Capacity Results for Block-Stationary Gaussian Fading Channels With a Peak Power Constraint

Jun Chen, *Member, IEEE*, and Venugopal V. Veeravalli, *Fellow, IEEE*

***Abstract*—A peak-power-limited single-antenna block-stationary Gaussian fading channel is studied, where neither the transmitter nor the receiver knows the channel state information, but both know the channel statistics. This model subsumes most previously studied Gaussian fading models. The asymptotic channel capacity in the high signal-to-noise ratio (SNR) regime is first computed, and it is shown that the behavior of the channel capacity depends critically on the channel model. For the special case where the fading process is symbol-by-symbol stationary, it is shown that the codeword length must scale at least logarithmically with SNR in order to guarantee that the communication rate can grow logarithmically with SNR with decoding error probability bounded away from one. An expression for the capacity per unit energy is also derived. Furthermore, it is shown that the capacity per unit energy is achievable using temporal ON–OFF signaling with optimally allocated ON symbols, where the optimal ON-symbol allocation scheme may depend on the peak power constraint.**

***Index Terms*—Asymptotic capacity, block fading, capacity per unit cost, noncoherent capacity, wireless channels.**

I. Introduction

THE capacity analysis of *noncoherent* fading channels has received considerable attention in recent years since it provides the ultimate limit on the rate of reliable communication on such channels. Proposed approaches to modeling noncoherent fading channels can be classified into two broad categories. The first is to model the fading process as a *block-independent* process. In the standard version of this model [1], the channel remains constant over blocks consisting of T symbol periods, and changes independently from block to block. The second is to model the fading process as a symbol-by-symbol *stationary* process. In this model, the independence assumption is removed, but the block structure is not allowed.[1] Somewhat surprisingly, these two models lead to very different capacity results. For the standard block-fading model, the capacity is shown [1], [2] to grow logarithmically with the signal-to-noise ratio (SNR), while for the symbol-by-symbol stationary model, the capacity grows only double-logarithmically in SNR at high SNR if the fading process is *regular* [3]–[5]. For symbol-by-symbol stationary fading channels, if the Lebesgue measure of the set of harmonics where the spectral density of the fading process is zero is positive, the fading process is *nonregular* and the capacity grows logarithmically with SNR [6], [7]. This result is consistent with the capacity result for block-independent fading channels in the sense that the logarithmic growth with SNR in the high-SNR regime results from the rank deficiency of the correlation matrix of the fading process. This point was elucidated in [8] where a *time-selective* block-fading model was considered in which the rank of the correlation within the block is allowed to be any number between one and the block length.

However, the mechanisms that cause the rank deficiency in the block-independent fading and nonregular symbol-by-symbol stationary models are different. For the block-independent fading model, the rank deficiency happens within each block. But for the nonregular symbol-by-symbol stationary fading channel model, the correlation matrix of the fading process over any finite block can still be full-rank; the rank deficiency in this case is in the asymptotic sense. In general, the rank deficiency of the correlation matrix can be affected by both the short time-scale correlation of the fading process as in the block-independent fading model and large time-scale correlation as in the symbol-by-symbol stationary channel model. In order to capture both of these effects, we model the fading process as a *block-stationary* Gaussian process.

The block-stationary model was introduced and justified in [8]. We summarize the main points of the justification here. In the block-independent fading model, the channel is assumed to change in an independent and identically distributed (i.i.d.) manner from block to block. The independence can be justified in certain time-division or frequency-hopping systems, where the blocks are separated sufficiently in time or frequency to undergo independent fading. The independence assumption is also convenient for information-theoretic analysis as it allows us to focus on one block in studying the capacity. If the blocks are not separated far enough in time or frequency, the fading process can be correlated across blocks and the block-stationary model is more appropriate in this scenario. Without time or frequency hopping, the channel variations from one block to the next are dictated by the long-term variations in the scattering environment. If we assume that the variations in average channel power are compensated for by other means such as power control, it is reasonable to model the variation from block to block as stationary and ergodic.

Manuscript received May 30, 2006; revised July 25, 2007. This work was supported by the National Science Foundation under Grants CCF 0049089 and CCF 0431088, through the University of Illinois at Urbana-Champaign. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Seattle, WA, July 2006.

J. Chen was with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA. He is now with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada (e-mail: junchen@ece.mcmaster.ca).

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

Communicated by G. Kramer, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2007.909083

[1]Here the block structure refers to the nature of the fading process, not the way the channel is used. For example, in a time-division system, each user may experience a block-stationary channel even if the underlying fading process is symbol-by-symbol stationary. However, the existing capacity results derived for the symbol-by-symbol stationary fading model do not apply to this scenario.

Remark 1: The block-stationary model does not imply that the fading process is stationary on a symbol-by-symbol basis as in the analysis of [3], [6]. But as explained in [8], the symbol-by-symbol stationary model is not realistic for time intervals that are larger than the time it takes for the mobile to traverse a distance that is of the order of a few carrier wavelengths. This is because the number of paths joining the transmitter and receiver, their strengths, and their relative delays can all change significantly for movements of the order of a few wavelengths. For this reason, it may be more accurate to model the fading process using a block-fading model with possible correlation across blocks than it is to model it as a symbol-by-symbol stationary process. From the viewpoint of analysis, the block-stationary model generalizes all previously considered models discussed above and therefore so do the capacity results for this model. More importantly, the block-stationary model provides us with a framework to study the interplay between many aspects of fading channels which are not captured in the symbol-by-symbol stationary and block-independent models, and allows us to identify the properties that are shared by the different models and the properties that are model dependent.

The channel capacity for the block-stationary model was only studied in [8] under certain constraints on the correlation structure across blocks, which essentially disallow rank deficiency over the large time scale. In this paper, we conduct a more complete study of the capacity for this channel model.

The remainder of this paper is organized as follows. In Section II, we describe the notation used in the paper and the system model. In Section III, we establish single-letter upper and lower bounds on channel capacity, and use these bounds to analyze the asymptotic capacity in the high-SNR regime. In Section IV, we discuss the robustness of the asymptotic capacity results, and the interplay between the codeword length, communication rate, and decoding error probability. In Section V, we adapt the formula of Verdú for capacity per unit cost [9] to our channel model, and use it to derive an expression for the capacity per unit energy in the presence of a peak power constraint. We summarize our results in Section VI.

II. Notation and System Model

A. Notation

The following notation is used in paper. For deterministic objects, upper case letters denote matrices, lower case letters denote scalars, and underlined lower case letters denote vectors. Random objects are identified by corresponding boldfaced letters. For example, $\boldsymbol{X}$ denotes a random matrix, X denotes the realization of $\boldsymbol{X}$, $\underline{\boldsymbol{x}}$ denotes a random vector, and $\boldsymbol{x}$ denotes a random scalar. For simplicity, sometimes we also use $\boldsymbol{x}^n$ to denote the random vector $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)^\top$. Although upper case letters are typically used for matrices, there are some exceptions, and these exceptions are noted explicitly in the paper. The operators $\det$, tr, $*$, $\top$, and $\dagger$ denote determinant, trace, conjugate, transpose, and conjugate transpose, respectively. For positive integer M, the $M \times M$ identity matrix is denoted by I_M, and for random vectors $\underline{\boldsymbol{a}}$ and $\underline{\boldsymbol{b}}$, $\mathbb{E}[(\underline{\boldsymbol{a}} - \mathbb{E}(\underline{\boldsymbol{a}}|\underline{\boldsymbol{b}}))(\underline{\boldsymbol{a}} - \mathbb{E}(\underline{\boldsymbol{a}}|\underline{\boldsymbol{b}}))^\dagger]$ is denoted by $\mathrm{var}(\underline{\boldsymbol{a}}|\underline{\boldsymbol{b}})$. Here $\mathbb{E}$ is the expectation operation and $\mathbb{E}(\underline{\boldsymbol{a}}|\underline{\boldsymbol{b}})$ denotes the conditional expectation of $\underline{\boldsymbol{a}}$ given $\underline{\boldsymbol{b}}$. All logarithms in this paper are to the natural base.

B. System Model

We consider a discrete-time channel whose time-t complex-valued output $\boldsymbol{y}_t \in \mathbb{C}$ is given by

$$\boldsymbol{y}_t = \boldsymbol{h}_t \boldsymbol{x}_t + \boldsymbol{z}_t \tag{1}$$

where $\boldsymbol{x}_t \in \mathbb{C}$ is the input at time t with peak power constraint $|\boldsymbol{x}_t|^2 \leq \mathsf{SNR}$, $\{\boldsymbol{h}_t\}$ models the fading process, and $\{\boldsymbol{z}_t\}$ models additive noise. We assume that the processes $\{\boldsymbol{h}_t\}$ and $\{\boldsymbol{z}_t\}$ are independent and have a joint distribution that does not depend on the input $\{\boldsymbol{x}_t\}$. We assume that $\{\boldsymbol{z}_t\}$ is a sequence of i.i.d. random variables with $\boldsymbol{z}_t \sim \mathcal{CN}(0,1)$, where we use the notation $\boldsymbol{v} \sim \mathcal{CN}(0,1)$ to indicate that $\boldsymbol{v}$ has a zero-mean unit-variance circularly symmetric complex-Gaussian distribution. We assume that the fading process $\{\boldsymbol{h}_t\}$ is a block-stationary process with $\boldsymbol{h}_t \sim \mathcal{CN}(0,1)$ and block length T, i.e., $\left\{\underline{\boldsymbol{h}}_k = (\boldsymbol{h}_{kT+1}, \boldsymbol{h}_{kT+2}, \ldots, \boldsymbol{h}_{kT+T})^\top\right\}_k$ is a vector-valued stationary process. Furthermore, we assume that $\{\underline{\boldsymbol{h}}_k\}$ is an ergodic process with a matrix spectral density function $S(e^{j\omega})$, $-\pi \leq \omega \leq \pi$. Specifically

$$S(e^{j\omega}) = \sum_{i=-\infty}^{\infty} R(i) e^{-j\omega i}$$

where $R(i) = \mathbb{E}\underline{\boldsymbol{h}}_k \underline{\boldsymbol{h}}_{k-i}^\dagger$. Since $R(i) = R^\dagger(-i)$, $i \in \mathbb{Z}$, it is not difficult to check that $S(e^{j\omega})$ is Hermitian, i.e., $S(e^{j\omega}) = S^\dagger(e^{j\omega})$. Moreover, we have $S(e^{j\omega}) \succcurlyeq 0$ $(-\pi \leq \omega \leq \pi)$, i.e., $S(e^{j\omega})$ is a positive semi-definite matrix.

There is an interesting relation between the matrix spectral density function and the asymptotic prediction error. Specifically, for the block stationary process $\left\{\boldsymbol{h}_t + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}$, define the following prediction error covariance matrices shown at the bottom of the page. Then $\Sigma(\mathsf{SNR})$ and $\Sigma(\infty)$ are related to the matrix spectral density function $S(e^{j\omega})$ of $\{\boldsymbol{h}_t\}$ as [10]

$$\det\left[\Sigma(\mathsf{SNR})\right] = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det\left[S(e^{j\omega}) + \frac{1}{\mathsf{SNR}} I_T\right] \mathrm{d}\omega\right\} \tag{2}$$

$$\det\left[\Sigma(\infty)\right] = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det\left[S(e^{j\omega})\right] \mathrm{d}\omega\right\}. \tag{3}$$

$$\Sigma(\mathsf{SNR}) \triangleq \mathrm{var}\left(\left(\boldsymbol{h}_1 + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_1, \boldsymbol{h}_2 + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_2, \ldots, \boldsymbol{h}_T + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_T\right)^\top \middle| \left\{\boldsymbol{h}_t + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{0}\right)$$

$$\Sigma(\infty) \triangleq \mathrm{var}\left((\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_T)^\top \middle| \{\boldsymbol{h}_t\}_{t=-\infty}^{0}\right).$$

III. Asymptotic Capacity at High SNR

We denote the capacity with peak power constraint SNR by $C(\mathsf{SNR})$. For any $n \in \mathbb{N}$ and $\mathsf{SNR} > 0$, let

$$\mathbb{D}_n(\mathsf{SNR}) = \left\{ x^n \in \mathbb{C}^n : \max_{1 \le t \le n} |x_t|^2 \le \mathsf{SNR} \right\}.$$

Let $\mathcal{P}_n(\mathsf{SNR})$ be the set of probability distributions on $\mathbb{D}_n(\mathsf{SNR})$. Since the channel is block-wise stationary and ergodic, a coding theorem exists [11], [12] and we have

$$C(\mathsf{SNR}) = \lim_{n\to\infty} \sup_{p_{\boldsymbol{x}^n} \in \mathcal{P}_n(\mathsf{SNR})} \frac{1}{n} I(\boldsymbol{x}^n; \boldsymbol{y}^n).$$

A. Lower Bound and Upper Bound

To derive a lower bound on $C(\mathsf{SNR})$ for the channel model given in (1), we adopt the interleaved decision-oriented training scheme proposed in [13] with some modifications. This scheme can also be viewed as a way of interpreting the computations in [3, Sec. IV-E].

Let $p_{\boldsymbol{x}}$ be a probability distribution on $\mathbb{C}$ with $|\boldsymbol{x}|^2 \in [x_{\min}^2, \mathsf{SNR}]$. Construct the codebook $\mathcal{C} = \mathcal{C}_1 \times \mathcal{C}_2 \times \cdots \times \mathcal{C}_K$ with K subcodebooks $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_K$, where codebook $\mathcal{C}_i$ $(i = 1, \ldots, K)$ contains 2^{nR_i} codewords of length n generated independently symbol by symbol using distribution $p_{\boldsymbol{x}}$. Assume that K is a multiple of the block length T, i.e., $K = rT$ for some positive integer r. Now we multiplex (or interleave) these K codebooks. Specifically, codebook $\mathcal{C}_i$ $(i = 1, \ldots, K)$ is used at time instants $i, i+K, i+2K, \ldots, i+(n-1)K$. For codebook $\mathcal{C}_1$, its codeword can be successfully decoded if

$$R_1 \le \frac{1}{n} I\left(\{\boldsymbol{x}_{1+jK}\}_{j=0}^{n-1}; \{\boldsymbol{y}_{1+jK}\}_{j=0}^{n-1}\right)$$

for sufficiently large n. Furthermore, using the facts that $\{\boldsymbol{x}_{1+jK}\}_{j=0}^{n-1}$ are i.i.d. and that the channel is stationary over time instants $1, 1+K, 1+2K, \ldots, 1+(n-1)K$, we get

$$\begin{aligned} &\frac{1}{n} I\left(\{\boldsymbol{x}_{1+jK}\}_{j=0}^{n-1}; \{\boldsymbol{y}_{1+jK}\}_{j=0}^{n-1}\right) \\ &= \frac{1}{n} \sum_{j=0}^{n-1} I\left(\boldsymbol{x}_{1+jK}; \{\boldsymbol{y}_{1+iK}\}_{i=0}^{n-1} | \{\boldsymbol{x}_{1+iK}\}_{i=j+1}^{n-1}\right) \\ &= \frac{1}{n} \sum_{j=0}^{n-1} I\left(\boldsymbol{x}_{1+jK}; \{\boldsymbol{y}_{1+iK}\}_{i=0}^{n-1}, \{\boldsymbol{x}_{1+iK}\}_{i=j+1}^{n-1}\right) \\ &\ge \frac{1}{n} \sum_{j=0}^{n-1} I(\boldsymbol{x}_{1+jK}; \boldsymbol{y}_{1+jK}) \\ &= I(\boldsymbol{x}_1; \boldsymbol{y}_1). \end{aligned} \tag{4}$$

This is to be expected since a channel with memory has a higher reliable communication rate than the memoryless channel with the same marginal transition probability. Thus, reliable communication at rate $R_1 = I(\boldsymbol{x}_1; \boldsymbol{y}_1)$ is possible for subcodebook $\mathcal{C}_1$. After $\{\boldsymbol{x}_{1+jK}\}_{j=0}^{n-1}$ is successfully decoded, the receiver can use these values as well as $\{\boldsymbol{y}_{1+jK}\}_{j=0}^{n-1}$ to estimate $\{\boldsymbol{h}_{2+jK}\}_{j=0}^{n-1}$. Specifically, $(\boldsymbol{x}_{1+jK}, \boldsymbol{y}_{1+jK})$ is used to estimate $\boldsymbol{h}_{2+jK}, j = 0, 1, \ldots, n-1$. To facilitate the calculation, we assume [2] that $\boldsymbol{h}_{1+jK} + \frac{1}{x_{\min}} \boldsymbol{z}_{1+jK}$ is used to estimate $\boldsymbol{h}_{2+jK}$ by forming the minimum mean-square error (MMSE) estimate

$$\mathbb{E}\left(\boldsymbol{h}_{2+jK} \middle| \boldsymbol{h}_{1+jK} + \frac{1}{x_{\min}} \boldsymbol{z}_{1+jK}\right), \qquad j = 0, 1, \ldots, n-1.$$

The receiver decodes the codeword in codebook $\mathcal{C}_2$ using $\left\{\mathbb{E}\left(\boldsymbol{h}_{2+jK} \middle| \boldsymbol{h}_{1+jK} + \frac{1}{x_{\min}} \boldsymbol{z}_{1+jK}\right)\right\}_{j=0}^{n-1}$ as side information. Successful decoding is possible if

$$\begin{aligned} R_2 \le \frac{1}{n} \sum_{j=0}^{n-1} I\Bigg(&\{\boldsymbol{x}_{2+jK}\}_{j=0}^{n-1}; \{\boldsymbol{y}_{2+jK}\}_{j=0}^{n-1}, \\ &\left\{\mathbb{E}\left(\boldsymbol{h}_{2+jK} \middle| \boldsymbol{h}_{1+jK} + \frac{1}{x_{\min}} \boldsymbol{z}_{1+jK}\right)\right\}_{j=0}^{n-1}\Bigg). \end{aligned}$$

Similar to (4), we can use the following lower bound to show that reliable communication at rate

$$R_2 = I\left(\boldsymbol{x}_2; \boldsymbol{y}_2 \middle| \mathbb{E}\left(\boldsymbol{h}_2 \middle| \boldsymbol{h}_1 + \frac{1}{x_{\min}} \boldsymbol{z}_1\right)\right)$$

is possible for subcodebook $\mathcal{C}_2$

$$\begin{aligned} &\frac{1}{n} \sum_{j=0}^{n-1} I\Bigg(\{\boldsymbol{x}_{2+jK}\}_{j=0}^{n-1}; \{\boldsymbol{y}_{2+jK}\}_{j=0}^{n-1}, \\ &\qquad \left\{\mathbb{E}\left(\boldsymbol{h}_{2+jK} \middle| \boldsymbol{h}_{1+jK} + \frac{1}{x_{\min}} \boldsymbol{z}_{1+jK}\right)\right\}_{j=0}^{n-1}\Bigg) \\ &\ge I\left(\boldsymbol{x}_2; \boldsymbol{y}_2, \mathbb{E}\left(\boldsymbol{h}_2 \middle| \boldsymbol{h}_1 + \frac{1}{x_{\min}} \boldsymbol{z}_1\right)\right) \\ &= I\left(\boldsymbol{x}_2; \boldsymbol{y}_2 \middle| \mathbb{E}\left(\boldsymbol{h}_2 \middle| \boldsymbol{h}_1 + \frac{1}{x_{\min}} \boldsymbol{z}_1\right)\right). \end{aligned}$$

By applying this procedure successively, we can conclude that for codebook $\mathcal{C}_i$, reliable communication is possible at rate

$$R_i = I\left(\boldsymbol{x}_i; \boldsymbol{y}_i \middle| \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_j + \frac{1}{x_{\min}} \boldsymbol{z}_j\right\}_{j=1}^{i-1}\right)\right), \quad i = 1, 2, \ldots, K.$$

[2]This assumption can be justified by the following argument. The receiver first generates a sequence of i.i.d. circularly symmetric complex Gaussian random variables $\{\boldsymbol{n}_{1+jK}\}_{j=0}^{n-1}$ (with $\boldsymbol{n}_{1+jK} \sim \mathcal{CN}(0,1)$) that are independent of everything else. Since $|\boldsymbol{x}_{1+jK}|^2 \in [x_{\min}^2, \mathsf{SNR}]$, the receiver can construct $\boldsymbol{h}_{1+jK} + \frac{1}{x_{\min}} \boldsymbol{z}'_{1+jK}$ using $(\boldsymbol{x}_{1+jK}, \boldsymbol{y}_{1+jK}, \boldsymbol{n}_{1+jK})$, where

$$\boldsymbol{z}'_{1+jK} = \frac{x_{\min}}{\boldsymbol{x}_{1+jK}} \boldsymbol{z}_{1+jK} + x_{\min} \sqrt{\frac{1}{x_{\min}^2} - \frac{1}{|\boldsymbol{x}_{1+jK}|^2}}\, \boldsymbol{n}_{1+jK}, \quad j = 0, 1, \ldots, n-1.$$

It can be verified that the estimation based on $\boldsymbol{h}_{1+jK} + \frac{1}{x_{\min}} \boldsymbol{z}_{1+jK}$ is equivalent to that based on $\boldsymbol{h}_{1+jK} + \frac{1}{x_{\min}} \boldsymbol{z}'_{1+jK}$, $j = 0, 1, \ldots, n-1$.

