On the Asymptotic Capacity of Stationary Gaussian Fading Channels

Amos Lapidoth, *Fellow, IEEE*

***Abstract*—We consider a peak-power-limited single-antenna flat complex-Gaussian fading channel where the receiver and transmitter, while fully cognizant of the distribution of the fading process, have no knowledge of its realization. Upper and lower bounds on channel capacity are derived, with special emphasis on tightness in the high signal-to-noise ratio (SNR) regime. Necessary and sufficient conditions (in terms of the autocorrelation of the fading process) are derived for capacity to grow double-logarithmically in the SNR. For cases in which capacity increases logarithmically in the SNR, we provide an expression for the "pre-log," i.e., for the asymptotic ratio between channel capacity and the logarithm of the SNR. This ratio is given by the Lebesgue measure of the set of harmonics where the spectral density of the fading process is zero. We finally demonstrate that the asymptotic dependence of channel capacity on the SNR need not be limited to logarithmic or double-logarithmic behaviors. We exhibit power spectra for which capacity grows as a fractional power of the logarithm of the SNR.**

***Index Terms*—Asymptotic expansion, channel capacity, fading channels, high signal-to-noise ratio (SNR), multiplexing gain, noncoherent, Rayleigh, Rice, time-selective.**

I. Introduction

IN this paper, we study the capacity of a single-antenna discrete-time flat-fading channel. We assume that the fading process is a stationary circularly symmetric complex-Gaussian process whose law (i.e., mean and autocorrelation function)—but not realization—is known to the transmitter and receiver. Some authors refer to models, such as ours, where the realization of the fading is unknown to the receiver and transmitter as "noncoherent" models. Our channel model includes as special cases the Rayleigh and Ricean channel models that correspond to zero-mean (Rayleigh) and non-zero-mean (Rice) independent and identically distributed (i.i.d.) fading. Our emphasis here will, however, be on the case where the fading process has memory (is not i.i.d.) and thus introduces memory into the channel model. The fading is thus "time-selective." This memory can be exploited by the system designer to allow for the receiver to track the fading level and to thus achieve higher communication rates. While we do not preclude the possibility of the use of training sequences to learn the channel, we view this possibility as a special case of coding. Thus, the capacity of this channel is the ultimate limit on the rate of reliable communication on this channel irrespective of the type of coding employed, be it via training sequences or not.

Even in the absence of memory, this channel model does not lead to explicit expressions for channel capacity, and it is thus not surprising that previous analyses of this model were mostly based on a further simplification of the model. A commonly used simplification is the block-constant fading model [1]. In this model, the fading is no longer assumed stationary.[1] Instead, it is assumed that it is drawn independently every T symbols and then remains constant for the duration of T symbols. The capacity of this simplified model was studied in [2] in the high-signal-to-noise ratio (SNR) regime, where capacity was shown to increase logarithmically with the SNR, with the "pre-log"[2] being given (for $T \geq 2$) by $(T-1)/T$. A different simplified model—one that generalizes the block-constant model—was recently proposed in [3]. Here the fading is still nonstationary but it has a more intricate structure. The fading is i.i.d. in blocks of size T, but within the block the fading need not be constant; an arbitrary covariance structure is allowed. The high-SNR analysis shows that unless the covariance matrix is of full rank, capacity grows logarithmically in the SNR with the pre-log determined by the rank Q of the covariance matrix. For $Q < T$ the pre-log is $(T-Q)/T$.

To the best of our knowledge, the only study that addresses our model without any simplifications is by Lapidoth and Moser [4]. There, it was shown that if the fading process is regular in the sense that its "present" cannot be predicted precisely from its "past," then capacity grows double-logarithmically in the SNR.[3] This was perhaps the first indication that the high-SNR behavior of channel capacity can depend critically on the model, and that simplifications of the model may lead to completely different asymptotic behaviors.

[1] The fading in the Marzetta–Hochwald model [1] is cyclostationary. By introducing a random time shift that is uniformly distributed over the duration of the block one can stationarize the process, but the resulting process is no longer Gaussian.

[2] By the "pre-log" we refer to the limiting ratio of channel capacity to the logarithm of the SNR. Some authors refer to this as "multiplexing gain," but this latter expression seems more appropriate for multiple-antenna systems where it can be greater than one.

