, 1989.

[10] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, Vol. 1, No. 2, pp. 41–59, Autumn 1996.

[11] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, Vol. 6, No. 3, pp. 311–335, March 1998.

[12] G. J. Foschini and R. A. Valenzuela, "Initial estimation of communication efficiency of indoor wireless channels," *Wireless Networks*, Vol. 3, No. 2, pp. 141–154, 1997.

[13] R. G. Gallager, *Information Theory and Reliable Communication*. New York: J. Wiley & Sons, 1968.

[14] R. M. Gray, *Probability, Random Processes, and Ergodic Properties*. New York: Springer Verlag, 1988.

[15] J.-C. Guey, M. P. Fitz, M. R. Bell, and W.-Y. Kuo, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," *IEEE Trans. Commun.*, Vol. 47, No. 4, pp. 527–537, April 1999.

[16] S. Hanly and D. Tse, "Multi-access fading channels. Part II: Delay-limited capacities," *IEEE Trans. Inform. Theory*, Vol. 44, No. 7, pp. 2816–2831, Nov. 1998.

[17] R. Horn and C. Johnson, *Matrix Analysis*. New York: Cambridge University Press, 1985.

[18] R. Knopp, *Coding and Multiple-Accessing over Fading Channels*, PhD dissertation, EPFL, Lausanne (Switzerland), and Institut Eurécom, Sophia Antipolis (France), 1997.

[19] R. Knopp and P. A. Humblet, "On coding for block-fading channels," submitted to *IEEE Trans. on Inform. Theory*, Dec. 1997.

[20] E. Malkamäki and H. Leib, "Coded diversity schemes on block fading Rayleigh channels," *Proc. of ICUPC'97*, San Diego, CA, Oct. 1997.

[21] E. Malkamäki and H. Leib, "Coded diversity on block-fading channels," *IEEE Trans. Inform. Theory*, Vol. 45, No. 2, pp. 771–781, March 1999.

[22] R. McEliece and W. Stark, "Channels with block interference," *IEEE Trans. Inform. Theory*, Vol. 30, No. 1, pp. 44–53, Jan. 1984.

[23] A. Narula, M. D. Trott, and G. Wornell, "Information theoretic analysis of multiple antenna transmitter diversity," *IEEE Trans. Inform. Theory*, Vol. 45, No. 7, pp. 2418–2433, Nov. 1999.

[24] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, Vol. 45, No. 1, pp. 139–157, January 1999.

[25] L. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Vehic. Tech.*, Vol. 43, No. 2, May 1994.

[26] G. Raleigh and J. Cioffi, "Spatio-temporal coding for wireless communication,"*IEEE Trans. on Commun.*, Vol. 46, No. 3, pp. 357 – 366, March 1998.

[27] T. Rappaport, *Wireless Communications*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[28] D. Shiu, G. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multi-element antenna systems," to appear in *IEEE Trans. Commun.*.

[29] D. Shiu and J. M. Kahn, "Power allocation strategies for wireless systems with multiple transmit antennas," submitted to *IEEE Trans. Commun.*

[30] D. Shiu and J. M. Kahn, "Design of high-throughput codes for multiple-antenna wireless systems," submitted to *IEEE Trans. Commun.*

[31] D. Shiu and J. M. Kahn, "Layered space-time codes for wireless communications using multiple transmit antennas," *IEEE International Conf. Telecomm. (ICC'99)*, Vancouver, BC, June 6–10, 1999.

[32] A. S. Stefanov and T. M. Duman, "Turbo coded modulation for systems with transmit and receive antenna diversity," *Proc. of GLOBECOM '99*, Rio de Janeiro (Brasil), 1999.

[33] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. on Inform. Theory* Vol. 44, No. 2, pp. 744 – 765, March 1998.

[34] "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, Vol. 45, No. 5, pp. 1456–1467, July 1999.

[35] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," to appear in *European Trans. Telecomm.*.

[36] D. Tse and S. Hanly, "Multi-access fading channels—Part I: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inform. Theory*, Vol. 44, No. 7, pp. 2796–2815, November 1998.

[37] J. Ventura-Traveset, G. Caire, E. Biglieri, and G. Taricco, "Impact of diversity reception on fading channels with coded modulation—Part I: Coherent detection," *IEEE Trans. Commun.* Vol. 45, No. 5, pp. 563–572, May 1997.

[38] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, Vol. 40, No. 4, pp. 1147–1157, July 1994.

[39] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inform. Theory*, Vol. 45, No. 2, pp. 622–640, March 1999.

[40] ARIB, *Personal Handy Phone System ARIB Standard, Version 2*, 1995.