Thus, using this interleaved decision-oriented training scheme, we can have reliable communication at an overall rate of

$$R = \frac{1}{K}\sum_{i=1}^{K} R_i = \frac{1}{K}\sum_{i=1}^{K} I\left(\boldsymbol{x}_i;\boldsymbol{y}_i \middle| \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_j + \frac{1}{x_{\min}}\boldsymbol{z}_j\right\}_{j=1}^{i-1}\right)\right).$$

We show in Appendix I that

$$\left\{I\left(\boldsymbol{x}_{i+jT};\boldsymbol{y}_{i+jT} \middle| \mathbb{E}\left(\boldsymbol{h}_{i+jT} \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=1}^{i+jT-1}\right)\right)\right\}_j$$

is a monotone increasing sequence with the limit shown in the first equation at the bottom of the page. Now we let K go to infinity (i.e., we let $r \to \infty$ since T is fixed), and we obtain the second equation at the bottom of the page. This yields the single-letter lower bound

$$C(\mathsf{SNR}) \geq \frac{1}{T}\sum_{i=1}^{T} I\left(\boldsymbol{x}_i;\boldsymbol{h}_i\boldsymbol{x}_i + \boldsymbol{z}_i \middle| \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right)\right) \quad (5)$$

where $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T$ all have the same distribution $p_{\boldsymbol{x}}$, which is to be optimized later.

Remark 2: Although channel estimation and communication are intertwined in this interleaved decision-oriented training scheme, the effect of channel memory is isolated from channel coding through interleaving. This is because when K is large enough, $\boldsymbol{h}_i, \boldsymbol{h}_{i+K}, \boldsymbol{h}_{i+2K}, \ldots, \boldsymbol{h}_{i+(n-1)K}$ are roughly independent. Thus, the codeword in codebook $\mathcal{C}_i$, which is transmitted over time instants $i, i+K, i+2K, \ldots, i+(n-1)K$, essentially experiences a memoryless channel. This also suggests that as K goes to infinity, the single-letter lower bound (5) provides a correct estimate of the rate supported by this interleaved decision-oriented training scheme. We can see that the channel memory manifests itself in the lower bound (5) only through

$$\mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right).$$

Furthermore, in (5), we can write $\boldsymbol{h}_i$ as the sum of two independent random variables: the coherent fading component

$$\mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right)$$

which is known to the receiver, and the noncoherent fading component

$$\boldsymbol{h}_i - \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right)$$

which is unknown. Isolating the effect of channel memory facilitates the channel code design: we only need to design channel codes for memoryless fading channels with different coherent and noncoherent components, instead of designing different codes for channels with different memory structures.

To derive a single-letter upper bound on $C(\mathsf{SNR})$, we follow the approach in [6]. The capacity $C(\mathsf{SNR})$ is given by

$$C(\mathsf{SNR}) = \lim_{n\to\infty} \sup_{p_{\boldsymbol{x}^n} \in \mathcal{P}_n(\mathsf{SNR})} \frac{1}{n} I(\boldsymbol{x}^n;\boldsymbol{y}^n).$$

By the chain rule

$$I(\boldsymbol{x}^n;\boldsymbol{y}^n) = \sum_{k=1}^{n} I(\boldsymbol{x}^n;\boldsymbol{y}_k|\boldsymbol{y}^{k-1}).$$

$$\lim_{j\to\infty} I\left(\boldsymbol{x}_{i+jT};\boldsymbol{y}_{i+jT} \middle| \mathbb{E}\left(\boldsymbol{h}_{i+jT} \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=1}^{i+jT-1}\right)\right) = I\left(\boldsymbol{x}_i;\boldsymbol{y}_i \middle| \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right)\right).$$

$$\begin{aligned}
\lim_{K\to\infty} R &= \lim_{K\to\infty} \frac{1}{K}\sum_{i=1}^{K} I\left(\boldsymbol{x}_i;\boldsymbol{y}_i \middle| \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_j + \frac{1}{x_{\min}}\boldsymbol{z}_j\right\}_{j=1}^{i-1}\right)\right) \\
&= \lim_{K\to\infty} \frac{1}{K}\sum_{i=1}^{T}\sum_{j=0}^{r-1} I\left(\boldsymbol{x}_{i+jT};\boldsymbol{y}_{i+jT} \middle| \mathbb{E}\left(\boldsymbol{h}_{i+jT} \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=1}^{i+jT-1}\right)\right) \\
&= \frac{1}{T}\sum_{i=1}^{T}\left[\lim_{r\to\infty} \frac{1}{r}\sum_{j=0}^{r-1} I\left(\boldsymbol{x}_{i+jT};\boldsymbol{y}_{i+jT} \middle| \mathbb{E}\left(\boldsymbol{h}_{i+jT} \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=1}^{i+jT-1}\right)\right)\right] \\
&= \frac{1}{T}\sum_{i=1}^{T} I\left(\boldsymbol{x}_i;\boldsymbol{y}_i \middle| \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right)\right) \\
&= \frac{1}{T}\sum_{i=1}^{T} I\left(\boldsymbol{x}_i;\boldsymbol{h}_i\boldsymbol{x}_i + \boldsymbol{z}_i \middle| \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right)\right).
\end{aligned}$$

We can upper-bound $I\left(\boldsymbol{x}^n; \boldsymbol{y}_k | \boldsymbol{y}^{k-1}\right)$ as shown at the bottom of the page. Since

$$\left(\boldsymbol{x}_k, \mathbb{E}\left(\boldsymbol{h}_k \left| \left\{\boldsymbol{h}_t + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{k-1}\right.\right)\right)$$

is a sufficient statistic for estimating $\boldsymbol{y}_k$ from

$$\left(\boldsymbol{x}_k, \left\{\boldsymbol{h}_t + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{k-1}\right)$$

it follows that

$$I\left(\boldsymbol{x}_k, \left\{\boldsymbol{h}_t + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{k-1}; \boldsymbol{y}_k\right) = I\left(\boldsymbol{x}_k, \mathbb{E}\left(\boldsymbol{h}_k \left| \left\{\boldsymbol{h}_t + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{k-1}\right.\right); \boldsymbol{y}_k\right).$$

Note that by the block stationarity of the fading process,

$$\sup_{p_{\boldsymbol{x}_k} \in \mathcal{P}_1(\mathsf{SNR})} I\left(\boldsymbol{x}_k, \mathbb{E}\left(\boldsymbol{h}_k \left| \left\{\boldsymbol{h}_t + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{k-1}\right.\right); \boldsymbol{y}_k\right)$$

depends on k only through $(k \bmod T)$. Therefore, we obtain the single-letter upper bound

$$C(\mathsf{SNR}) \leq \frac{1}{T}\sum_{k=1}^{T} \sup_{p_{\boldsymbol{x}_k} \in \mathcal{P}_1(\mathsf{SNR})} I\left(\boldsymbol{x}_k, \mathbb{E}\left(\boldsymbol{h}_k \left| \left\{\boldsymbol{h}_t + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{k-1}\right.\right); \boldsymbol{y}_k\right). \quad (8)$$

B. Asymptotic Analysis

Now we proceed to show that the lower bound (5) and upper bound (8) together characterize the asymptotic behavior of $C(\mathsf{SNR})$ in the high-SNR regime.

Lemma 1: For every $\xi \in [\xi_0, \xi_1]$, let $A(\xi)$ be an $M \times M$ symmetric positive semidefinite matrix. We have

$$\lim_{\epsilon \to 0} \frac{\int_{\xi_0}^{\xi_1} \log\det\left[A(\xi) + \epsilon I_M\right] \mathrm{d}\xi}{\log \epsilon} = \sum_{i=0}^{M}(M-i)\mu\left(\mathrm{rank}\left(A(\xi)\right) = i\right)$$

where $\mu(\mathrm{rank}(A(\xi)) = i)$ is the Lebesgue measure of the set $\{\xi : \mathrm{rank}(A(\xi)) = i\}$.

For the special case where $A(\xi) = A$ for all $\xi \in [\xi_0, \xi_1]$ and $\xi_1 - \xi_0 = 1$, we get

$$\lim_{\epsilon \to 0} \frac{\log\det\left[A + \epsilon I_M\right]}{\log \epsilon} = M - \mathrm{rank}(A).$$

Proof: See Appendix II. □

Lemma 2 ([6, Sec. IV]): If $\boldsymbol{x}$ is uniformly distributed over the set $\{z \in \mathbb{C} : \frac{\sqrt{\mathsf{SNR}}}{2} \leq |z| \leq \sqrt{\mathsf{SNR}}\}$, $\widehat{\boldsymbol{h}} \sim \mathcal{CN}\left(0, 1 - \mathbb{E}\left|\widetilde{\boldsymbol{h}}\right|^2\right)$, $\widetilde{\boldsymbol{h}} \sim \mathcal{CN}\left(0, \mathbb{E}|\widetilde{\boldsymbol{h}}|^2\right)$, $\boldsymbol{z} \sim \mathcal{CN}(0,1)$; and $\boldsymbol{x}$, $\widehat{\boldsymbol{h}}$, $\widetilde{\boldsymbol{h}}$, $\boldsymbol{z}$ are all independent, then

$$I\left(\boldsymbol{x}; \left(\widehat{\boldsymbol{h}} + \widetilde{\boldsymbol{h}}\right)\boldsymbol{x} + \boldsymbol{z} \middle| \widehat{\boldsymbol{h}}\right) \geq -\log\left(\mathbb{E}\left|\widetilde{\boldsymbol{h}}\right|^2 + \frac{8}{5\mathsf{SNR}}\right) + \log\left(1 - \mathbb{E}\left|\widetilde{\boldsymbol{h}}\right|^2\right) - \gamma - \log\frac{5e}{6} \quad (9)$$

where γ is the Euler constant.

$$\begin{aligned}
I(\boldsymbol{x}^n; \boldsymbol{y}_k | \boldsymbol{y}^{k-1}) &= I(\boldsymbol{x}^n, \boldsymbol{y}^{k-1}; \boldsymbol{y}_k) - I(\boldsymbol{y}_k; \boldsymbol{y}^{k-1}) \\
&\leq I(\boldsymbol{x}^n, \boldsymbol{y}^{k-1}; \boldsymbol{y}_k) \\
&= I(\boldsymbol{x}^k, \boldsymbol{y}^{k-1}; \boldsymbol{y}_k) \\
&= h(\boldsymbol{y}_k) - h(\boldsymbol{y}_k | \boldsymbol{x}^k, \boldsymbol{y}^{k-1}) \\
&= h(\boldsymbol{y}_k) - \int_{\mathbb{D}_k(\mathsf{SNR})} p_{\boldsymbol{x}^k}\left(\boldsymbol{x}^k = x^k\right) h(\boldsymbol{y}_k | \boldsymbol{x}^k = x^k, \boldsymbol{y}^{k-1}) \mathrm{d}x^k \\
&= h(\boldsymbol{y}_k) - \int_{\mathbb{D}_k(\mathsf{SNR})} p_{\boldsymbol{x}^k}\left(\boldsymbol{x}^k = x^k\right) \log\left(\pi e \,\mathrm{var}\left(\boldsymbol{y}_k \left| \boldsymbol{x}^k = x^k, \boldsymbol{y}^{k-1}\right.\right)\right) \mathrm{d}x^k \\
&\leq h(\boldsymbol{y}_k) - \int_{\mathbb{D}_k(\mathsf{SNR})} p_{\boldsymbol{x}^k}\left(\boldsymbol{x}^k = x^k\right) \log\left(\pi e \,\mathrm{var}\left(\boldsymbol{y}_k \left| \boldsymbol{x}_1 = \sqrt{\mathsf{SNR}}, \ldots, \boldsymbol{x}_{k-1} = \sqrt{\mathsf{SNR}}, \boldsymbol{x}_k = x_k, \boldsymbol{y}^{k-1}\right.\right)\right) \mathrm{d}x^k \\
&= h(\boldsymbol{y}_k) - \int_{\mathbb{D}_1(\mathsf{SNR})} p_{\boldsymbol{x}_k}\left(\boldsymbol{x}_k = x_k\right) \log\left(\pi e \,\mathrm{var}\left(\boldsymbol{y}_k \left| \boldsymbol{x}_1 = \sqrt{\mathsf{SNR}}, \ldots, \boldsymbol{x}_{k-1} = \sqrt{\mathsf{SNR}}, \boldsymbol{x}_k = x_k, \boldsymbol{y}^{k-1}\right.\right)\right) \mathrm{d}x_k \\
&= h(\boldsymbol{y}_k) - h\left(\boldsymbol{y}_k \left| \boldsymbol{x}_k, \boldsymbol{h}_{k-1} + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_{k-1}, \ldots, \boldsymbol{h}_1 + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_1\right.\right) \\
&= I\left(\boldsymbol{x}_k, \boldsymbol{h}_{k-1} + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_{k-1}, \ldots, \boldsymbol{h}_1 + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_1; \boldsymbol{y}_k\right) && (6) \\
&\leq I\left(\boldsymbol{x}_k, \left\{\boldsymbol{h}_t + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{k-1}; \boldsymbol{y}_k\right). && (7)
\end{aligned}$$

Theorem 1: For the block-stationary Gaussian fading channel model given in (1)

$$\lim_{\mathsf{SNR}\to\infty}\frac{C(\mathsf{SNR})}{\log\mathsf{SNR}}=\lim_{\mathsf{SNR}\to\infty}\frac{-\log\det[\Sigma(\mathsf{SNR})]}{T\log\mathsf{SNR}}=\frac{1}{2\pi T}\sum_{i=0}^{T}(T-i)\mu\left(\mathrm{rank}(S(e^{j\omega}))=i\right). \quad (10)$$

Remark 3: The second equality in (10) follows from (2) and Lemma 1.

Proof: Below we provide an intuitive explanation of this theorem based on the lower bound (5). The details of the proof are left to Appendix III.

In the lower bound (5), let $\boldsymbol{x}_i$ be uniformly distributed over the set $\left\{z\in\mathbb{C}:\frac{\sqrt{\mathsf{SNR}}}{2}\le|z|\le\sqrt{\mathsf{SNR}}\right\}$, and write $\boldsymbol{h}_i$ as $\boldsymbol{h}_i=\widehat{\boldsymbol{h}}_i+\widetilde{\boldsymbol{h}}_i$ where

$$\widehat{\boldsymbol{h}}_i=\mathbb{E}\left(\boldsymbol{h}_i\left|\left\{\boldsymbol{h}_k+\frac{2}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right.\right).$$

Suppose $\mathbb{E}|\widetilde{\boldsymbol{h}}_i|^2\approx\mathsf{SNR}^{-r_i}$, $i=1,2,\ldots,T$. We can then write $\boldsymbol{y}_i$ as $\boldsymbol{y}_i=\widehat{\boldsymbol{h}}_i\boldsymbol{x}_i+\boldsymbol{w}_i$, where $\boldsymbol{w}_i=\widetilde{\boldsymbol{h}}_i\boldsymbol{x}_i+\boldsymbol{z}_i$ with $\mathbb{E}|\boldsymbol{w}_i|^2\approx\mathsf{SNR}^{1-r_i}$. By viewing $\boldsymbol{y}_i$ as the output of a coherent fading channel with the fading $\widehat{\boldsymbol{h}}_i$, which is known at the receiver, and noise $\boldsymbol{w}_i$, we get

$$\begin{aligned}I&\left(\boldsymbol{x}_i;\boldsymbol{h}_i\boldsymbol{x}_i+\boldsymbol{z}_i\left|\mathbb{E}\left(\boldsymbol{h}_i\left|\left\{\boldsymbol{h}_k+\frac{2}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right.\right)\right.\right)\\&=I\left(\boldsymbol{x}_i;\widehat{\boldsymbol{h}}_i\boldsymbol{x}_i+\boldsymbol{w}_i\left|\widehat{\boldsymbol{h}}_i\right.\right)\\&\approx\log\frac{\mathsf{SNR}}{\mathsf{SNR}^{1-r_i}}\\&=r_i\log\mathsf{SNR}.\end{aligned}$$

Thus, the lower bound (5) can be approximated by

$$\frac{1}{T}\sum_{i=1}^{T}I\left(\boldsymbol{x}_i;\boldsymbol{h}_i\boldsymbol{x}_i+\boldsymbol{z}_i\left|\mathbb{E}\left(\boldsymbol{h}_i\left|\left\{\boldsymbol{h}_k+\frac{2}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right.\right)\right.\right)\approx\frac{1}{T}\sum_{i=1}^{T}r_i\log\mathsf{SNR}.$$

We then complete the proof by showing that $\sum_{i=1}^{T}r_i$ is related to the matrix spectral density function $S(e^{j\omega})$ through the equation

$$\sum_{i=1}^{T}r_i=\frac{1}{2\pi}\sum_{i=0}^{T}(T-i)\mu\left(\mathrm{rank}(S(e^{j\omega}))=i\right). \qquad \square$$

Theorem 1 generalizes many previous results on the noncoherent capacity for Gaussian channels in the high-SNR regime as we illustrate in Section III-C.

C. Previous Results as Special Cases of Theorem 1

Example 1: Constant Fading Within Block: For the special case where the fading remains constant within a block, i.e., $\boldsymbol{h}_{kT+1}=\boldsymbol{h}_{kT+2}=\cdots=\boldsymbol{h}_{kT+T}$, for all $k\in\mathbb{Z}$, all the entries of $R(i)$, for any fixed i, are identical. This implies that, for any fixed ω, all the entries of $S(e^{j\omega})$ are identical, which we shall denote by $s(e^{j\omega})$. It is easy to see that $s(e^{j\omega})$ is essentially the spectral density function of $\{\boldsymbol{h}_{kT}\}_k$. The rank of $S(e^{j\omega})$ is 1 if $s(e^{j\omega})>0$, and is 0 if $s(e^{j\omega})=0$. We therefore have

$$\begin{aligned}\lim_{\mathsf{SNR}\to\infty}\frac{C(\mathsf{SNR})}{\log\mathsf{SNR}}&=\frac{1}{2\pi T}\sum_{i=0}^{T}(T-i)\mu(\mathrm{rank}(S(e^{j\omega}))=i)\\&=\frac{1}{2\pi T}\sum_{i=0}^{1}(T-i)\mu(\mathrm{rank}(S(e^{j\omega}))=i)\\&=1-\frac{\mu(s(e^{j\omega})>0)}{2\pi T}.\end{aligned}$$

When $T=1$, we recover the result in [6] that

$$\lim_{\mathsf{SNR}\to\infty}\frac{C(\mathsf{SNR})}{\log\mathsf{SNR}}=\frac{\mu(s(e^{j\omega})=0)}{2\pi} \quad (11)$$

which illustrates the effect of large time scale correlation of the fading process on the pre-log term of the channel capacity in the high-SNR regime. When the fading is independent from block to block, we have $\mu(s(e^{j\omega})>0)=2\pi$, and thus recover the result in [1], [2] that

$$\lim_{\mathsf{SNR}\to\infty}\frac{C(\mathsf{SNR})}{\log\mathsf{SNR}}=\frac{T-1}{T}$$

which illustrates the effect of short time scale correlation of the fading process on the pre-log term of the capacity at high SNR.