[3] A special case of regular Gaussian fading is the first-order Gauss–Markov fading model, which has been recently studied in [5] and [6]. Chen *et al.*[5] focused on the mutual information that is achievable with fixed (SNR-independent) input distributions, whereas Etkin and Tse [6] focused on the asymptotics of channel capacity (input distribution allowed to depend on the SNR). Note that the asymptotics studied by Etkin and Tse are different from the ones that are of interest to us here. Whereas we fix the channel and study the limiting behavior of channel capacity as the SNR tends to infinity, Etkin and Tse study a *double limit*: the SNR tending to infinity and the variance of the fading innovations tending to zero. Thus, their asymptotics correspond to the limiting behavior as the SNR tends to infinity and the fading becomes more and more deterministic.

Manuscript received May 23, 2003; revised September 23, 2004. The material in this paper was presented in part at the 41st Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, October 2003.

The author is with the Swiss Federal Institute of Technology (ETH), CH-8092 Zurich, Switzerland (e-mail: lapidoth@isi.ee.ethz.ch).

Communicated by İ. E. Telatar, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2004.840889

In order to better understand channel capacity at high SNR and in an attempt to bridge the gap between the double-logarithmic and the logarithmic behaviors discussed above, we extend here the study of [4] to the case where the fading is not regular, i.e., when the present fading can be determined precisely from the past values of the fading. We shall derive upper (33) and lower (47) bounds on the capacity of this channel with a view to an understanding of channel capacity at high SNR (48), (53). With the aid of these bound we shall obtain the following.

- A characterization (in terms of the power spectral density of the fading process) of the fading processes that lead to a double-logarithmic dependence of channel capacity on the SNR. See Section VII (55).
- An expression for the pre-log when capacity grows logarithmically in the SNR. See Section VIII (61).
- Examples of fading processes that lead to other asymptotic behaviors, e.g., processes for which capacity grows like a fractional power of the logarithm of the SNR. See Section IX (74).

It should be emphasized that this paper deals only with single-antenna communications. This is not to imply that extensions to the multiple-antenna scenario are straightforward. On the contrary, the multiple-antenna scenario seems to be much more complicated. Some recent advances in the analysis of multiple-antenna fading channels with memory include an exact computation of the fading number for regular single-input multiple-output (SIMO) systems [7]; and the extension of the calculation of the present paper to the pre-log for some Gaussian multiple-input single-output (MISO) fading channels [8].[4]

The rest of this paper is organized as follows. In Section II, we describe the channel model and define its capacity. In Section III, we discuss the classical prediction problem and the noisy prediction problem for stationary circularly symmetric Gaussian processes. In Section IV, we propose upper bounds on channel capacity, and in Section V, lower bounds. An asymptotic analysis of these bounds is performed in Section VI. This analysis is used in Section VII to derive necessary and sufficient conditions for capacity to grow double-logarithmically in the SNR. The study of the pre-log is the subject of Section VIII and asymptotic behaviors other than logarithmic or double-logarithmic are presented in Section IX. Section X discusses the application of the new bounds to the finer analysis of regular fading channels. The paper concludes with a brief summary and some conclusions in Section XI.

II. Channel Model

We consider a discrete-time channel whose time-k complex-valued output $Y_k \in \mathbb{C}$ is given by

$$Y_k = (d + H_k)x_k + Z_k \tag{1}$$

where $x_k \in \mathbb{C}$ is the complex-valued channel input at time k; the constant $d \in \mathbb{C}$ is a deterministic complex number; the complex process $\{H_k\}$ models multiplicative noise; and the complex process $\{Z_k\}$ models additive noise. The processes $\{H_k\}$ and $\{Z_k\}$ are assumed to be independent and of a joint law that does not depend on the input sequence $\{x_k\}$.

We shall assume that the sequence $\{Z_k\}$ is a sequence of i.i.d. circularly symmetric complex-Gaussian random variables of zero mean and variance σ^2. Thus, $Z_k \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2)$ where we use the notation $W \sim \mathcal{N}_{\mathbb{C}}(\mu, \sigma^2)$ to indicate that $W - \mu$ has a zero-mean variance-σ^2 circularly symmetric complex-Gaussian distribution, i.e., to indicate that the density $f_W(w)$ of W is given by

$$f_W(w) = \frac{1}{\pi\sigma^2} e^{-\frac{|w-\mu|^2}{\sigma^2}}, \qquad w \in \mathbb{C}. \tag{2}$$