Example 2: Time-Selectivity Within Block: In this example, we recover the main result in [8] concerning the case where rank deficiency is caused purely by the correlation within a block. If $\mathrm{rank}(\Sigma(\infty))=\mathrm{rank}(R(0))$, then[3]

$$\lim_{\mathsf{SNR}\to\infty}\frac{C(\mathsf{SNR})}{\log\mathsf{SNR}}=\frac{T-\mathrm{rank}(R(0))}{T}. \quad (12)$$

To prove (12), we first note the equation at the top of the following page, where $\preccurlyeq$ denotes positive semidefinite ordering: $A\preccurlyeq B$ means that $B-A$ is a positive semidefinite matrix. We therefore have the bound

$$\begin{aligned}&\lim_{\mathsf{SNR}\to\infty}\frac{\log\det\left[\Sigma(\infty)+\frac{1}{\mathsf{SNR}}I_T\right]}{\log\mathsf{SNR}}\\&\quad\le\lim_{\mathsf{SNR}\to\infty}\frac{\log\det\Sigma(\mathsf{SNR})}{\log\mathsf{SNR}}\\&\quad\le\lim_{\mathsf{SNR}\to\infty}\frac{\log\det\left[R(0)+\frac{1}{\mathsf{SNR}}I_T\right]}{\log\mathsf{SNR}}.\end{aligned}$$

By Lemma 1

$$\begin{aligned}&\lim_{\mathsf{SNR}\to\infty}\frac{\log\det\left[\Sigma(\infty)+\frac{1}{\mathsf{SNR}}I_T\right]}{\log\mathsf{SNR}}\\&\quad=\lim_{\mathsf{SNR}\to\infty}\frac{\log\det\left[R(0)+\frac{1}{\mathsf{SNR}}I_T\right]}{\log\mathsf{SNR}}\\&\quad=-T+\mathrm{rank}(R(0))\end{aligned}$$

which implies that

$$\lim_{\mathsf{SNR}\to\infty}\frac{\log\det\Sigma(\mathsf{SNR})}{\log\mathsf{SNR}}=-T+\mathrm{rank}(R(0)).$$

[3]The condition $\mathrm{rank}(\Sigma(\infty))=\mathrm{rank}(R(0))$ is satisfied, for instance, when the fading process is independent from block to block.

$$\begin{aligned}\Sigma(\infty) + \frac{1}{\mathsf{SNR}} I_T &= \mathrm{var}\left((\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_T)^\top \middle| \{\boldsymbol{h}_t\}_{k=-\infty}^{0} \right) + \frac{1}{\mathsf{SNR}} I_T \\ &= \mathrm{var}\left(\left(\boldsymbol{h}_1 + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_1, \ldots, \boldsymbol{h}_T + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_T \right)^\top \middle| \{\boldsymbol{h}_k\}_{k=-\infty}^{0} \right) \\ &\preccurlyeq \mathrm{var}\left(\left(\boldsymbol{h}_1 + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_1, \ldots, \boldsymbol{h}_T + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_T \right)^\top \middle| \left\{ \boldsymbol{h}_k + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_k \right\}_{k=-\infty}^{0} \right) \\ &= \Sigma(\mathsf{SNR}) \\ &\preccurlyeq \mathrm{var}\left(\left(\boldsymbol{h}_1 + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_1, \ldots, \boldsymbol{h}_T + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_T \right)^\top \right) \\ &= R(0) + \frac{1}{\mathsf{SNR}} I_T\end{aligned}$$

Therefore, by Theorem 1

$$\begin{aligned}\lim_{\mathsf{SNR}\to\infty} \frac{C(\mathsf{SNR})}{\log \mathsf{SNR}} &= \lim_{\mathsf{SNR}\to\infty} \frac{-\log\det\Sigma(\mathsf{SNR})}{T\log\mathsf{SNR}} \\ &= \frac{T - \mathrm{rank}(R(0))}{T}.\end{aligned}$$

It is worth noting that in this case the pre-log term of the capacity can be achieved by a scheme simpler than the aforementioned interleaved decision-oriented training scheme. Suppose the rank of $R(0)$ is Q, so that $R(0)$ has $Q \times Q$ positive-definite principal submatrix. Without loss of generality, suppose this submatrix is the covariance matrix of $(\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_Q)^\top$. Then $\boldsymbol{h}_{kT+i}$ can be represented as a linear combination of $\boldsymbol{h}_{kT+1}, \boldsymbol{h}_{kT_2}, \ldots, \boldsymbol{h}_{kT+Q}$ for any $k \in \mathbb{Z}$ and $i \in \{Q+1, Q+2, \ldots, T\}$. The simpler scheme is described as follows.

The transmitter sends deterministic training symbols with maximum power at time instants $kT+1, kT+2, \ldots, kT+Q$, i.e.,

$$\boldsymbol{x}_{kT+1} = \boldsymbol{x}_{kT+2} = \cdots = \boldsymbol{x}_{kT+Q} = \sqrt{\mathsf{SNR}}$$

where $k = 0, 1, 2, \ldots$. The receiver can form the MMSE estimates

$$\mathbb{E}\left(\boldsymbol{h}_{kT+i} \middle| \left\{ \boldsymbol{h}_{kT+j} + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_{kT+j} \right\}_{j=1}^{Q} \right)$$

for $i = Q+1, Q+2, \ldots, T$ and $k = 0, 1, 2, \ldots$. Clearly, we have

$$\mathrm{var}\left(\boldsymbol{h}_{kT+i} \middle| \left\{ \boldsymbol{h}_{kT+j} + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_{kT+j} \right\}_{j=1}^{Q} \right) = O\left(\frac{1}{\mathsf{SNR}} \right).$$

With the side information

$$\left\{ \mathbb{E}\left(\boldsymbol{h}_{kT+i} \middle| \left\{ \boldsymbol{h}_{kT+j} + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_{kT+j} \right\}_{j=1}^{Q} \right) \right\}_{k=0}^{\infty}$$

at the receiver, we can communicate reliably at time instants $i, T+i, 2T+i, \ldots$ with rate at least

$$I\left(\boldsymbol{x}_i; \boldsymbol{y}_i \middle| \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{ \boldsymbol{h}_j + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_j \right\}_{j=1}^{Q} \right) \right).$$

Let $\boldsymbol{x}_i$ be uniformly distributed over the set

$$\left\{ z \in \mathbb{C} : \frac{\sqrt{\mathsf{SNR}}}{2} \le |z| \le \sqrt{\mathsf{SNR}} \right\}.$$

By Lemma 2

$$\begin{aligned}&I\left(\boldsymbol{x}_i; \boldsymbol{y}_i \middle| \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{ \boldsymbol{h}_j + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_j \right\}_{j=1}^{Q} \right) \right) \\ &\quad \ge -\log\left[\mathrm{var}\left(\boldsymbol{h}_i \middle| \left\{ \boldsymbol{h}_j + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_j \right\}_{j=1}^{Q} \right) + \frac{8}{5\mathsf{SNR}} \right] \\ &\quad\quad + \log\left(1 - \mathrm{var}\left(\boldsymbol{h}_i \middle| \left\{ \boldsymbol{h}_j + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_j \right\}_{j=1}^{Q} \right) \right) \\ &\quad\quad - \gamma - \log\frac{5e}{6} \\ &\quad = \log\mathsf{SNR} + o(\log\mathsf{SNR}).\end{aligned}$$

Therefore, the overall rate is lower-bounded by

$$\begin{aligned}&\frac{1}{T}\sum_{i=Q+1}^{T} I\left(\boldsymbol{x}_i; \boldsymbol{y}_i \middle| \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{ \boldsymbol{h}_j + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_j \right\}_{j=1}^{Q} \right) \right) \\ &\quad \ge \frac{T-Q}{T}\log\mathsf{SNR} + o(\log\mathsf{SNR}) \\ &\quad = \frac{T - \mathrm{rank}(R(0))}{T}\log\mathsf{SNR} + o(\log\mathsf{SNR})\end{aligned}$$

and the pre-log term is achieved. This scheme has the following obvious advantages over the interleaved decision-oriented training scheme: i) channel estimation and communication are completely decoupled; and ii) channel estimation is done locally since the estimate

$$\mathbb{E}\left(\boldsymbol{h}_{kT+i} \middle| \left\{ \boldsymbol{h}_{kT+j} + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_{kT+j} \right\}_{j=1}^{Q} \right), \quad i = Q+1, Q+2, \ldots, T$$

only depends $\boldsymbol{h}_{kT+1}, \boldsymbol{h}_{kT+2}, \ldots, \boldsymbol{h}_{kT+Q}$.

D. Regular Block-Stationary Process

The following theorem generalizes [3, Corollary 4.42] for regular Gaussian fading processes to the block-stationary case.

Theorem 2: If $\det(\Sigma(\infty)) > 0$, then

$$\lim_{\mathsf{SNR}\to\infty} [C(\mathsf{SNR}) - \log\log \mathsf{SNR}] = -1 - \gamma - \frac{1}{T}\log\det(\Sigma(\infty)) = -1 - \gamma - \frac{1}{2\pi T}\int_{-\pi}^{\pi} \log\det\left[S(e^{j\omega})\right] \mathrm{d}\omega. \quad (13)$$

Remark 4: The second equality in (13) follows from (3).

Proof: See Appendix IV. □

Example 3: Gauss–Markov Process: Suppose $\{\boldsymbol{h}_t\}_{t=-\infty}^{\infty}$ is a Gauss–Markov process with $E(\boldsymbol{h}_{t+1}\boldsymbol{h}_t^*) = \rho_1$ if $(t \bmod T) = 0$, and $= \rho_2$ otherwise. Here ρ_1, ρ_2 are complex numbers with $\max(|\rho_1|, |\rho_2|) < 1$. In this case, we have

$$\det(\Sigma(\infty)) = \left(1 - |\rho_1|^2\right)\left(1 - |\rho_2|^2\right)^{T-1}.$$

Therefore, by Theorem 2

$$\lim_{\mathsf{SNR}\to\infty} [C(\mathsf{SNR}) - \log\log \mathsf{SNR}] = -1 - \gamma - \log\left(1 - |\rho_2|^2\right) - \frac{\log(1 - |\rho_1|^2) - \log\left(1 - |\rho_2|^2\right)}{T}.$$

If $|\rho_1| < 1$ and $|\rho_2| = 1$, then it follows from (12) that

$$\lim_{\mathsf{SNR}\to\infty} \frac{C(\mathsf{SNR})}{\log \mathsf{SNR}} = \frac{T-1}{T}.$$

Therefore, a small perturbation of a parameter in the channel model can dramatically affect the asymptotic capacity. However, this phenomenon needs to be interpreted with great caution. It should be noted that the sensitivity of the capacity asymptotics to the channel modeling does not imply the sensitivity of the channel capacity at a fixed SNR level. Intuitively, if the parameter perturbation is small enough, its effect on the channel capacity at a fixed SNR level should be negligible. In the above example, if $|\rho_2|$ varies from 1 to $1 - \epsilon$ (where ϵ is a small positive number), it is natural to expect that the channel capacity only changes slightly over a wide range of SNR, and that a significant difference only appears at sufficiently high SNR.

IV. Symbol-By-Symbol Stationary Fading Model

For simplicity, we assume in this section that the fading process is symbol-by-symbol stationary, i.e., $T = 1$. In this case, Theorem 1 is specialized to (11).

A. Best and Worst Case Spectral Densities

We can see that two fading processes with spectral density functions $s_1(e^{j\omega})$ and $s_2(e^{j\omega})$ can induce the same pre-log term in the high-SNR regime as long as $\mu(s_1(e^{j\omega}) = 0) = \mu(s_2(e^{j\omega}) = 0)$. But in the nonasymptotic regime, the capacities of these two channels may behave very differently. So, for a fixed $\mu(s(e^{j\omega}) = 0)$, it is natural to ask the question: which spectral density function $s(e^{j\omega})$ gives the largest (or smallest) channel capacity at a given SNR? This question is difficult to answer since we do not have a closed-form expression for noncoherent channel capacity. We therefore turn to the lower bound (5) to formulate a closely related problem.

When $T = 1$, the lower bound (5) can be reduced to

$$C(\mathsf{SNR}) \geq I\left(\boldsymbol{x}_1; \boldsymbol{h}_1\boldsymbol{x}_1 + \boldsymbol{z}_1 \,\middle|\, \mathbb{E}\left(\boldsymbol{h}_1 \,\middle|\, \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_1\right\}_{k=-\infty}^{0}\right)\right). \quad (14)$$

We can see that the lower bound (14) depends on $s(e^{j\omega})$ only through

$$\mathbb{E}\left(\boldsymbol{h}_1 \,\middle|\, \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_1\right\}_{k=-\infty}^{0}\right).$$

Furthermore, if we fix the input distribution $p_{\boldsymbol{x}_1}$, then it follows from Proposition 2 in Appendix I that

$$\operatorname{var}\left(\boldsymbol{h}_1 \,\middle|\, \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_1\right\}_{k=-\infty}^{0}\right)\Bigg|_{s_1(e^{j\omega})} \leq \operatorname{var}\left(\boldsymbol{h}_1 \,\middle|\, \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_1\right\}_{k=-\infty}^{0}\right)\Bigg|_{s_2(e^{j\omega})}$$

implies the equation at the bottom of the page . We can therefore ask which $s(e^{j\omega})$ yields the largest (or smallest) value for

$$\operatorname{var}\left(\boldsymbol{h}_1 \,\middle|\, \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_1\right\}_{k=-\infty}^{0}\right)\Bigg|_{s(e^{j\omega})}.$$

More precisely, since

$$\operatorname{var}\left(\boldsymbol{h}_1 \,\middle|\, \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_1\right\}_{k=-\infty}^{0}\right)\Bigg|_{s(e^{j\omega})} = \operatorname{var}\left(\boldsymbol{h}_1 + \frac{1}{x_{\min}}\boldsymbol{z}_1 \,\middle|\, \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_1\right\}_{k=-\infty}^{0}\right)\Bigg|_{s(e^{j\omega})} - \frac{1}{x_{\min}^2} = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left[s(e^{j\omega}) + \frac{1}{x_{\min}^2}\right]\mathrm{d}\omega\right\} - \frac{1}{x_{\min}^2}$$

$$I\left(\boldsymbol{x}_1; \boldsymbol{h}_1\boldsymbol{x}_1 + \boldsymbol{z}_1 \,\middle|\, \mathbb{E}\left(\boldsymbol{h}_1 \,\middle|\, \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_1\right\}_{k=-\infty}^{0}\right)\right)\Bigg|_{s_1(e^{j\omega})} \geq I\left(\boldsymbol{x}_1; \boldsymbol{h}_1\boldsymbol{x}_1 + \boldsymbol{z}_1 \,\middle|\, \mathbb{E}\left(\boldsymbol{h}_1 \,\middle|\, \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_1\right\}_{k=-\infty}^{0}\right)\right)\Bigg|_{s_2(e^{j\omega})}.$$

we can formulate the problem in the following form:

$$\arg\max_{s(e^{j\omega})}(\text{or } \arg\min_{s(e^{j\omega})})\int_{-\pi}^{\pi}\log\left[s(e^{j\omega})+\frac{1}{x_{\min}^2}\right]\mathrm{d}\omega \quad (15)$$

subject to

$$s(e^{j\omega})\geq 0,\quad \frac{1}{2\pi}\int_{-\pi}^{\pi}s(e^{j\omega})\mathrm{d}\omega=1,\quad \mu(s(e^{j\omega})=0)=\alpha$$

where $\alpha\in[0,2\pi)$. Due to the strict concavity of $\log(\cdot)$, it is easy to show that the maximizers of the optimization problem (15) are the set of spectral density functions $s(e^{j\omega})$ with the property

$$\mu\left(s(e^{j\omega})=\frac{2\pi}{2\pi-\alpha}\right)=2\pi-\alpha,\quad \mu(s(e^{j\omega})=0)=\alpha.$$

This solution has the following interpretation. Without constraints on the spectral density function, the worst fading process is the i.i.d. Gaussian process whose spectral density function is flat. With the constraint $\mu(s(e^{j\omega})=0)=\alpha$, the spectral density function $s(e^{j\omega})$ cannot be completely flat, but the worst fading process should have a spectral density function that is as flat as possible, i.e., the correlation in the time domain should be the weakest possible. Note that the solution does not depend on $x_{\min}$. We can use this fact to derive a universal lower bound on $C(\mathsf{SNR})$ for the class of spectral density functions $\{s(e^{j\omega}) : \mu(s(e^{j\omega})=0)=\alpha\}$, which has further implications for the high-SNR asymptotic behavior of $C(\mathsf{SNR})$. Let $s_{\max}(e^{j\omega})$ be a maximizer of (15). We have

$$\begin{aligned}&\mathrm{var}\left(\boldsymbol{h}_1\left|\left\{\boldsymbol{h}_k+\frac{1}{x_{\min}}\boldsymbol{z}_1\right\}_{k=-\infty}^{0}\right.\right)\Bigg|_{s_{\max}(e^{j\omega})}\\&=\left(\frac{2\pi}{2\pi-\alpha}+\frac{1}{x_{\min}^2}\right)^{\frac{2\pi-\alpha}{2\pi}}\left(\frac{1}{x_{\min}^2}\right)^{\frac{\alpha}{2\pi}}-\frac{1}{x_{\min}^2}\\&\triangleq\phi(\alpha,x_{\min}).\end{aligned}$$

For any spectral density function $s(e^{j\omega})$ with $\mu(s(e^{j\omega})=0)=\alpha$, we get (16) at the bottom of the page ,where $\boldsymbol{x}$ is distributed over the set $\{z\in\mathbb{C} : x_{\min}\leq|z|\leq\sqrt{\mathsf{SNR}}\}$, $\widehat{\boldsymbol{h}}\sim\mathcal{CN}(0,1-\phi(\alpha,x_{\min}))$, $\widetilde{\boldsymbol{h}}\sim\mathcal{CN}(0,\phi(\alpha,x_{\min}))$, $\boldsymbol{z}\sim\mathcal{CN}(0,1)$, and $\boldsymbol{x},\widehat{\boldsymbol{h}},\widetilde{\boldsymbol{h}},\boldsymbol{z}$ are all independent. We can further optimize over $p(\boldsymbol{x})$ to tighten the lower bound (16).

Minimizers of (15) do not exist. Consider the following set spectral density functions $\{s_\theta(e^{j\omega})\}_\theta$ given by

$$s_\theta(e^{j\omega})=\begin{cases}0, & |\omega|\leq\frac{\alpha}{2}\\ \frac{1}{\theta}, & |\omega|\in(\frac{\alpha}{2},\pi-\frac{1}{2\theta}]\\ \frac{2\pi\theta^2-2\pi\theta+\alpha\theta+1}{\theta}, & |\omega|\in(\pi-\frac{1}{2\theta},\pi]\end{cases}$$

where $\theta\geq\theta_0$ with

$$\theta_0=\begin{cases}\frac{1}{2\pi-\alpha}, & (2\pi-\alpha)^2<8\pi\\ \frac{2\pi-\alpha+\sqrt{(2\pi-\alpha)^2-8\pi}}{4\pi}, & (2\pi-\alpha)^2\geq 8\pi.\end{cases}$$

We can compute

$$\begin{aligned}&\lim_{\theta\to\infty}\int_{-\pi}^{\pi}\log\left[s_\theta(e^{j\omega})+\frac{1}{x_{\min}^2}\right]\mathrm{d}\omega\\&=\lim_{\theta\to\infty}\Bigg[-2\alpha\log x_{\min}+\left(2\pi-\alpha-\frac{1}{\theta}\right)\log\left(\frac{1}{\theta}+\frac{1}{x_{\min}^2}\right)\\&\qquad+\frac{1}{\theta}\log\left(\frac{2\pi\theta^2-2\pi\theta+\alpha\theta+1}{\theta}+x_{\min}^2\right)\Bigg]\\&=-4\pi\log x_{\min}.\end{aligned}$$

Note that

$$\int_{-\pi}^{\pi}\log\left[s(e^{j\omega})+\frac{1}{x_{\min}^2}\right]\mathrm{d}\omega>-4\pi\log x_{\min}.$$

Therefore, as θ goes to infinity, $\int_{-\pi}^{\pi}\log\left[s(e^{j\omega})+\frac{1}{x_{\min}^2}\right]\mathrm{d}\omega$ approaches the lower bound that is not attainable by any spectral density function. Intuitively, the fading process associated with $s_\theta(e^{j\omega})$ becomes more and more deterministic as θ gets larger, and it can be verified that

$$\lim_{\theta\to\infty}\mathrm{var}\left(\boldsymbol{h}_1\left|\left\{\boldsymbol{h}_k+\frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{0}\right.\right)\Bigg|_{s_\theta(e^{j\omega})}=0.$$

This result has interesting implications for the channel capacity.

Proposition 1: For any $r>0$,

$$\lim_{\mathsf{SNR}\to\infty,\theta=\mathsf{SNR}^r}\frac{C(\mathsf{SNR})|_{s_\theta(e^{j\omega})}}{\log\mathsf{SNR}}=\frac{\alpha+\min(r,1)(2\pi-\alpha)}{2\pi}.$$

Proof: See Appendix V. □

Remark 5: Although by Theorem 1, for any fixed θ, the ratio between $C(\mathsf{SNR})|_{s_\theta(e^{j\omega})}$ and $\log\mathsf{SNR}$ converges to $\frac{\alpha}{2\pi}$, Proposition 1 says that the convergence is not uniform with respect to θ. This is intuitively clear because when θ is large, we have $s_\theta(e^{j\omega})\approx 0$ for $|\omega|\in[0,\pi-\frac{1}{2\theta}]$. Therefore, it can be expected that for a large range of SNR, the channel capacity $C(\mathsf{SNR})|_{s_\theta(e^{j\omega})}$ behaves like $\left(1-\frac{1}{2\pi\theta}\right)\log\mathsf{SNR}$, which could be significantly larger than $\frac{\alpha}{2\pi}\log\mathsf{SNR}$. For the extreme case where $\alpha=0$, by Theorem 2 (also see [3, Corollary 4.42]), for any fixed θ, the capacity $C(\mathsf{SNR})|_{s(e^{j\omega})}$ grows like $\log\log\mathsf{SNR}$ at high SNR. But Proposition 1 implies that even in this extreme case, the capacity $C(\mathsf{SNR})|_{s_\theta(e^{j\omega})}$ for some θ can grow linearly with $\log\mathsf{SNR}$ for a large range of SNR. This is consistent with the result in [13] where it was shown that for the Gauss–Markov

$$\begin{aligned}C(\mathsf{SNR})|_{s(e^{j\omega})}&\geq I\left(\boldsymbol{x}_1;\boldsymbol{h}_1\boldsymbol{x}_1+\boldsymbol{z}_1\left|\mathbb{E}\left(\boldsymbol{h}_1\left|\left\{\boldsymbol{h}_k+\frac{1}{x_{\min}}\boldsymbol{z}_1\right\}_{k=-\infty}^{0}\right.\right)\right.\right)\Bigg|_{s_{\max}(e^{j\omega})}\\&=I\left(\boldsymbol{x};\left(\widehat{\boldsymbol{h}}+\widetilde{\boldsymbol{h}}\right)\boldsymbol{x}+\boldsymbol{z}\middle|\widehat{\boldsymbol{h}}\right)\end{aligned} \quad (16)$$

process with $E(\boldsymbol{h}_{i+1}\boldsymbol{h}_i^*) = \rho$, the capacity $C(\mathsf{SNR})$ grows like $\log \mathsf{SNR}$ for a wide range of SNR levels if $|\rho|$ is close to 1. An intuitive explanation for this behavior is that if $|\rho|$ is close to 1, the spectrum

$$s(e^{j\omega}) = \frac{1-|\rho|^2}{1-2\mathrm{Re}(\rho e^{-j\omega})+|\rho|^2}$$

is approximately zero for all values of ω except those around zero, and we can expect from (11) that $C(\mathsf{SNR})$ should grow like $\log \mathsf{SNR}$ for a wide range of SNR. But it should be noted that as opposed to a Gauss–Markov process, a general Gaussian stationary process cannot be characterized by a single parameter, and the behavior of $C(\mathsf{SNR})$ can be more complicated as shown in the following example.

Example 4: Consider the spectral density function

$$s(e^{j\omega}) = \begin{cases} \epsilon_1, & |\omega| \le \pi\alpha_1 \\ \epsilon_2, & |\omega| \in (\pi\alpha_1, \pi\alpha_2] \\ \frac{1-\alpha_1\epsilon_1-(\alpha_2-\alpha_1)\epsilon_2}{1-\alpha_2}, & |\omega| \in (\pi\alpha_2, \pi] \end{cases}$$

where $\frac{1}{3} \le \alpha_1 < \alpha_2 \le \frac{2}{3}$, $1 \ll \log\frac{1}{\epsilon_2} \ll \log\frac{1}{\epsilon_1}$, and $\epsilon_1 > 0$ (Note: For two positive numbers a and b, $a \ll b$ means $\frac{b}{a}$ is much greater than 1). We show in Appendix VI that $\frac{C(\mathsf{SNR})}{\log \mathsf{SNR}}$ is approximately equal to α_2 for $1 \ll \log \mathsf{SNR} \le \frac{1}{\epsilon_2}$, and gradually decreases to α_1 as SNR approaches $\frac{1}{\epsilon_1}$. By Theorem 2 (also see [3, Corollary 4.42]), $C(\mathsf{SNR})$ eventually grows like

$$\log\log \mathsf{SNR} - 1 - \gamma - \alpha_1 \log \epsilon_1 - (\alpha_2-\alpha_1) \times \log \epsilon_2 - (1-\alpha_2)\log\left[\frac{1-\alpha_1\epsilon_1-(\alpha_2-\alpha_1)\epsilon_2}{1-\alpha_2}\right] + o(1)$$

and consequently $\frac{C(\mathsf{SNR})}{\log \mathsf{SNR}}$ converges to 0.

This example shows that $\frac{C(\mathsf{SNR})}{\log \mathsf{SNR}}$ can be highly SNR dependent. For a regular Gaussian fading process, the noncoherent capacity can be approximated by $\log\log \mathsf{SNR} + c$ in the high-SNR limit, where the constant c is termed the *fading number* [3]. However, in order for this approximation to be accurate, one needs $\log\log \mathsf{SNR}$ to be at least comparable with the fading number, which may require large values of SNR if the fading number is large. The behavior of $C(\mathsf{SNR})$ at moderate SNR levels (more precisely, in the regime where $\log\log \mathsf{SNR}$ is not significantly larger than the fading number) may be highly dependent on the spectral density function. A related and detailed discussion can also be found in [4, Sec. III-B].

Overall, these examples suggest that great caution should be exercised when using the asymptotic results in Theorems 1 and 2 to approximate the channel capacity $C(\mathsf{SNR})$ at a finite SNR level. The analysis of these examples also indicates that more information about the channel capacity is contained in the noisy prediction error $\Sigma(\mathsf{SNR})$ than in the capacity asymptotics. Therefore, as also suggested in [14], it is important to study the noisy prediction error to better understand the moderate SNR regime.

B. Finite Codeword-Length Behavior

In the capacity analysis, it is assumed that the codeword is of infinite length. But when the length of codewords is finite, the behavior of the communication rate as a function of SNR can be quite different. By Fano's inequality, the communication rate R is upper-bounded by

$$R \le \frac{I(\boldsymbol{x}^n;\boldsymbol{y}^n)+1}{n(1-P_e)}$$

where n is the codeword length and P_e is the decoding error probability. Suppose we fix n and P_e, and let SNR go to infinity. For a symbol-by-symbol stationary Gaussian fading process, even if $\mu(s(e^{j\omega})=0) > 0$, the correlation matrix of the fading process over any finite block length can still be full-rank. Note that $\frac{1}{n}I(\boldsymbol{x}^n;\boldsymbol{y}^n)$ is upper-bounded by the capacity of a block-independent Gaussian fading channel with the correlation matrix of each block given by $\mathbb{E}[\boldsymbol{h}^n(\boldsymbol{h}^n)^\dagger]$. Since $\mathbb{E}[\boldsymbol{h}^n(\boldsymbol{h}^n)^\dagger]$ is full-rank, it follows from Theorem 2 that $\frac{1}{n}I(\boldsymbol{x}^n;\boldsymbol{y}^n)$, and hence, R, grows at most like $\log\log \mathsf{SNR}$ as SNR goes to infinity. Therefore, there is no nontrivial tradeoff between diversity and multiplexing in the sense of [15]. If we want R to grow linearly with $\log \mathsf{SNR}$ while having the decoding error probability P_e bounded away from 1, the codeword length n must scale with SNR. It is of interest to determine how fast the codeword n should scale with SNR in order to guarantee that the rate R can grow as $\log \mathsf{SNR}$ with the decoding error probability not approaching 1. More precisely, letting the rate $R(\mathsf{SNR})$, codeword length $n(\mathsf{SNR})$, and decoding error probability $P_e(\mathsf{SNR})$ all depend on SNR, we wish to determine conditions on $n(\mathsf{SNR})$ that guarantee the existence of a sequence of codebooks (indexed by SNR) with rate $R(\mathsf{SNR})$ and codeword length $n(\mathsf{SNR})$ such that

$$\liminf_{\mathsf{SNR}\to\infty} \frac{R(\mathsf{SNR})}{\log \mathsf{SNR}} \ge r$$

and

$$\limsup_{\mathsf{SNR}\to\infty} P_e(\mathsf{SNR}) \le P_e$$

where $r > 0$ and $P_e \in (0,1)$.

Now we proceed to derive a necessary condition on the growth rate of $n(\mathsf{SNR})$. It follows from the chain rule that

$$I(\boldsymbol{x}^{n(\mathsf{SNR})};\boldsymbol{y}^{n(\mathsf{SNR})}) = \sum_{k=1}^{n(\mathsf{SNR})} I\left(\boldsymbol{x}^{n(\mathsf{SNR})};\boldsymbol{y}_k \,\middle|\, \boldsymbol{y}^{k-1}\right).$$

By (6), we can upper-bound $I\left(\boldsymbol{x}^{n(\mathsf{SNR})};\boldsymbol{y}_k \,\middle|\, \boldsymbol{y}^{k-1}\right)$ as shown in the equation at the bottom of the page. Since $\sup_{p_{\boldsymbol{x}_k}\in\mathcal{P}_1(\mathsf{SNR})} I(\boldsymbol{x}_k;\boldsymbol{y}_k) = o(\log \mathsf{SNR})$, and the equation

$$I\left(\boldsymbol{x}^{n(\mathsf{SNR})};\boldsymbol{y}_k \,\middle|\, \boldsymbol{y}^{k-1}\right) \le \sup_{p_{\boldsymbol{x}_k}\in\mathcal{P}_1(\mathsf{SNR})} I\left(\boldsymbol{x}_k, \boldsymbol{h}_{k-1}+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_{k-1},\ldots,\boldsymbol{h}_1+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_1;\boldsymbol{y}_k\right)$$
$$\le \sup_{p_{\boldsymbol{x}_k}\in\mathcal{P}_1(\mathsf{SNR})} I(\boldsymbol{x}_k;\boldsymbol{y}_k) + \sup_{p_{\boldsymbol{x}_k}\in\mathcal{P}_1(\mathsf{SNR})} I\left(\boldsymbol{h}_{k-1}+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_{k-1},\ldots,\boldsymbol{h}_1+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_1;\boldsymbol{y}_k \,\middle|\, \boldsymbol{x}_k\right).$$

$$\begin{aligned} I\left(\boldsymbol{h}_{k-1}+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_{k-1},\ldots,\boldsymbol{h}_1+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_1;\boldsymbol{y}_k\middle|\boldsymbol{x}_k\right) &= \mathbb{E}\left\{\log\left[\frac{1+|\boldsymbol{x}_k|^2}{1+|\boldsymbol{x}_k|^2\cdot\operatorname{var}\left(\boldsymbol{h}_k\middle|\left\{\boldsymbol{h}_v+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_v\right\}_{v=1}^{k-1}\right)}\right]\right\} \\ &\le \log\left[\frac{1+\mathsf{SNR}}{1+\mathsf{SNR}\cdot\operatorname{var}\left(\boldsymbol{h}_k\middle|\left\{\boldsymbol{h}_v+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_v\right\}_{v=1}^{k-1}\right)}\right] \\ &= \log\left(\frac{1+\mathsf{SNR}}{\mathsf{SNR}}\right)-\log\operatorname{var}\left(\boldsymbol{h}_k+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_k\middle|\left\{\boldsymbol{h}_v+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_v\right\}_{v=1}^{k-1}\right) \\ &\le \log\left(\frac{1+\mathsf{SNR}}{\mathsf{SNR}}\right)-\log\operatorname{var}\left(\boldsymbol{h}_0+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_0\middle|\left\{\boldsymbol{h}_v+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_v\right\}_{v=-n(\mathsf{SNR})}^{-1}\right) \end{aligned}$$

at the top of the page, we get, by Fano's inequality and the condition $\limsup_{\mathsf{SNR}\to\infty} P_e(\mathsf{SNR}) \le P_e$ the first equation at the bottom of the page. Therefore, in order for

$$\liminf_{\mathsf{SNR}\to\infty}\frac{R(\mathsf{SNR})}{\log\mathsf{SNR}} \ge r,$$

we must have have (17), also at the bottom of the page. Since

$$-\log\operatorname{var}\left(\boldsymbol{h}_0+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_0\middle|\left\{\boldsymbol{h}_v+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_v\right\}_{v=-n}^{-1}\right)$$

is a monotone increasing function of n, it is easy to see that (17) implicitly provides us with a lower bound on the scaling rate of $n(\mathsf{SNR})$.

In order to derive an explicit lower bound on the scaling rate of $n(\mathsf{SNR})$, we need to introduce a concept called *transfinite diameter* [16].

Definition 1: Let $\mathcal{S}$ be a compact set in the plane. Set

$$\begin{aligned} V(z_1,\ldots,z_n) &= \prod_{j>k}(z_j-z_k), \qquad n\ge 2,\ z_i\in\mathcal{S} \\ V_n(\mathcal{S}) &= \max_{z_1,\ldots,z_n\in\mathcal{S}}|V(z_1,\ldots,z_n)| \end{aligned}$$

and

$$\tau_n(\mathcal{S}) = [V_n(\mathcal{S})]^{\frac{2}{n(n-1)}}.$$

The transfinite diameter of $\mathcal{S}$ is defined by

$$\tau(\mathcal{S}) = \lim_{n\to\infty}\tau_n(\mathcal{S}).$$

We need the following facts regarding the transfinite diameter.

i) For two compact sets $\mathcal{S}_1$ and $\mathcal{S}_2$ with $\mathcal{S}_1 \subseteq \mathcal{S}_2$, we have $\tau(\mathcal{S}_1) \le \tau(\mathcal{S}_2)$.
ii) The diameter of the unit circle is 1. More generally, the diameter of an arc of central angle θ on the unit circle is $\sin\left(\frac{\theta}{4}\right)$.
iii) The transfinite diameter of any closed proper subset of the unit circle is less than 1.

A full discussion of the transfinite diameter can be found in [16].

Now return to the original problem. Since

$$\operatorname{var}\left(\boldsymbol{h}_0+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_0\middle|\left\{\boldsymbol{h}_v+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_v\right\}_{v=-n(\mathsf{SNR})}^{-1}\right) \ge \operatorname{var}\left(\boldsymbol{h}_0\middle|\{\boldsymbol{h}_v\}_{v=-n(\mathsf{SNR})}^{-1}\right)$$

$$\begin{aligned} \liminf_{\mathsf{SNR}\to\infty}\frac{R(\mathsf{SNR})}{\log\mathsf{SNR}} &\le \liminf_{\mathsf{SNR}\to\infty}\frac{I(\boldsymbol{x}^{n(\mathsf{SNR})};\boldsymbol{y}^{n(\mathsf{SNR})})+1}{n(\mathsf{SNR})(1-P_e(\mathsf{SNR}))\log\mathsf{SNR}} \\ &\le \liminf_{\mathsf{SNR}\to\infty}\frac{-\log\operatorname{var}\left(\boldsymbol{h}_0+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_0\middle|\left\{\boldsymbol{h}_v+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_v\right\}_{v=-n(\mathsf{SNR})}^{-1}\right)}{(1-P_e(\mathsf{SNR}))\log\mathsf{SNR}} \\ &\le \liminf_{\mathsf{SNR}\to\infty}\frac{-\log\operatorname{var}\left(\boldsymbol{h}_0+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_0\middle|\left\{\boldsymbol{h}_v+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_v\right\}_{v=-n(\mathsf{SNR})}^{-1}\right)}{(1-P_e)\log\mathsf{SNR}}. \end{aligned}$$

$$\liminf_{\mathsf{SNR}\to\infty}\frac{-\log\operatorname{var}\left(\boldsymbol{h}_0+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_0\middle|\left\{\boldsymbol{h}_v+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_v\right\}_{v=-n(\mathsf{SNR})}^{-1}\right)}{\log\mathsf{SNR}} \ge r(1-P_e). \tag{17}$$

we have the equation at the bottom of the page. Let $\mathcal{S} = \{e^{j\omega} : s(e^{j\omega}) > 0\}$. It was shown in [17] that if the set $\mathcal{S}$ consists of a finite number of arcs of the unit circle, then

$$\lim_{n\to\infty}\left(\mathrm{var}\left(\boldsymbol{h}_0 \,\Big|\, \{\boldsymbol{h}_v\}_{v=-n}^{-1}\right)\right)^{\frac{1}{n}} = \tau(\mathcal{S}).$$

Under the conditions

a) the set $\mathcal{S}$ consists of a finite number of arcs of the unit circle;

b) the set $\mathcal{S}$ is a closed proper subset of the unit circle;

it can be shown by using Facts i)–iii) that

$$0 < \tau(\mathcal{S}) < 1.$$

Therefore, under Conditions a) and b), we have

$$\begin{aligned}
&\liminf_{\mathsf{SNR}\to\infty} \frac{-\log \mathrm{var}\left(\boldsymbol{h}_0 \,\Big|\, \{\boldsymbol{h}_v\}_{v=-n(\mathsf{SNR})}^{-1}\right)}{\log \mathsf{SNR}}\\
&= \liminf_{\mathsf{SNR}\to\infty} \frac{-\frac{1}{n(\mathsf{SNR})}\log \mathrm{var}\left(\boldsymbol{h}_0 \,\Big|\, \{\boldsymbol{h}_v\}_{v=-n(\mathsf{SNR})}^{-1}\right)}{\frac{1}{n(\mathsf{SNR})}\log \mathsf{SNR}}\\
&= \liminf_{\mathsf{SNR}\to\infty} \frac{-n(\mathsf{SNR})\log\tau(\mathcal{S})}{\log \mathsf{SNR}}. \qquad (18)
\end{aligned}$$

It is clear that in order to guarantee that (18) is greater than or equal to $r(1-P_e)$, we must have

$$\liminf_{\mathsf{SNR}\to\infty} \frac{n(\mathsf{SNR})}{\log \mathsf{SNR}} \geq -\frac{r(1-P_e)}{\log\tau(\mathcal{S})}$$

which is a necessary condition on the scaling rate of $n(\mathsf{SNR})$.

In contrast to this result, we show in Appendices VII and VIII that for the additive white Gaussian noise (AWGN) channel and memoryless coherent Rayleigh-fading channel, it is possible to have the rate $R(\mathsf{SNR})$ grow linearly with $\log\mathsf{SNR}$ with fixed codeword length n and bounded decoding error probability at high SNR. For these two cases, to facilitate the calculation, we adopt the average power constraint. But our main conclusion holds also under the peak power constraint.

It can be seen from (2) and (3) that $\det[\Sigma(\mathsf{SNR})]$ and $\det[\Sigma(\infty)]$ are invariant under a reordering of the set of harmonics $[-\pi,\pi]$. In view of Theorems 1 and 2, this invariance property is inherited by the capacity asymptotics. In fact, many existing nonasymptotic capacity bounds depend on the spectral density function only through $\det[\Sigma(\mathsf{SNR})]$ or $\det[\Sigma(\infty)]$, and consequently possess the same invariance property. In contrast, the transfinite diameter $\tau(\mathcal{S})$ can be affected by a reordering of the set of harmonics. Therefore, more information about the spectral density function is reflected in the finite codeword-length setting than in the capacity results.

V. Capacity Per Unit Energy

In the preceding sections, we focused on the channel capacity in the high-SNR regime. Now we proceed to characterize the behavior of channel capacity in the low average power regime for the block-stationary Gaussian fading channel model. To this end, we shall study the capacity per unit energy (which is denoted by $C_p(\mathsf{SNR})$) due to its intrinsic connection with the channel capacity in this regime. The following theorem provides a general expression for the capacity per unit energy.

Theorem 3 ([9], [18]):

$$C_p(\mathsf{SNR}) = \lim_{n\to\infty} \sup_{\boldsymbol{x}^n\in\mathbb{D}_n(\mathsf{SNR})} \frac{D(p_{\boldsymbol{y}^n|\boldsymbol{x}^n}|D(p_{\boldsymbol{y}^n|0^n})}{|\boldsymbol{x}^n|_2^2}.$$

Furthermore, the capacity per unit energy is related to the capacity by

$$C_p(\mathsf{SNR}) = \sup_{\mathsf{P}>0}\frac{C(\mathsf{P},\mathsf{SNR})}{\mathsf{P}} = \lim_{\mathsf{P}\to 0}\frac{C(\mathsf{P},\mathsf{SNR})}{\mathsf{P}}$$

where $C(\mathsf{P},\mathsf{SNR})$ is the channel capacity with average power constraint P and peak power constraint SNR.

The following theorem is an extension of [18, Proposition 3.1] for the symbol–symbol stationary channel model to the block-stationary model.

Theorem 4: For the block-stationary Gaussian fading channel model given in (1)

$$C_p(\mathsf{SNR}) = 1 - \frac{1}{2\pi\mathsf{SNR}} \min_{\mathcal{M}\subseteq\{1,\ldots,T\}} \Psi(\mathcal{M},\mathsf{SNR})$$

where

$$\Psi(\mathcal{M},\mathsf{SNR}) = \frac{1}{|\mathcal{M}|}\int_{-\pi}^{\pi}\log\det\left[I_{|\mathcal{M}|} + \mathsf{SNR}S_{\mathcal{M}}(e^{j\omega})\right]\mathrm{d}\omega$$

and $S_{\mathcal{M}}(e^{j\omega})$ is an $|\mathcal{M}|\times|\mathcal{M}|$ principal minor of $S(e^{j\omega})$ with the indices of columns and rows specified by $\mathcal{M}$.

Proof: The proof is omitted since it is almost identical to that for the symbol-by-symbol stationary fading channel [18]. The only difference is that although the capacity per unit energy of the block-stationary fading channel can be asymptotically achieved by temporal ON–OFF signaling, we have to determine how to allocate the ON symbols in a block. It can be shown that the optimal allocation scheme is given by $\mathcal{M}^*$, which is the minimizer of $\min_{\mathcal{M}\in\{1,\ldots,T\}}\Psi(\mathcal{M})$. Here $\mathcal{M}^*$ might not be unique. □

We note that $C_p(\mathsf{SNR})$ is a monotonically increasing function of SNR. It is easy to see that $C_p(\mathsf{SNR})$ goes to 1 as $\mathsf{SNR}\to\infty$, and goes to 0 as $\mathsf{SNR}\to 0$. The following result provides a more precise characterization of the convergence behavior.

$$\liminf_{\mathsf{SNR}\to\infty}\frac{-\log\mathrm{var}\left(\boldsymbol{h}_0+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_0\,\Big|\,\left\{\boldsymbol{h}_v+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_v\right\}_{v=-n(\mathsf{SNR})}^{-1}\right)}{\log\mathsf{SNR}} \leq \liminf_{\mathsf{SNR}\to\infty}\frac{-\log\mathrm{var}\left(\boldsymbol{h}_0\,\Big|\,\{\boldsymbol{h}_v\}_{v=-n(\mathsf{SNR})}^{-1}\right)}{\log\mathsf{SNR}}.$$

Corollary 1: At high SNR, we get (19) at the bottom of the page. At low SNR, if

$$\int_{-\pi}^{\pi} \mathrm{tr}\left[S^2(e^{j\omega})\right] \mathrm{d}\omega < \infty$$

then

$$C_p(\mathsf{SNR}) = \frac{\mathsf{SNR}}{4\pi} \max_{\mathcal{M}\subseteq\{1,\dots,T\}} \frac{1}{|\mathcal{M}|} \times \int_{-\pi}^{\pi} \mathrm{tr}\left[S_{\mathcal{M}}^2(e^{j\omega})\right] \mathrm{d}\omega + o(\mathsf{SNR}). \quad (20)$$

Proof: By Lemma 1, at high SNR

$$\int_{-\pi}^{\pi} \log\det\left[\frac{1}{\mathsf{SNR}} I_{|\mathcal{M}|} + S_{\mathcal{M}}(e^{j\omega})\right] \mathrm{d}\omega = -\sum_{i=0}^{|\mathcal{M}|} (|\mathcal{M}| - i)\mu(\mathrm{rank}(S_{\mathcal{M}}(e^{j\omega})) = i) \times \log \mathsf{SNR} + o(\log \mathsf{SNR}).$$

Therefore, we get the second equation at the bottom of the page. At low SNR, using the second-order approximation [19], we obtain

$$\log\det\left[I_{|\mathcal{M}|} + \mathsf{SNR} S_{\mathcal{M}}(e^{j\omega})\right] = \mathrm{tr}\left[S_{\mathcal{M}}(e^{j\omega})\right]\mathsf{SNR} - \frac{1}{2}\mathrm{tr}\left[S_{\mathcal{M}}^2(e^{j\omega})\right]\mathsf{SNR}^2 + o(\mathsf{SNR}^2).$$

Therefore, we have the third equation at the bottom of the page, where the last equality follows from the fact that

$$\frac{1}{2\pi|\mathcal{M}|}\int_{-\pi}^{\pi} \mathrm{tr}\left[S_{\mathcal{M}}(e^{j\omega})\right] \mathrm{d}\omega = 1. \qquad \square$$

Remark 6: Using the inequality[4]

$$\log\det\left[I_{|\mathcal{M}|} + \mathsf{SNR} S_{\mathcal{M}}(e^{j\omega})\right] \geq \mathrm{tr}\left[S_{\mathcal{M}}(e^{j\omega})\right]\mathsf{SNR} - \frac{1}{2}\mathrm{tr}\left[S_{\mathcal{M}}^2(e^{j\omega})\right]\mathsf{SNR}^2$$

we can upper-bound $C_p(\mathsf{SNR})$ by

$$C_p(\mathsf{SNR}) \leq \frac{\mathsf{SNR}}{4\pi} \max_{\mathcal{M}\subseteq\{1,\dots,T\}} \frac{1}{|\mathcal{M}|}\int_{-\pi}^{\pi} \mathrm{tr}[S_{\mathcal{M}}^2(e^{j\omega})]\mathrm{d}\omega.$$

It can be seen from Corollary 1 that this upper bound is a good approximation of $C_p(\mathsf{SNR})$ in the low-SNR regime.

Now we proceed to compute $C_p(\mathsf{SNR})$ in the following examples.

Example 5: When the channel changes independently from block to block, $C_p(\mathsf{SNR})$ is equal to

$$1 - \min_{\mathcal{M}\subseteq\{1,\dots,T\}} \frac{1}{|\mathcal{M}|\mathsf{SNR}} \log\det\left[I_{|\mathcal{M}|} + \mathsf{SNR} R_{\mathcal{M}}(0)\right]$$

where $R_{\mathcal{M}}(0)$ is an $|\mathcal{M}| \times |\mathcal{M}|$ principal minor of $R(0)$ with the indices of columns and rows specified by $\mathcal{M}$. If we further

[4]This equality can be proved by applying the eigenvalue decomposition to $S_{\mathcal{M}}(e^{j\omega})$ and then using the inequality $\log(1+x) \geq x - \frac{1}{2}x^2$.

$$C_p(\mathsf{SNR}) = 1 - \min_{\mathcal{M}\subseteq\{1,\dots,T\}} \frac{\sum_{i=0}^{|\mathcal{M}|} i\mu(\mathrm{rank}(S_{\mathcal{M}}(e^{j\omega})) = i)\log\mathsf{SNR}}{2\pi|\mathcal{M}|\mathsf{SNR}} + o\left(\frac{\log\mathsf{SNR}}{\mathsf{SNR}}\right). \quad (19)$$

$$\begin{aligned}
C_p(\mathsf{SNR}) &= 1 - \frac{1}{2\pi\mathsf{SNR}} \min_{\mathcal{M}\subseteq\{1,\dots,T\}} \Psi(\mathcal{M}, \mathsf{SNR}) \\
&= 1 - \frac{1}{2\pi\mathsf{SNR}} \min_{\mathcal{M}\subseteq\{1,\dots,T\}} \left\{\frac{1}{|\mathcal{M}|}\int_{-\pi}^{\pi} \log\det\left[\frac{1}{\mathsf{SNR}} I_{|\mathcal{M}|} + S_{\mathcal{M}}(e^{j\omega})\right]\mathrm{d}\omega + 2\pi\log\mathsf{SNR}\right\} \\
&= 1 + \max_{\mathcal{M}\subseteq\{1,\dots,T\}} \frac{\sum_{i=0}^{|\mathcal{M}|}(|\mathcal{M}| - i)\mu\left(\mathrm{rank}(S_{\mathcal{M}}(e^{j\omega})) = i\right)\log\mathsf{SNR}}{2\pi|\mathcal{M}|\mathsf{SNR}} - \frac{\log\mathsf{SNR}}{\mathsf{SNR}} + o\left(\frac{\log\mathsf{SNR}}{\mathsf{SNR}}\right) \\
&= 1 - \min_{\mathcal{M}\subseteq\{1,\dots,T\}} \frac{\sum_{i=0}^{|\mathcal{M}|} i\mu\left(\mathrm{rank}(S_{\mathcal{M}}(e^{j\omega})) = i\right)\log\mathsf{SNR}}{2\pi|\mathcal{M}|\mathsf{SNR}} + o\left(\frac{\log\mathsf{SNR}}{\mathsf{SNR}}\right).
\end{aligned}$$

$$\begin{aligned}
C_p(\mathsf{SNR}) &= 1 - \frac{1}{2\pi\mathsf{SNR}} \min_{\mathcal{M}\subseteq\{1,\dots,T\}} \Psi(\mathcal{M}, \mathsf{SNR}) \\
&= 1 - \frac{1}{2\pi} \min_{\mathcal{M}\subseteq\{1,\dots,T\}} \frac{1}{|\mathcal{M}|}\left\{\int_{-\pi}^{\pi} \mathrm{tr}\left[S_{\mathcal{M}}(e^{j\omega})\right]\mathrm{d}\omega - \frac{1}{2}\int_{-\pi}^{\pi} \mathrm{tr}\left[S_{\mathcal{M}}^2(e^{j\omega})\right]\mathsf{SNR}\mathrm{d}\omega\right\} + o(\mathsf{SNR}) \\
&= \frac{\mathsf{SNR}}{4\pi} \max_{\mathcal{M}\subseteq\{1,\dots,T\}} \frac{1}{|\mathcal{M}|}\int_{-\pi}^{\pi} \mathrm{tr}\left[S_{\mathcal{M}}^2(e^{j\omega})\right]\mathrm{d}\omega + o(\mathsf{SNR})
\end{aligned}$$

let the fading remain constant within a block, then all the entries of $R(0)$ are one. It is not difficult to show that

$$\frac{1}{|\mathcal{M}|\mathsf{SNR}}\log\det\left[I_{|\mathcal{M}|}+\mathsf{SNR}R_{\mathcal{M}}(0)\right]=\frac{\log(1+|\mathcal{M}|\mathsf{SNR})}{|\mathcal{M}|\mathsf{SNR}}$$

which is minimized when $|\mathcal{M}|=T$, i.e., $\mathcal{M}=\{1,2,\ldots,T\}$. So we have

$$C_p(\mathsf{SNR})=1-\frac{\log(1+T\mathsf{SNR})}{T\mathsf{SNR}}$$

as shown in [18].

Example 6: Consider the case in which the fading process satisfies the following conditions:

1) all the off-diagonal entries of $R(0)$ are equal to α, where $\alpha\in\mathbb{C}$ is a constant;
2) all the entries of $R(i)$ are equal to β_i for any nonzero integer i, where $\beta_i\in\mathbb{C}$ is a constant that depends only on i.

We also know that the diagonal entries of $R(0)$ are all one. So for any fixed ω ($-\pi\le\omega\le\pi$), all the diagonal entries of $I+\mathsf{SNR}S(e^{j\omega})$ are identical, and all the off-diagonal entries of $I+\mathsf{SNR}S(e^{j\omega})$ are identical. It then follows from Szasz's inequality [20] that for any $\omega\in[-\pi,\pi]$

$$\left\{\det[I_{|\mathcal{M}|}+\mathsf{SNR}S_{\mathcal{M}}(e^{j\omega})\right\}^{\frac{1}{|\mathcal{M}|}}$$

is minimized when $\mathcal{M}=\{1,2,\ldots,T\}$. In this case, we therefore have

$$C_p(\mathsf{SNR})=1-\frac{1}{2\pi T\mathsf{SNR}}\int_{-\pi}^{\pi}\log\det\left(I+\mathsf{SNR}S(e^{j\omega})\right)\mathrm{d}\omega.$$

If the fading remains constant within a block, then for any fixed ω, all the entries of $S(e^{j\omega})$ are identical, which we shall denote by $s(e^{j\omega})$. It can be shown that

$$\det\left[I+\mathsf{SNR}S(e^{j\omega})\right]=1+T\mathsf{SNR}s(e^{j\omega})$$

which yields

$$C_p(\mathsf{SNR})=1-\frac{1}{2\pi T\mathsf{SNR}}\int_{-\pi}^{\pi}\log\left[1+T\mathsf{SNR}s(e^{j\omega})\right]\mathrm{d}\omega. \quad (21)$$

We can see from (21) that $C_p(\mathsf{SNR})$ is a monotonically increasing function of T and SNR. Intuitively, as T is increased, the receiver can estimate the channel more accurately, and thus the capacity per unit energy of the noncoherent channel should converge to that of the coherent channel, which is equal to one. As SNR goes to infinity, $C_p(\mathsf{SNR})$ should also converge to one since flash signaling can be used if there is no peak power constraint (i.e., $\mathsf{SNR}=\infty$) [21]. Moreover, (21) provides a precise characterization of the interplay between the coherence time and signal peakiness, stating that the capacity per unit energy is unaffected as long as the product of T and SNR is fixed. See [22], [23] for a related discussion.

For the special case, where the fading is a block Gauss–Markov process, i.e., all the entries of $R(i)$ are equal to ρ^i if $i\ge 0$, and equal to $(\rho^*)^{-i}$ if $i<0$, for some $\rho\in\mathbb{C}$, with $0\le|\rho|<1$, we have

$$1+T\mathsf{SNR}s(e^{j\omega})=|\varphi(e^{j\omega})|$$

where

$$\varphi(z)=\frac{(\rho^*z-\gamma_0)^2}{\gamma_0|\rho|^2\left(z-\frac{1}{\rho^*}\right)^2}$$

and

$$\gamma_0=\frac{b+\sqrt{b^2-4|\rho|^2}}{2}$$

with $b=1+T\mathsf{SNR}+|\rho|^2(1-T\mathsf{SNR})$. The function φ is analytic and nonzero in a neighborhood of the unit disk. Thus, by Jensen's formula

$$\begin{aligned}C_p(\mathsf{SNR})&=1-\frac{1}{2\pi T\mathsf{SNR}}\int_{-\pi}^{\pi}\log|\varphi(e^{j\omega})|\mathrm{d}\omega\\&=1-\frac{1}{T\mathsf{SNR}}\log|\varphi(0)|\\&=1-\frac{1}{T\mathsf{SNR}}\log\gamma_0\end{aligned}$$

from which we can recover [18, Corollary 4.1] by setting $T=1$.

Finding the optimal $\mathcal{M}^*$ is a difficult problem in general. Moreover, as shown in the following example, the optimal $\mathcal{M}^*$ may depend on the SNR level.

Example 7: Let the fading process be independent from block to block with

$$S(e^{j\omega})=R(0)=\begin{pmatrix}1&1&\rho^*\\1&1&\rho^*\\\rho&\rho&1\end{pmatrix}$$

where $|\rho|\in[0,1]$.

The following is shown in Appendix IX.

1) When $0\le|\rho|\le\frac{1}{2}$, the optimal $\mathcal{M}^*$ is $\{1,2\}$, and
$$\begin{aligned}C_p(\mathsf{SNR})&=1-\frac{1}{2\pi\mathsf{SNR}}\Psi(\{1,2\},\mathsf{SNR})\\&=1-\frac{\log(1+2\mathsf{SNR})}{2\mathsf{SNR}}.\end{aligned}$$
2) When $\frac{1}{2}<|\rho|<1$, we get the equation at the bottom of the page.
3) When $|\rho|=1$, the optimal $\mathcal{M}^*$ is $\{1,2,3\}$, and
$$C_p(\mathsf{SNR})=1-\frac{1}{2\pi\mathsf{SNR}}\Psi(\{1,2,3\},\mathsf{SNR})$$

$$\mathcal{M}^*=\begin{cases}\{1,2,3\}, & \mathsf{SNR}<\frac{2|\rho|-1}{2(1-|\rho|)^2}\\ \{1,2\}\text{ or }\{1,2,3\}, & \mathsf{SNR}=\frac{2|\rho|-1}{2(1-|\rho|)^2}\\ \{1,2\}, & \mathsf{SNR}>\frac{2|\rho|-1}{2(1-|\rho|)^2}\end{cases}$$

and

$$C_p(\mathsf{SNR})=\begin{cases}1-\frac{\log(1+3\mathsf{SNR}+2\mathsf{SNR}^2-2|\rho|^2\mathsf{SNR}^2)}{3\mathsf{SNR}}, & \mathsf{SNR}<\frac{2|\rho|-1}{2(1-|\rho|)^2}\\ 1-\frac{\log(1+2\mathsf{SNR})}{2\mathsf{SNR}}, & \mathsf{SNR}\ge\frac{2|\rho|-1}{2(1-|\rho|)^2}.\end{cases}$$

$$= 1 - \frac{\log(1+3\mathsf{SNR})}{3\mathsf{SNR}}.$$

It can be verified that this result is consistent with the asymptotic analysis in Corollary 1. Since $S_{\mathcal{M}}(e^{j\omega}) = R_{\mathcal{M}}(0)$ for any $\mathcal{M} \subseteq \{1,\ldots,T\}$, it is easy to see that

$$\sum_{i=0}^{|\mathcal{M}|} \frac{i}{|\mathcal{M}|}\mu(\operatorname{rank}(S_{\mathcal{M}}(e^{j\omega})) = i)$$

is minimized at $\mathcal{M} = \{1,2\}$ if $|\rho| < 1$, and minimized at $\mathcal{M} = \{1,2,3\}$ if $|\rho| = 1$. Therefore, by (19), the optimal $\mathcal{M}^*$ at high SNR should be $\{1,2\}$ if $|\rho| < 1$, and should be $\{1,2,3\}$ if $|\rho| = 1$. Since

$$\begin{aligned}
S^2_{\{1\}}(e^{j\omega}) &= S^2_{\{2\}}(e^{j\omega}) = S^2_{\{3\}}(e^{j\omega}) = 1\\
S^2_{\{1,3\}}(e^{j\omega}) &= S^2_{\{2,3\}}(e^{j\omega}) = \begin{pmatrix} 1+|\rho|^2 & 2\rho^* \\ 2\rho & 1+|\rho|^2 \end{pmatrix}\\
S^2_{\{1,2\}}(e^{j\omega}) &= \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}\\
S^2_{\{1,2,3\}}(e^{j\omega}) &= \begin{pmatrix} 2+|\rho|^2 & 2+|\rho|^2 & 3\rho^* \\ 2+|\rho|^2 & 2+|\rho|^2 & 3\rho^* \\ 3\rho & 3\rho & 1+2|\rho|^2 \end{pmatrix}
\end{aligned}$$

it follows that $\frac{1}{|\mathcal{M}|}\operatorname{tr}\left[S^2_{\mathcal{M}}(e^{j\omega})\right]$ is maximized at $\mathcal{M} = \{1,2\}$ if $|\rho| < \frac{1}{2}$, and maximized at $\mathcal{M} = \{1,2,3\}$ if $|\rho| > \frac{1}{2}$. Therefore, by (20), the optimal $\mathcal{M}^*$ at low SNR should be $\{1,2\}$ if $|\rho| < \frac{1}{2}$, and should be $\{1,2,3\}$ if $|\rho| > \frac{1}{2}$.

Intuitively, if $|\rho|$ is close to 1, we can approximate $R(0)$ by the all-one matrix, and then it follows from Example 5 that the optimal $\mathcal{M}^*$ is $\{1,2,3\}$. The approximation breaks down at high SNR since Corollary 1 implies that the optimal $\mathcal{M}^*$ should be $\{1,2\}$ as $\mathsf{SNR} \to \infty$.

VI. Conclusion

We conducted a detailed study of the block-stationary Gaussian fading channel model introduced in [8]. We derived single-letter upper and lower bounds on channel capacity, and used these bounds to characterize the asymptotic behavior of channel capacity. Specifically, we computed the asymptotic ratio between the noncoherent channel capacity and the logarithm of the SNR in the high-SNR regime. This result generalizes many previous results on noncoherent capacity. We showed that the behavior of the channel capacity depends critically on channel modeling. We also derived an expression for the capacity per unit energy for the block-stationary fading model. It is clearly of interest to generalize these results to the multiple-antenna scenario [2], [5], but such an extension seems technically nontrivial.

Another direction that we explored was the interplay between the codeword length, SNR level, and decoding error probability. We showed that for noncoherent symbol-by-symbol stationary fading channels, the codeword length must scale with SNR in order to guarantee that the communication rate can grow linearly with $\log \mathsf{SNR}$ with decoding error probability bounded away from one, and we found a necessary condition for the growth rate of the codeword length. We believe that a more complete characterization of the interplay between the codeword length, SNR level, and decoding error probability would be of theoretical significance and practical value.

Appendix I
Proof of Monotonicity

Proposition 2: Let $\boldsymbol{x}$, $\boldsymbol{z}$, and $(\boldsymbol{h},\boldsymbol{h}_1,\boldsymbol{h}_2)$ be independent random variables. Suppose $\mathbb{E}\boldsymbol{h} = \mathbb{E}\boldsymbol{h}_1 = \mathbb{E}\boldsymbol{h}_2 = 0$, and $(\boldsymbol{h},\boldsymbol{h}_i)$ are jointly Gaussian, $i = 1,2$. If $\operatorname{var}(\boldsymbol{h}|\boldsymbol{h}_1) \le \operatorname{var}(\boldsymbol{h}|\boldsymbol{h}_2)$, then

$$I(\boldsymbol{x};\boldsymbol{hx}+\boldsymbol{z}|\boldsymbol{h}_1) \ge I(\boldsymbol{x};\boldsymbol{hx}+\boldsymbol{z}|\boldsymbol{h}_2).$$

Proof: If $\operatorname{var}(\boldsymbol{h}|\boldsymbol{h}_1) \le \operatorname{var}(\boldsymbol{h}|\boldsymbol{h}_2)$, then we can construct a zero-mean Gaussian random variable $\Delta\boldsymbol{h}$ independent of everything else such that $(\boldsymbol{x},\boldsymbol{hx}+\boldsymbol{z},\mathbb{E}(\boldsymbol{h}|\boldsymbol{h}_1+\Delta\boldsymbol{h}))$ have the same joint distribution as $(\boldsymbol{x},\boldsymbol{hx}+\boldsymbol{z},\mathbb{E}(\boldsymbol{h}|\boldsymbol{h}_2))$. Now we have

$$\begin{aligned}
I(\boldsymbol{x};\boldsymbol{hx}+\boldsymbol{z}|\boldsymbol{h}_2) &= I(\boldsymbol{x};\boldsymbol{hx}+\boldsymbol{z}|\mathbb{E}(\boldsymbol{h}|\boldsymbol{h}_2))\\
&= I(\boldsymbol{x};\boldsymbol{hx}+\boldsymbol{z}|\mathbb{E}(\boldsymbol{h}|\boldsymbol{h}_1+\Delta\boldsymbol{h}))\\
&= I(\boldsymbol{x};\boldsymbol{hx}+\boldsymbol{z}|\boldsymbol{h}_1+\Delta\boldsymbol{h})\\
&= I(\boldsymbol{x};\boldsymbol{hx}+\boldsymbol{z},\boldsymbol{h}_1+\Delta\boldsymbol{h})
\end{aligned}$$

where the first and third equalities follow from the fact that $\mathbb{E}(\boldsymbol{h}|\boldsymbol{h}_2)$ and $\mathbb{E}(\boldsymbol{h}|\boldsymbol{h}_1+\Delta\boldsymbol{h})$ are just scaled versions of $\boldsymbol{h}_2$ and $\boldsymbol{h}_1+\Delta\boldsymbol{h}$, respectively. To complete the proof, we note that the data processing theorem implies

$$\begin{aligned}
I(\boldsymbol{x};\boldsymbol{hx}+\boldsymbol{z},\boldsymbol{h}_1+\Delta\boldsymbol{h}) &\le I(\boldsymbol{x};\boldsymbol{hx}+\boldsymbol{z},\boldsymbol{h}_1)\\
&= I(\boldsymbol{x};\boldsymbol{hx}+\boldsymbol{z}|\boldsymbol{h}_1). \qquad \square
\end{aligned}$$

Now we proceed to prove the monotonicity. By the block stationarity of the fading process, we have

$$\begin{aligned}
I&\left(\boldsymbol{x}_{i+jT};\boldsymbol{y}_{i+jT}\middle|\mathbb{E}\left(\boldsymbol{h}_{i+jT}\middle|\left\{\boldsymbol{h}_k+\frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=1}^{i+jT-1}\right)\right)\\
&= I\left(\boldsymbol{x}_i;\boldsymbol{y}_i\middle|\mathbb{E}\left(\boldsymbol{h}_i\middle|\left\{\boldsymbol{h}_k+\frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=1-jT}^{i-1}\right)\right). \qquad (22)
\end{aligned}$$

It follows by Proposition 2 that for any $j_1 < j_2$

$$\begin{aligned}
I&\left(\boldsymbol{x}_i;\boldsymbol{y}_i\middle|\mathbb{E}\left(\boldsymbol{h}_i\middle|\left\{\boldsymbol{h}_k+\frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=1-j_1T}^{i-1}\right)\right)\\
&\le I\left(\boldsymbol{x}_i;\boldsymbol{y}_i\middle|\mathbb{E}\left(\boldsymbol{h}_i\middle|\left\{\boldsymbol{h}_k+\frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=1-j_2T}^{i-1}\right)\right). \qquad (23)
\end{aligned}$$

Equations (22) and (23) together imply that

$$\left\{I\left(\boldsymbol{x}_{i+jT};\boldsymbol{y}_{i+jT}\middle|\mathbb{E}\left(\boldsymbol{h}_{i+jT}\middle|\left\{\boldsymbol{h}_k+\frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=1}^{i+jT-1}\right)\right)\right\}_j$$

is a monotone increasing sequence.

For every

$$\mathbb{E}\left(\boldsymbol{h}_i\middle|\left\{\boldsymbol{h}_k+\frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=1-jT}^{i-1}\right)$$

we can construct a random variable $\boldsymbol{\Delta}_j \sim \mathcal{CN}(0, \delta_j)$ independent of everything else such that

$$\mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=1-jT}^{i-1}\right) = \mathbb{E}\left(\boldsymbol{h}_i \middle| \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right) + \Delta_j\right)$$

in distribution. Clearly, $\delta_j \to 0$ as $j \to \infty$. Moreover, it is easy to see that

$$I\left(\boldsymbol{x}_i; \boldsymbol{y}_i \middle| \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=1-jT}^{i-1}\right)\right) = I\left(\boldsymbol{x}_i; \boldsymbol{y}_i \middle| \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right) + \Delta_j\right). \tag{24}$$

Combining (22) and (24), we get

$$\begin{aligned}\lim_{j\to\infty} I\left(\boldsymbol{x}_{i+jT}; \boldsymbol{y}_{i+jT} \middle| \mathbb{E}\left(\boldsymbol{h}_{i+jT} \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=1}^{i+jT-1}\right)\right) &= \lim_{j\to\infty} I\left(\boldsymbol{x}_i; \boldsymbol{y}_i \middle| \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=1-jT}^{i-1}\right)\right)\\ &= \lim_{j\to\infty} I\left(\boldsymbol{x}_i; \boldsymbol{y}_i \middle| \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right) + \Delta_j\right).\end{aligned}$$

Note that

$$\begin{aligned}I\left(\boldsymbol{x}_i; \boldsymbol{y}_i \middle| \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right) + \Delta_j\right) &= I\left(\boldsymbol{x}_i; \boldsymbol{y}_i, \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right) + \Delta_j\right)\\ &= h\left(\boldsymbol{y}_i, \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right) + \Delta_j\right)\\ &\quad - h\left(\boldsymbol{y}_i, \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right) + \Delta_j \middle| \boldsymbol{x}_i\right).\end{aligned}$$

By [3, Lemma 6.11], we get

$$\lim_{j\to\infty} h\left(\boldsymbol{y}_i, \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right) + \Delta_j\right) = h\left(\boldsymbol{y}_i, \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right)\right).$$

Since conditioned on $\boldsymbol{x}_i$

$$\left(\boldsymbol{y}_i, \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right) + \Delta_j\right)$$

are jointly Gaussian with uniformly bounded differential entropy for any realization of $\boldsymbol{x}_i$ (note: $|\boldsymbol{x}_i|^2 \le \mathsf{SNR}$), it follows by the dominated convergence theorem that

$$\lim_{j\to\infty} h\left(\boldsymbol{y}_i, \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right) + \Delta_j \middle| \boldsymbol{x}_i\right) = h\left(\boldsymbol{y}_i, \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right) \middle| \boldsymbol{x}_i\right).$$

Therefore

$$\lim_{j\to\infty} I\left(\boldsymbol{x}_{i+jT}; \boldsymbol{y}_{i+jT} \middle| \mathbb{E}\left(\boldsymbol{h}_{i+jT} \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=1}^{i+jT-1}\right)\right) = \lim_{j\to\infty} I\left(\boldsymbol{x}_i; \boldsymbol{y}_i \middle| \mathbb{E}\left(\boldsymbol{h}_i \middle| \left\{\boldsymbol{h}_k + \frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right)\right).$$

Appendix II
Proof of Lemma 1

By eigenvalue decomposition, we write

$$A(\xi) = U(\xi)D(\xi)U^\dagger(\xi),$$

and

$$A(\xi) + \epsilon I_M = U(\xi)(D(\xi) + \epsilon I_M)U^\dagger(\xi)$$

where $U(\xi)$ is a unitary matrix, and $D(\xi)$ is a diagonal matrix with nonnegative diagonal entries. Since $\mathrm{rank}(A(\xi)) = \mathrm{rank}(D(\xi))$, define

$$\Omega_i = \{\xi : \mathrm{rank}(A(\xi)) = \mathrm{rank}(D(\xi)) = i\}, \qquad 0 \le i \le M.$$

We have

$$\begin{aligned}\lim_{\epsilon\to 0} \frac{\int_{\xi_0}^{\xi_1} \log\det[A(\xi) + \epsilon I_M]\,\mathrm{d}\xi}{\log\epsilon} &= \lim_{\epsilon\to 0} \frac{\int_{\xi_1}^{\xi_2} \log\det[D(\xi) + \epsilon I_M]\,\mathrm{d}\xi}{\log\epsilon}\\ &= \lim_{\epsilon\to 0} \frac{-\sum_{i=0}^{M} \int_{\Omega_i} \log\det[D(\xi) + \epsilon I_M]\,\mathrm{d}\xi}{\log\epsilon}.\end{aligned}$$

For $\xi \in \Omega_i$ (possibly after permutating diagonal entries), we can write $D(\xi) = \mathrm{diag}\{d_1(\xi), \ldots, d_i(\xi), 0, \ldots, 0\}$, where $d_j(\xi) > 0$, $1 \le j \le i$. Therefore

$$\begin{aligned}\lim_{\epsilon\to 0} \frac{\int_{\xi_0}^{\xi_1} \log\det[A(\xi) + \epsilon I_M]\,\mathrm{d}\xi}{\log\epsilon} &= \lim_{\epsilon\to 0} \frac{-\sum_{i=0}^{M} \int_{\Omega_i} \sum_{j=1}^{i} \log[d_j(\xi) + \epsilon]\mathrm{d}\xi}{\log\epsilon}\\ &\quad + \sum_{i=0}^{M} (M - i)\mu(\mathrm{rank}(A(\xi)) = i)\end{aligned}$$

where $\mu(\mathrm{rank}(A(\xi)) = i)$ is the Lebesgue measure of Ω_i. By the argument in [6, Sec. VIII], it can be shown that

$$\lim_{\epsilon\to 0} \frac{\int_{\Omega_i} \log[d_i(\xi)+\epsilon]\mathrm{d}\xi}{\log\epsilon} = 0.$$

So we have

$$\lim_{\epsilon\to 0} \frac{\int_{\xi_0}^{\xi_1} \log\det\left[A(\xi)+\epsilon I\right]\mathrm{d}\xi}{\log\epsilon} = \sum_{i=0}^{M}(M-i)\mu(\mathrm{rank}(A(\xi))=i).$$

Appendix III
Proof of Theorem 1

Define

$$\sigma_i(\mathsf{SNR}) = \mathrm{var}\left(\boldsymbol{h}_i + \frac{1}{\sqrt{\mathsf{SNR}}} \times \boldsymbol{z}_i \,\middle|\, \left\{\boldsymbol{h}_k + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right), \quad i=1,2,\ldots,T.$$

In the lower bound (5), let $\boldsymbol{x}_i$ be uniformly distributed over the set $\left\{z\in\mathbb{C} : \frac{\sqrt{\mathsf{SNR}}}{2} \le \|z\| \le \sqrt{\mathsf{SNR}}\right\}$. By Lemma 2

$$\begin{aligned} I\left(\boldsymbol{x}_i;\boldsymbol{h}_i\boldsymbol{x}_i+\boldsymbol{z}_i \,\middle|\, \mathbb{E}\left(\boldsymbol{h}_i \,\middle|\, \left\{\boldsymbol{h}_k + \frac{4}{\mathsf{SNR}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right)\right) &\ge -\log\left[\sigma_i\left(\frac{\mathsf{SNR}}{4}\right) - \frac{12}{5\mathsf{SNR}}\right] \\ &\quad + \log\left(1-\sigma_i\left(\frac{\mathsf{SNR}}{4}\right)+\frac{4}{\mathsf{SNR}}\right) - \gamma - \log\frac{5e}{6} \\ &= -\log\left[\sigma_i\left(\frac{\mathsf{SNR}}{4}\right) - \frac{12}{5\mathsf{SNR}}\right] + o(\log\mathsf{SNR}). \end{aligned}$$

Since $\sigma_i\left(\frac{\mathsf{SNR}}{4}\right) \ge \frac{4}{\mathsf{SNR}}$, it follows that

$$\liminf_{\mathsf{SNR}\to\infty} \frac{I\left(\boldsymbol{x}_i;\boldsymbol{h}_i\boldsymbol{x}_i+\boldsymbol{z}_i \,\middle|\, \mathbb{E}\left(\boldsymbol{h}_i \,\middle|\, \left\{\boldsymbol{h}_k + \frac{4}{\mathsf{SNR}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right)\right)}{\log\mathsf{SNR}} \ge \liminf_{\mathsf{SNR}\to\infty} \frac{-\log\left[\sigma_i\left(\frac{\mathsf{SNR}}{4}\right)\right]}{\log\mathsf{SNR}}.$$

Let

$$\Sigma\left(\frac{\mathsf{SNR}}{4}\right) = L\left(\frac{\mathsf{SNR}}{4}\right)\Lambda\left(\frac{\mathsf{SNR}}{4}\right)L^{\dagger}\left(\frac{\mathsf{SNR}}{4}\right)$$

where $L\left(\frac{\mathsf{SNR}}{4}\right)$ is a lower triangular matrix with unit diagonal entries, and

$$\Lambda\left(\frac{\mathsf{SNR}}{4}\right) = \mathrm{diag}\left\{\sigma_1\left(\frac{\mathsf{SNR}}{4}\right), \sigma_2\left(\frac{\mathsf{SNR}}{4}\right), \ldots, \sigma_T\left(\frac{\mathsf{SNR}}{4}\right)\right\}.$$

We have

$$\begin{aligned} \det\left[\Sigma\left(\frac{\mathsf{SNR}}{4}\right)\right] &= \det\left[L\left(\frac{\mathsf{SNR}}{4}\right)\right] \times \det\left[\Lambda\left(\frac{\mathsf{SNR}}{4}\right)\right]\det\left[L^{\dagger}\left(\frac{\mathsf{SNR}}{4}\right)\right] \\ &= \det\left[\Lambda\left(\frac{\mathsf{SNR}}{4}\right)\right] \\ &= \prod_{i=1}^{T}\sigma_i\left(\frac{\mathsf{SNR}}{4}\right). \end{aligned} \tag{25}$$

Therefore

$$\begin{aligned} \liminf_{\mathsf{SNR}\to\infty} \frac{C(\mathsf{SNR})}{\log\mathsf{SNR}} &\ge, \liminf_{\mathsf{SNR}\to\infty} \frac{\sum_{i=1}^{T} I\left(\boldsymbol{x}_i;\boldsymbol{h}_i\boldsymbol{x}_i+\boldsymbol{z}_i \,\middle|\, \mathbb{E}\left(\boldsymbol{h}_i \,\middle|\, \left\{\boldsymbol{h}_k+\frac{4}{\mathsf{SNR}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right)\right)}{T\log\mathsf{SNR}} \\ &\ge \liminf_{\mathsf{SNR}\to\infty} \frac{-\log\left[\prod_{i=1}^{T}\sigma_i\left(\frac{\mathsf{SNR}}{4}\right)\right]}{\log\mathsf{SNR}} \\ &= \liminf_{\mathsf{SNR}\to\infty} \frac{-\log\det\left[\Sigma\left(\frac{\mathsf{SNR}}{4}\right)\right]}{\log\mathsf{SNR}} \\ &= \liminf_{\mathsf{SNR}\to\infty} \frac{-\log\det\left[\Sigma\left(\mathsf{SNR}\right)\right]}{\log\mathsf{SNR}}. \end{aligned} \tag{26}$$

We use (8) to derive an upper bound on $\frac{\log C(\mathsf{SNR})}{\log\mathsf{SNR}}$. First it is easy to see the inequality in (27), shown at the bottom of the page. Note that $\sup_{p_{\boldsymbol{x}_k}\in\mathcal{P}_1(\mathsf{SNR})} I(\boldsymbol{x}_k;\boldsymbol{y}_k)$ is the capacity of the memoryless noncoherent Rayleigh-fading channel (see [3, eq. (141)] for a nonasymptotic upper bound), and we have

$$\sup_{p_{\boldsymbol{x}_k}\in\mathcal{P}_1(\mathsf{SNR})} I(\boldsymbol{x}_k;\boldsymbol{y}_k) = o(\log\mathsf{SNR}). \tag{28}$$

Now we proceed to upper-bound the first term in (27) as in (29) at the top of the following page. Therefore, we get (30), also at the top of the following page. The desired result follows by combining (26) and (30).

Appendix IV
Proof of Theorem 2

Define

$$\sigma_i(\infty) = \mathrm{var}\left(\boldsymbol{h}_i \,\middle|\, \{\boldsymbol{h}_k\}_{k=-\infty}^{i-1}\right) \quad i=1,2,\ldots,T.$$

$$\begin{aligned} &\sup_{p_{\boldsymbol{x}_k}\in\mathcal{P}_1(\mathsf{SNR})} I\left(\boldsymbol{x}_k, \mathbb{E}\left(\boldsymbol{h}_k \,\middle|\, \left\{\boldsymbol{h}_t+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{k-1}\right);\boldsymbol{y}_k\right) \\ &\quad\le \sup_{p_{\boldsymbol{x}_k}\in\mathcal{P}_1(\mathsf{SNR})} I\left(\mathbb{E}\left(\boldsymbol{h}_k \,\middle|\, \left\{\boldsymbol{h}_t+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{k-1}\right);\boldsymbol{y}_k \,\middle|\, \boldsymbol{x}_k\right) + \sup_{p_{\boldsymbol{x}_k}\in\mathcal{P}_1(\mathsf{SNR})} I(\boldsymbol{x}_k;\boldsymbol{y}_k). \end{aligned} \tag{27}$$

$$\begin{aligned}
&\sup_{p_{\boldsymbol{x}_k}\in\mathcal{P}_1(\mathsf{SNR})} I\left(\mathbb{E}\left(\boldsymbol{h}_k \middle| \left\{\boldsymbol{h}_t+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{k-1}\right);\boldsymbol{y}_k\middle|\boldsymbol{x}_k\right)\\
&\quad=\sup_{p_{\boldsymbol{x}_k}\in\mathcal{P}_1(\mathsf{SNR})}\mathbb{E}\left\{\log\left[\frac{1+|\boldsymbol{x}_k|^2}{1+|\boldsymbol{x}_k|^2\cdot\mathrm{var}\left(\boldsymbol{h}_k\middle|\left\{\boldsymbol{h}_t+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{k-1}\right)}\right]\right\}\\
&\quad\le\log\left[\frac{1+\mathsf{SNR}}{1+\mathsf{SNR}\cdot\mathrm{var}\left(\boldsymbol{h}_k\middle|\left\{\boldsymbol{h}_t+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{k-1}\right)}\right]\\
&\quad=\log\left[\frac{1+\mathsf{SNR}}{\mathsf{SNR}\cdot\sigma_k(\mathsf{SNR})}\right].
\end{aligned}\tag{29}$$

$$\begin{aligned}
\limsup_{\mathsf{SNR}\to\infty}\frac{C(\mathsf{SNR})}{\log\mathsf{SNR}}&\le\limsup_{\mathsf{SNR}\to\infty}\frac{\sum_{k=1}^{T}\sup_{p_{\boldsymbol{x}_k}\in\mathcal{P}_1(\mathsf{SNR})} I\left(\mathbb{E}\left(\boldsymbol{h}_k \middle| \left\{\boldsymbol{h}_t+\frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{k-1}\right);\boldsymbol{y}_k\middle|\boldsymbol{x}_k\right)}{T\log\mathsf{SNR}}\\
&\le\limsup_{\mathsf{SNR}\to\infty}\frac{-\log\left[\prod_{k=1}^{T}\sigma_k(\mathsf{SNR})\right]}{\log\mathsf{SNR}}\\
&=\limsup_{\mathsf{SNR}\to\infty}\frac{-\log\det\left[\Sigma(\mathsf{SNR})\right]}{\log\mathsf{SNR}}.
\end{aligned}\tag{30}$$

Similar to (25), we have

$$\det\left[\Sigma(\infty)\right]=\prod_{i=1}^{T}\sigma_i(\infty).$$

Therefore, $\det[\Sigma(\infty)]>0$ implies $\sigma_i(\infty)>0$ for all i.

Note that if $\frac{1}{x_{\min}}<\delta$ for some $\delta>0$, then it follows from Proposition 2 that

$$\begin{aligned}
&I\left(\boldsymbol{x}_i;\boldsymbol{h}_i\boldsymbol{x}_i+\boldsymbol{z}_i\middle|\mathbb{E}\left(\boldsymbol{h}_i\middle|\left\{\boldsymbol{h}_k+\frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right)\right)\\
&\qquad\ge I\left(\boldsymbol{x}_i;\boldsymbol{h}_i\boldsymbol{x}_i+\boldsymbol{z}_i\middle|\mathbb{E}\left(\boldsymbol{h}_i\middle|\left\{\boldsymbol{h}_k+\delta\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right)\right).
\end{aligned}$$

In the lower bound (5), let $\log|\boldsymbol{x}_i|^2$ be uniformly distributed over the interval $[\log x_{\min}^2,\log\mathsf{SNR}]$. As $\log x_{\min}^2$ grows sublinearly in $\log\mathsf{SNR}$ to infinity, we get the first equation at the bottom of the page, where the last equality follows from [3, Proposition 4.23]. Therefore, we get (31), also at the bottom of the page. Since (31) holds for arbitrary positive δ, it follows that

$$\begin{aligned}
&\liminf_{\mathsf{SNR}\to\infty}\left[C(\mathsf{SNR})-\log\log\mathsf{SNR}\right]\\
&\quad\ge-1-\gamma-\frac{1}{T}\lim_{\delta\to0}\sum_{i=1}^{T}\log\mathrm{var}\left(\boldsymbol{h}_i\middle|\left\{\boldsymbol{h}_j+\delta\boldsymbol{z}_j\right\}_{j=-\infty}^{i-1}\right)\\
&\quad=-1-\gamma-\frac{1}{T}\log\det\left[\Sigma(\infty)\right].
\end{aligned}\tag{32}$$

$$\begin{aligned}
&\liminf_{\mathsf{SNR}\to\infty}\left[I\left(\boldsymbol{x}_i;\boldsymbol{h}_i\boldsymbol{x}_i+\boldsymbol{z}_i\middle|\mathbb{E}\left(\boldsymbol{h}_i\middle|\left\{\boldsymbol{h}_k+\frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right)\right)-\log\log\mathsf{SNR}\right]\\
&\quad\ge\liminf_{\mathsf{SNR}\to\infty}\left[I\left(\boldsymbol{x}_i;\boldsymbol{h}_i\boldsymbol{x}_i+\boldsymbol{z}_i\middle|\mathbb{E}\left(\boldsymbol{h}_i\middle|\left\{\boldsymbol{h}_k+\delta\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right)\right)-\log\log\mathsf{SNR}\right]\\
&\quad=-1-\gamma-\log\mathrm{var}\left(\boldsymbol{h}_i\middle|\left\{\boldsymbol{h}_j+\delta\boldsymbol{z}_j\right\}_{j=-\infty}^{i-1}\right)
\end{aligned}$$

$$\begin{aligned}
\liminf_{\mathsf{SNR}\to\infty}\left[C(\mathsf{SNR})-\log\log\mathsf{SNR}\right]&\ge\liminf_{\mathsf{SNR}\to\infty}\left[\frac{1}{T}\sum_{i=1}^{T}I\left(\boldsymbol{x}_i;\boldsymbol{h}_i\boldsymbol{x}_i+\boldsymbol{z}_i\middle|\mathbb{E}\left(\boldsymbol{h}_i\middle|\left\{\boldsymbol{h}_k+\frac{1}{x_{\min}}\boldsymbol{z}_k\right\}_{k=-\infty}^{i-1}\right)\right)-\log\log\mathsf{SNR}\right]\\
&=-1-\gamma-\frac{1}{T}\sum_{i=1}^{T}\log\mathrm{var}\left(\boldsymbol{h}_i\middle|\left\{\boldsymbol{h}_j+\delta\boldsymbol{z}_j\right\}_{j=-\infty}^{i-1}\right).
\end{aligned}\tag{31}$$

From the upper bound (8), we have (33) at the bottom of the page, where (33) follows from the fact that

$$\left(\boldsymbol{x}_k, \mathbb{E}\left(\boldsymbol{h}_k \middle| \left\{\boldsymbol{h}_t + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{k-1}\right)\right) \to \left(\boldsymbol{x}_k, \mathbb{E}\left(\boldsymbol{h}_k \left| \{\boldsymbol{h}_j\}_{j=-\infty}^{k-1}\right.\right)\right) \to \boldsymbol{y}_k$$

form a Markov chain.

It was shown in [3, Corollary 4.19] that

$$\lim_{\mathsf{SNR}\to\infty}\left[\sup_{p_{\boldsymbol{x}_k}\in\mathcal{P}_1(\mathsf{SNR})} I(\boldsymbol{x}_1;\boldsymbol{y}_1) - \log\log\mathsf{SNR}\right] = -1-\gamma.$$

Furthermore

$$\begin{aligned} I\left(\mathbb{E}\left(\boldsymbol{h}_k \left| \{\boldsymbol{h}_j\}_{j=-\infty}^{k-1}\right.\right); \boldsymbol{y}_k \middle| \boldsymbol{x}_k\right) &\le I\left(\mathbb{E}\left(\boldsymbol{h}_k \left| \{\boldsymbol{h}_j\}_{j=-\infty}^{k-1}\right.\right); \boldsymbol{h}_k, \boldsymbol{y}_k \middle| \boldsymbol{x}_k\right) \\ &= I\left(\mathbb{E}(\boldsymbol{h}_k|\{\boldsymbol{h}_j\}_{j=-\infty}^{k-1}); \boldsymbol{h}_k\right) \\ &\quad + I\left(\mathbb{E}(\boldsymbol{h}_k|\{\boldsymbol{h}_j\}_{j=-\infty}^{k-1}); \boldsymbol{y}_k|\boldsymbol{h}_k, \boldsymbol{x}_k\right) \\ &= I\left(\mathbb{E}(\boldsymbol{h}_k|\{\boldsymbol{h}_j\}_{j=-\infty}^{k-1}); \boldsymbol{h}_k\right) \qquad (34) \\ &= -\log\sigma_k(\infty) \end{aligned}$$

where (34) follows from the fact that $\mathbb{E}\left(\boldsymbol{h}_k \left| \{\boldsymbol{h}_j\}_{j=-\infty}^{k-1}\right.\right) \to (\boldsymbol{h}_k, \boldsymbol{x}_k) \to \boldsymbol{y}_k$ form a Markov chain. Therefore, we get

$$\begin{aligned} &\limsup_{\mathsf{SNR}\to\infty}[C(\mathsf{SNR}) - \log\log\mathsf{SNR}] \\ &\le -1-\gamma-\frac{1}{T}\sum_{k=1}^{T}\log\sigma_k(0) \\ &= -1-\gamma-\frac{1}{T}\log\det\left[\Sigma(\infty)\right]. \qquad (35) \end{aligned}$$

The desired result follows by combining (32) and (35).

Appendix V
Proof of Proposition 1

At high SNR, we have the second equation at the bottom of the page, and thus

$$\begin{aligned} &\mathrm{var}\left(\boldsymbol{h}_1 \middle| \left\{\boldsymbol{h}_k + \frac{a}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_k\right\}_{k=-\infty}^{0}\right)\Bigg|_{s_\theta(e^{j\omega}),\theta=\mathsf{SNR}^r} \\ &= \frac{e^{\frac{c(r,a)}{2\pi}}}{\mathsf{SNR}^\kappa} - \frac{a^2}{\mathsf{SNR}} + o\left(\frac{1}{\mathsf{SNR}^\kappa}\right) \qquad (36) \end{aligned}$$

where $a > 0$, $\kappa = \frac{\alpha+\min(r,1)(2\pi-\alpha)}{2\pi}$, and

$$c(r,a) = \begin{cases} 2\alpha\log a, & r\in(0,1) \\ 2\alpha\log a + (2\pi-\alpha)\log(1+a^2), & r=1 \\ 4\pi\log a, & r\in(1,\infty). \end{cases}$$

In the lower bound (14), let $\boldsymbol{x}_1$ be uniformly distributed over the set $\{z\in\mathbb{C} : \frac{\sqrt{\mathsf{SNR}}}{2} \le |z| \le \sqrt{\mathsf{SNR}}\}$. By Lemma 2 and (36), we get the third equation at the bottom of the page.

By specializing the upper bound (8) to the case where $T = 1$ and then invoking (27), (28), (29), and (36) we obtain the equation at the top of the following page. The proof is complete.

$$\begin{aligned} &\limsup_{\mathsf{SNR}\to\infty}[C(\mathsf{SNR}) - \log\log\mathsf{SNR}] \\ &\le \limsup_{\mathsf{SNR}\to\infty}\left[\frac{1}{T}\sum_{k=1}^{T}\sup_{p_{\boldsymbol{x}_k}\in\mathcal{P}_1(\mathsf{SNR})} I\left(\boldsymbol{x}_k, \mathbb{E}\left(\boldsymbol{h}_k \middle| \left\{\boldsymbol{h}_t + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{k-1}\right); \boldsymbol{y}_k\right) - \log\log\mathsf{SNR}\right] \\ &\le \limsup_{\mathsf{SNR}\to\infty}\left[\frac{1}{T}\sum_{k=1}^{T}\sup_{p_{\boldsymbol{x}_k}\in\mathcal{P}_1(\mathsf{SNR})} I\left(\boldsymbol{x}_k, \mathbb{E}\left(\boldsymbol{h}_k \left| \{\boldsymbol{h}_j\}_{j=-\infty}^{k-1}\right.\right); \boldsymbol{y}_k\right) - \log\log\mathsf{SNR}\right] \qquad (33) \\ &\le \limsup_{\mathsf{SNR}\to\infty}\left[\sup_{p_{\boldsymbol{x}_k}\in\mathcal{P}_1(\mathsf{SNR})} I(\boldsymbol{x}_1;\boldsymbol{y}_1) + \frac{1}{T}\sum_{k=1}^{T}\sup_{p_{\boldsymbol{x}_k}\in\mathcal{P}_1(\mathsf{SNR})} I\left(\mathbb{E}\left(\boldsymbol{h}_k \left| \{\boldsymbol{h}_j\}_{j=-\infty}^{k-1}\right.\right); \boldsymbol{y}_k \middle| \boldsymbol{x}_k\right) - \log\log\mathsf{SNR}\right] \end{aligned}$$

$$\begin{aligned} &\int_{-\pi}^{\pi}\log\left[s_\theta(e^{j\omega}) + \frac{a^2}{\mathsf{SNR}}\right]\mathrm{d}\omega\Bigg|_{\theta=\mathsf{SNR}^r} \\ &= \alpha\log\left(\frac{a^2}{\mathsf{SNR}}\right) + \left(2\pi-\alpha-\mathsf{SNR}^{-r}\right)\log\left(\mathsf{SNR}^{-r} + \frac{a^2}{\mathsf{SNR}}\right) + \mathsf{SNR}^{-r}\log\left(\frac{2\pi\mathsf{SNR}^{2r} - 2\pi\mathsf{SNR}^r + \alpha\mathsf{SNR}^r + 1}{\mathsf{SNR}^r} + \frac{a^2}{\mathsf{SNR}}\right) \\ &= -2\pi\kappa\log\mathsf{SNR} + c(r,a) + o(1) \end{aligned}$$

$$\begin{aligned} C(\mathsf{SNR}) &\ge I\left(\boldsymbol{x}_1; \boldsymbol{h}_1\boldsymbol{x}_1 + \boldsymbol{z}_1 \middle| \mathbb{E}\left(\boldsymbol{h}_1 \middle| \left\{\boldsymbol{h}_k + \frac{2}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_k\right\}_{k=-\infty}^{0}\right)\right) \\ &\ge -\log\left[\frac{e^{\frac{c(r,2)}{2\pi}}}{\mathsf{SNR}^\kappa} - \frac{4}{\mathsf{SNR}} + \frac{8}{5\mathsf{SNR}} + o\left(\frac{1}{\mathsf{SNR}^\kappa}\right)\right] + \log\left(1 - \frac{e^{\frac{c(r,2)}{2\pi}}}{\mathsf{SNR}} + \frac{4}{\mathsf{SNR}} - o\left(\frac{1}{\mathsf{SNR}^\kappa}\right)\right) - \gamma - \log\frac{5e}{6} \\ &= \kappa\log\mathsf{SNR} + o(\log\mathsf{SNR}). \end{aligned}$$

$$C(\mathsf{SNR}) \le \log\left[\frac{1+\mathsf{SNR}}{1+\mathsf{SNR}\cdot \mathrm{var}\left(\boldsymbol{h}_1 \middle| \left\{\boldsymbol{h}_t + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{0}\right)}\right] + o(\log \mathsf{SNR})$$
$$\le -\log\left(\mathrm{var}\left(\boldsymbol{h}_1 \middle| \left\{\boldsymbol{h}_t + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{0}\right)\right) + o(\log \mathsf{SNR})$$
$$= \kappa \log \mathsf{SNR} + o(\log \mathsf{SNR}).$$

Appendix VI
Example 4

In the lower bound (14), let $\boldsymbol{x}_1$ be uniformly distributed over the set $\left\{z \in \mathbb{C} : \frac{\sqrt{\mathsf{SNR}}}{2} \le |z| \le \sqrt{\mathsf{SNR}}\right\}$. By Lemma 2, we get (37) at the bottom of the page, where

$$\mathrm{var}\left(\boldsymbol{h}_1 \middle| \left\{\boldsymbol{h}_k + \frac{2}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_k\right\}_{k=-\infty}^{0}\right) = \left(\epsilon_1 + \frac{4}{\mathsf{SNR}}\right)^{\alpha_1}\left(\epsilon_2 + \frac{4}{\mathsf{SNR}}\right)^{\alpha_2-\alpha_1} \times \left[\frac{1-\alpha_1\epsilon_1-(\alpha_2-\alpha_1)\epsilon_2}{1-\alpha_2} + \frac{4}{\mathsf{SNR}}\right]^{1-\alpha_2} - \frac{4}{\mathsf{SNR}}. \tag{38}$$

By specializing the upper bound (8) to the case where $T=1$ and then invoking (27), (28), and (29), we obtain (39), also at the bottom of the page, where

$$\mathrm{var}\left(\boldsymbol{h}_1 \middle| \left\{\boldsymbol{h}_t + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{0}\right) = \left(\epsilon_1 + \frac{1}{\mathsf{SNR}}\right)^{\alpha_1}\left(\epsilon_2 + \frac{1}{\mathsf{SNR}}\right)^{\alpha_2-\alpha_1} \times \left[\frac{1-\alpha_1\epsilon_1-(\alpha_2-\alpha_1)\epsilon_2}{1-\alpha_2} + \frac{1}{\mathsf{SNR}}\right]^{1-\alpha_2} - \frac{1}{\mathsf{SNR}}.$$

It follows from (37) and (39) that at high SNR, $C(\mathsf{SNR})$ can be lower-bounded by

$$-\log\left(\mathrm{var}\left(\boldsymbol{h}_1 \middle| \left\{\boldsymbol{h}_k + \frac{2}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_1\right\}_{k=-\infty}^{0}\right) + \frac{8}{5\mathsf{SNR}}\right)$$

minus a constant term, and upper-bounded by

$$-\log\left(\mathrm{var}\left(\boldsymbol{h}_1 \middle| \left\{\boldsymbol{h}_t + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{0}\right)\right)$$

plus a term that is negligible compared with $\log \mathsf{SNR}$.

In view of (38) and the fact that $\frac{1}{3} \le \alpha_1 < \alpha_2 \le \frac{2}{3}$, we have, for $1 \ll \log \mathsf{SNR} \le \frac{1}{\epsilon_2}$, (40), shown at the top of the following page, and for $\log\frac{1}{\epsilon_2} \ll \log \mathsf{SNR} \le \log\frac{1}{\epsilon_1}$

$$-\log\left(\mathrm{var}\left(\boldsymbol{h}_1 \middle| \left\{\boldsymbol{h}_k + \frac{2}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_k\right\}_{k=-\infty}^{0}\right) + \frac{8}{5\mathsf{SNR}}\right)$$
$$\ge -\log\left(\left(\frac{5}{\mathsf{SNR}}\right)^{\alpha_1}(2\epsilon_2)^{\alpha_2-\alpha_1}\left[\frac{1}{1-\alpha_2}+4\right]^{1-\alpha_2} - \frac{12}{5\mathsf{SNR}}\right)$$
$$\ge -\log\left(\left(\frac{5}{\mathsf{SNR}}\right)^{\alpha_1}(2\epsilon_2)^{\alpha_2-\alpha_1}7^{\frac{2}{3}} - \frac{12}{5\mathsf{SNR}}\right)$$
$$= \alpha_1 \log \mathsf{SNR} + o(\log \mathsf{SNR}). \tag{41}$$

Similarly, we have, for $1 \ll \log \mathsf{SNR} \le \frac{1}{\epsilon_2}$, (42), also at the top of the following page, and $\log\frac{1}{\epsilon_2} \ll \log \mathsf{SNR} \le \log\frac{1}{\epsilon_1}$, in (43) also at the top of the following page.

$$C(\mathsf{SNR}) \ge I\left(\boldsymbol{x}_1; \boldsymbol{h}_1\boldsymbol{x}_1 + \boldsymbol{z}_1 \middle| \mathbb{E}\left(\boldsymbol{h}_1 \middle| \left\{\boldsymbol{h}_k + \frac{2}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_k\right\}_{k=-\infty}^{0}\right)\right)$$
$$\ge -\log\left(\mathrm{var}\left(\boldsymbol{h}_1 \middle| \left\{\boldsymbol{h}_k + \frac{2}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_k\right\}_{k=-\infty}^{0}\right) + \frac{8}{5\mathsf{SNR}}\right)$$
$$+ \log\left(1 - \mathrm{var}\left(\boldsymbol{h}_1 \middle| \left\{\boldsymbol{h}_k + \frac{2}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_k\right\}_{k=-\infty}^{0}\right)\right) - \gamma - \log\frac{5e}{6} \tag{37}$$

$$C(\mathsf{SNR}) \le \log\left[\frac{1+\mathsf{SNR}}{1+\mathsf{SNR}\cdot \mathrm{var}\left(\boldsymbol{h}_1 \middle| \left\{\boldsymbol{h}_t + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{0}\right)}\right] + o(\log \mathsf{SNR})$$
$$\le -\log\left(\mathrm{var}\left(\boldsymbol{h}_1 \middle| \left\{\boldsymbol{h}_t + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{0}\right)\right) + o(\log \mathsf{SNR}) \tag{39}$$

$$
\begin{aligned}
&-\log\left(\operatorname{var}\left(\boldsymbol{h}_1 \middle| \left\{\boldsymbol{h}_k + \frac{2}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_k\right\}_{k=-\infty}^{0}\right) + \frac{8}{5\mathsf{SNR}}\right)\\
&\quad= -\log\left(\left(\epsilon_1 + \frac{4}{\mathsf{SNR}}\right)^{\alpha_1}\left(\epsilon_2 + \frac{4}{\mathsf{SNR}}\right)^{\alpha_2-\alpha_1}\left[\frac{1-\alpha_1\epsilon_1 - (\alpha_2-\alpha_1)\epsilon_2}{1-\alpha_2} + \frac{4}{\mathsf{SNR}}\right]^{1-\alpha_2} - \frac{12}{5\mathsf{SNR}}\right)\\
&\quad\geq -\log\left(\left(\frac{5}{\mathsf{SNR}}\right)^{\alpha_1}\left(\frac{5}{\mathsf{SNR}}\right)^{\alpha_2-\alpha_1}\left[\frac{1}{1-\alpha_2} + 4\right]^{1-\alpha_2} - \frac{12}{5\mathsf{SNR}}\right)\\
&\quad\geq -\log\left(\left(\frac{5}{\mathsf{SNR}}\right)^{\alpha_1}\left(\frac{5}{\mathsf{SNR}}\right)^{\alpha_2-\alpha_1} 7^{\frac{2}{3}} - \frac{12}{5\mathsf{SNR}}\right)\\
&\quad= \alpha_2 \log \mathsf{SNR} + o(\log \mathsf{SNR})
\end{aligned}
\tag{40}
$$

$$
\begin{aligned}
&-\log\left(\operatorname{var}\left(\boldsymbol{h}_1 \middle| \left\{\boldsymbol{h}_t + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{0}\right)\right)\\
&\quad= -\log\left(\left(\epsilon_1 + \frac{1}{\mathsf{SNR}}\right)^{\alpha_1}\left(\epsilon_2 + \frac{1}{\mathsf{SNR}}\right)^{\alpha_2-\alpha_1}\left[\frac{1-\alpha_1\epsilon_1 - (\alpha_2-\alpha_1)\epsilon_2}{1-\alpha_2} + \frac{1}{\mathsf{SNR}}\right]^{1-\alpha_2} - \frac{1}{\mathsf{SNR}}\right)\\
&\quad\leq -\log\left(\left(\frac{1}{\mathsf{SNR}}\right)^{\alpha_1}\left(\frac{1}{\mathsf{SNR}}\right)^{\alpha_2-\alpha_1}\left[\frac{1-\alpha_1 - (\alpha_2-\alpha_1)}{1-\alpha_2}\right]^{1-\alpha_2} - \frac{1}{\mathsf{SNR}}\right)\\
&\quad= -\log\left(\left(\frac{1}{\mathsf{SNR}}\right)^{\alpha_1}\left(\frac{1}{\mathsf{SNR}}\right)^{\alpha_2-\alpha_1} - \frac{1}{\mathsf{SNR}}\right)\\
&\quad= \alpha_2 \log \mathsf{SNR} + o(\log \mathsf{SNR})
\end{aligned}
\tag{42}
$$

$$
\begin{aligned}
&-\log\left(\operatorname{var}\left(\boldsymbol{h}_1 \middle| \left\{\boldsymbol{h}_t + \frac{1}{\sqrt{\mathsf{SNR}}}\boldsymbol{z}_t\right\}_{t=-\infty}^{0}\right)\right)\\
&\quad\leq -\log\left(\left(\frac{1}{\mathsf{SNR}}\right)^{\alpha_1}(\epsilon_2)^{\alpha_2-\alpha_1}\left[\frac{1-\alpha_1 - (\alpha_2-\alpha_1)}{1-\alpha_2}\right]^{1-\alpha_2} - \frac{1}{\mathsf{SNR}}\right)\\
&\quad\leq -\log\left(\left(\frac{1}{\mathsf{SNR}}\right)^{\alpha_1}(\epsilon_2)^{\alpha_2-\alpha_1} - \frac{1}{\mathsf{SNR}}\right)\\
&\quad= \alpha_1 \log \mathsf{SNR} + o(\log \mathsf{SNR}).
\end{aligned}
\tag{43}
$$

In view of (40), (41), (42), and (43), we can conclude that $\frac{C(\mathsf{SNR})}{\log \mathsf{SNR}}$ is approximately equal to α_2 for $1 \ll \log \mathsf{SNR} \leq \frac{1}{\epsilon_2}$, and is approximately equal to α_1 for $\log \frac{1}{\epsilon_2} \ll \log \mathsf{SNR} \leq \log \frac{1}{\epsilon_1}$.

Appendix VII
AWGN Channel

By the random coding bound [24], we have

$$P_e \leq e^{-nE_r(R)}$$

where $E_r(R)$ is defined in (44) at the bottom of the page, if

$$\log\left[\frac{1}{2} + \frac{\mathsf{SNR}}{4} + \frac{1}{2}\sqrt{1 + \frac{\mathsf{SNR}^2}{4}}\right] \leq R \leq \log(1 + \mathsf{SNR})$$

and

$$
\begin{aligned}
E_r(R) = {}& 1 + \frac{\mathsf{SNR}}{2} - \sqrt{1 + \frac{\mathsf{SNR}^2}{4}}\\
&+ \log\left(\frac{1}{2} - \frac{\mathsf{SNR}}{4} + \frac{1}{2}\sqrt{1 + \frac{\mathsf{SNR}^2}{4}}\right)
\end{aligned}
$$

$$
\begin{aligned}
E_r(R) = {}& \frac{\mathsf{SNR}}{2e^R}\left[e^R + 1 - (e^R - 1)\sqrt{1 + \frac{4e^R}{\mathsf{SNR}(e^R - 1)}}\right]\\
&+ \log\left\{e^R - \frac{\mathsf{SNR}(e^R - 1)}{2}\left[\sqrt{1 + \frac{4e^R}{\mathsf{SNR}(e^R - 1)}} - 1\right]\right\}
\end{aligned}
\tag{44}
$$

$$+\log\left(\frac{1}{2}+\frac{\mathsf{SNR}}{4}+\frac{1}{2}\sqrt{1+\frac{\mathsf{SNR}^2}{4}}\right)-R \quad (45)$$

if

$$R<\log\left[\frac{1}{2}+\frac{\mathsf{SNR}}{4}+\frac{1}{2}\sqrt{1+\frac{\mathsf{SNR}^2}{4}}\right].$$

Let $R(\mathsf{SNR})=\log\mathsf{SNR}-\log\eta$, where $\eta>1$. By (44), we get the first equation at the bottom of the page. For any $P_e>0$, we can find an n such that

$$e^{-n(\eta-1-\log\eta)}<P_e.$$

Therefore, for any $P_e>0$, there exist a sequence of codebooks with rate $R(\mathsf{SNR})$ and fixed codeword length n such that

$$\lim_{\mathsf{SNR}\to\infty}\frac{R(\mathsf{SNR})}{\log\mathsf{SNR}}=1$$

and $\limsup_{\mathsf{SNR}\to\infty}P_e(\mathsf{SNR})\le P_e$.

Appendix VIII
Coherent Rayleigh-Fading Channel

It was shown in [25] that

$$E_r(R)=\max_{0\le\rho\le1}\left[-\log\mathbb{E}_{\boldsymbol{h}}\left(1+\frac{\mathsf{SNR}}{1+\rho}|\boldsymbol{h}|^2\right)^{-\rho}-\rho R\right]$$

where $\boldsymbol{h}\sim\mathcal{CN}(0,1)$.

Choosing $R(\mathsf{SNR})=\log\mathsf{SNR}-\log\log\mathsf{SNR}-c$ and $\rho=1$, we get the second equation at the bottom of the page, which is positive if $c>\log 2$.

Therefore, for any $P_e>0$, we can find a sequence of codebooks with rate $R(\mathsf{SNR})$ and fixed codeword length n such that

$$\lim_{\mathsf{SNR}\to\infty}\frac{R(\mathsf{SNR})}{\log\mathsf{SNR}}=1$$

and $\limsup_{\mathsf{SNR}\to\infty}P_e(\mathsf{SNR})\le P_e$.

Appendix IX
Example 7

We can compute

$$\begin{aligned}\Psi(\{1\},\mathsf{SNR})&=\Psi(\{2\},\mathsf{SNR})=\Psi(\{3\},\mathsf{SNR})\\&=2\pi\log(1+\mathsf{SNR})\\\Psi(\{1,3\},\mathsf{SNR})&=\Psi(\{2,3\},\mathsf{SNR})\\&=\pi\log(1+2\mathsf{SNR}+\mathsf{SNR}^2-|\rho|^2\mathsf{SNR})\\\Psi(\{1,2\},\mathsf{SNR})&\\&=\pi\log(1+2\mathsf{SNR})\\\Psi(\{1,2,3\},\mathsf{SNR})&\\&=\frac{2\pi}{3}\log(1+3\mathsf{SNR}+2\mathsf{SNR}^2-2|\rho|^2\mathsf{SNR}^2).\end{aligned}$$

$$\begin{aligned}\lim_{\mathsf{SNR}\to\infty}E_r(R(\mathsf{SNR}))&=\lim_{\mathsf{SNR}\to\infty}\frac{\eta}{2}\left[\frac{\mathsf{SNR}}{\eta}+1-\left(\frac{\mathsf{SNR}}{\eta}-1\right)\sqrt{1+\frac{4}{\mathsf{SNR}-\eta}}\right]\\&\quad+\log\left\{\frac{\mathsf{SNR}}{\eta}-\frac{\mathsf{SNR}(\mathsf{SNR}-\eta)}{2\eta}\left[\sqrt{1+\frac{4}{\mathsf{SNR}-\eta}}-1\right]\right\}\\&=\lim_{\mathsf{SNR}\to\infty}\frac{\eta}{2}\left[\frac{\mathsf{SNR}}{\eta}+1-\frac{\mathsf{SNR}-\eta}{\eta}\left(1+\frac{2}{\mathsf{SNR}-\eta}\right)\right]\\&\quad+\log\left\{\frac{\mathsf{SNR}}{\eta}-\frac{\mathsf{SNR}(\mathsf{SNR}-\eta)}{2\eta}\left[\frac{2}{\mathsf{SNR}-\eta}-\frac{2}{(\mathsf{SNR}-\eta)^2}\right]\right\}\\&=\eta-1-\log\eta>0.\end{aligned}$$

$$\begin{aligned}\liminf_{\mathsf{SNR}\to\infty}E_r(R(\mathsf{SNR}))&\ge\liminf_{\mathsf{SNR}\to\infty}\left[-\log\mathbb{E}_{\boldsymbol{h}}\left(1+\frac{\mathsf{SNR}}{2}|\boldsymbol{h}|^2\right)^{-1}-\log\mathsf{SNR}+\log\log\mathsf{SNR}+c\right]\\&=\liminf_{\mathsf{SNR}\to\infty}\left[-\log\mathbb{E}_{\boldsymbol{h}}\left(\frac{1}{\mathsf{SNR}}+\frac{1}{2}|\boldsymbol{h}|^2\right)^{-1}+\log\log\mathsf{SNR}+c\right]\\&=\liminf_{\mathsf{SNR}\to\infty}\left\{-\log\left[\int_0^\infty\left(\frac{1}{\mathsf{SNR}}+\frac{t}{2}\right)^{-1}e^{-t}\mathrm{d}t\right]+\log\log\mathsf{SNR}+c\right\}\\&=\liminf_{\mathsf{SNR}\to\infty}\left\{-\log\left[\int_0^1\left(\frac{1}{\mathsf{SNR}}+\frac{t}{2}\right)^{-1}e^{-t}\mathrm{d}t+\int_1^\infty\left(\frac{1}{\mathsf{SNR}}+\frac{t}{2}\right)^{-1}e^{-t}\mathrm{d}t\right]+\log\log\mathsf{SNR}+c\right\}\\&\ge\liminf_{\mathsf{SNR}\to\infty}\left\{-\log\left[\int_0^1\left(\frac{1}{\mathsf{SNR}}+\frac{t}{2}\right)^{-1}\mathrm{d}t+\int_1^\infty2e^{-t}\mathrm{d}t\right]+\log\log\mathsf{SNR}+c\right\}\\&=\liminf_{\mathsf{SNR}\to\infty}\left[-\log\left[2\log(2+2\mathsf{SNR}+\mathsf{SNR}^2)-2\log(2+2\mathsf{SNR})+2e^{-1}\right]+\log\log\mathsf{SNR}+c\right]\\&=-\log2+c\end{aligned}$$

It can be verified that

$$\Psi(\{1\},\mathsf{SNR}) = \Psi(\{2\},\mathsf{SNR}) = \Psi(\{3\},\mathsf{SNR}) \geq \Psi(\{1,3\},\mathsf{SNR}) = \Psi(\{2,3\},\mathsf{SNR}) \geq \Psi(\{1,2\},\mathsf{SNR}).$$

So the optimal $\mathcal{M}^*$ is either $\{1,2\}$ or $\{1,2,3\}$. Setting $\Psi(\{1,2\},\mathsf{SNR}) = \Psi(\{1,2,3\},\mathsf{SNR})$ yields

$$(1+3\mathsf{SNR}+2\mathsf{SNR}^2-2|\rho|^2\mathsf{SNR}^2)^2 = (1+2\mathsf{SNR})^3$$

which, after some algebraic manipulation, is equivalent to

$$1-4|\rho|^2+(4-12|\rho|^2)\mathsf{SNR}+(4-8|\rho|^2+4|\rho|^4)\mathsf{SNR}^2=0.$$

The above equation has two solutions

$$\mathsf{SNR}_1 = \frac{-2|\rho|-1}{2(1+|\rho|)^2}, \quad \mathsf{SNR}_2 = \frac{2|\rho|-1}{2(1-|\rho|)^2}.$$

SNR_1 can be discarded since it is always negative. SNR_2 is positive for $|\rho| \in \left(\frac{1}{2},1\right)$. When $|\rho| \in \left(\frac{1}{2},1\right)$, it can be verified that $\Psi(\{1,2\},\mathsf{SNR}) > \Psi(\{1,2,3\},\mathsf{SNR})$ if $\mathsf{SNR} < \mathsf{SNR}_2$, and $\Psi(\{1,2\},\mathsf{SNR}) < \Psi(\{1,2,3\},\mathsf{SNR})$ if $\mathsf{SNR} > \mathsf{SNR}_2$. When $|\rho| \in \left[0,\frac{1}{2}\right]$, SNR_2 is nonpositive. In this case, we have $\Psi(\{1,2\},\mathsf{SNR}) > \Psi(\{1,2,3\},\mathsf{SNR})$ for all $\mathsf{SNR} > 0$.

Acknowledgment

The authors wish to thank the Associate Editor Dr. Gerhard Kramer and the two anonymous reviewers for their valuable comments and suggestions.

References

[1] T. Marzetta and B. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.
[2] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.
[3] A. Lapidoth and S. Moser, "Capacity bounds via duality with applications to multi-antenna systems on flat fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.
[4] A. Lapidoth and S. Moser, "The fading number of single-input multiple-output fading channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 437–453, Feb. 2006.
[5] T. Koch and A. Lapidoth, "The fading number and degrees of freedom in non-coherent MIMO fading channels: A peace pipe," in *Proc. IEEE Int. Symp. Information Theory* , Adelaide, Australia, Sep. 2005, pp. 661–665.
[6] A. Lapidoth, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 437–446, Feb. 2005.
[7] T. Koch and A. Lapidoth, "Gaussian fading is the worst fading," in *Proc. IEEE Int. Symp. Information Theory*, Seattle, WA, Jul. 2006, pp. 26–29.
[8] Y. Liang and V. V. Veeravalli, "Capacity of noncoherent time-selective Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3095–3110, Dec. 2004.
[9] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, Sep. 1990.
[10] A. R. Yu, *Stationary Random Processes*. San Francisco, CA: Holden-Day, 1967.
[11] R. M. Gray, *Entropy and Information Theory*. New York: Springer-Verlag, 1990.
[12] R. L. Dobrushin, "General formulation of Shannon's main theorem in information theory," *Amer. Math. Sot. Trans.*, vol. 33, pp. 323–438, 1963.
[13] R. Etkin and D. N. C. Tse, "Degrees of freedom in some underspread MIMO fading channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1576–1608, Apr. 2006.
[14] A. Lapidoth, "On the high SNR capacity of stationary Gaussian fading channels," in *Proc. 41st Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2003.
[15] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.
[16] G. M. Goluzin, *The Geometrical Theory of Functions of a Complex Variable*. Moscow, U.S.S.R.: Nauka, 1966.
[17] N. M. Babayan, "On asymptotic behavior of the prediction error in the singular case," *Theory of Probability and Its Applications*, vol. 29, no. 1, pp. 147–150, 1985.
[18] V. Sethuraman and B. Hajek, "Capacity per unit energy of fading channels with a peak constraint," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3102–3120, Sep. 2005.
[19] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[20] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
[21] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 1319–1343, Feb. 2002.
[22] L. Zheng, D. N. C. Tse, and M. Médard, "Channel coherence in the low-SNR regime," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 976–997, Mar. 2007.
[23] S. Ray, M. Médard, and L. Zheng, "On noncoherent MIMO channels in the wideband regime: Capacity and reliability," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 1983–2009, Jun. 2007.
[24] R. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[25] İ.E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, 1999.