As to the "fading process" $\{H_k\}$, we shall assume that it is a zero-mean, unit-variance, stationary, circularly symmetric, Gaussian process of arbitrary spectral distribution function $F(\lambda)$, $-1/2 \le \lambda \le 1/2$. Thus, $F(\cdot)$ is a monotonically nondecreasing function on $[-1/2, 1/2]$ [9, Theorem 3.2, p. 474]

$$\mathsf{E}\left[H_{k+m}H_k^*\right] = \int_{-1/2}^{1/2} e^{i2\pi m\lambda} \mathrm{d}F(\lambda), \qquad k, m \in \mathbb{Z} \tag{3}$$

and

$$\mathsf{E}\left[|H_k|^2\right] = 1. \tag{4}$$

Notice that we do not assume that $F(\cdot)$ is absolutely continuous with respect to the Lebesgue measure on $[-1/2, 1/2]$, i.e., we do not assume that the process $\{H_k\}$ has a spectral density. Since $F(\lambda)$ is monotonic, it is almost everywhere differentiable, and we denote its derivate by $F'(\lambda)$. (At the discontinuity points of F, the derivative F' is undefined. We do not use Dirac's delta functions in this paper.)

Unless we restrict the channel inputs, the capacity of this channel is typically infinite. Usually one considers channel capacity under an energy constraint on the input but, to treat the problem analytically, we have chosen in this paper to consider the peak-power constraint

$$|x_k| \le A. \tag{5}$$

We define the SNR by

$$\text{SNR} = \frac{A^2}{\sigma^2}. \tag{6}$$

The subject of our investigation is the capacity $C(\text{SNR})$, which is defined by

$$C(\text{SNR}) = \lim_{n\to\infty} \frac{1}{n} \sup I(X_1, \dots, X_n; Y_1, \dots, Y_n) \tag{7}$$

where the supremum is over all joint distributions on $X_1, \dots, X_n$ satisfying the peak constraint (5), and where the limit exists because $\{H_k\}$ was assumed stationary.

It should be noted that $C(\text{SNR})$ need not have a coding theorem associated with it. A coding theorem will, however, hold if $\{H_k\}$ is *ergodic*, as is, for example, the case if $F(\cdot)$ is absolutely continuous, i.e., if $\{H_k\}$ has a *spectral density*.

[4]For *nonregular* fading, Koch [8] solves for the pre-log of peak-limited Gaussian MISO channels when the Gaussian fading processes from the different transmitters to the receiver are independent.

For *regular* fading, [8] computes the fading number for Gaussian MISO channels when the channels from the different transmitters to the receiver are independent and are additionally either all of zero mean (and arbitrary spectrum) or all of identical power spectra (and possibly different means.)

III. Noiseless and Noisy Prediction

As we shall see, the large-SNR behavior of $C(\mathrm{SNR})$ depends critically on the mean-squared error $\epsilon^2_{\mathrm{pred}}$ in predicting H_0 from past values $H_{-1}, H_{-2}, \ldots$ [9, Theorem 4.3]

$$\epsilon^2_{\mathrm{pred}} = \exp\left\{\int_{-1/2}^{1/2} \log F'(\lambda) \mathrm{d}\lambda\right\}. \tag{8}$$

If $\epsilon^2_{\mathrm{pred}} > 0$ then [4]

$$C(\mathrm{SNR}) = \log\log \mathrm{SNR} + \log |d|^2 - \mathrm{Ei}(-|d|^2) - 1 + \log \frac{1}{\epsilon^2_{\mathrm{pred}}} + o(1) \tag{9}$$

where $\mathrm{Ei}(\cdot)$ denotes the exponential integral function

$$\mathrm{Ei}(-x) = -\int_x^\infty \frac{e^{-t}}{t} \mathrm{d}t, \qquad x > 0 \tag{10}$$

and where the $o(1)$ term tends to zero as $\mathrm{SNR} \to \infty$.

We follow Doob [9, Sec. XII.2, p. 564] and refer to processes for which $\epsilon^2_{\mathrm{pred}} > 0$ as *regular* and to those for which $\epsilon^2_{\mathrm{pred}} = 0$ as *nonregular* or *deterministic*. Note, however, that Ibragimov and Rozanov [10] require that regular processes also have an absolutely continuous spectral distribution, i.e., possess a spectral density.

With (9) established, we shall focus in this paper on the case where $\epsilon^2_{\mathrm{pred}} = 0$. For the asymptotic analysis of this case we shall find it important to analyze the noisy prediction problem for $\{H_k\}$. This problem can be stated as follows. Let $\{W_k\}$ be a sequence of i.i.d. $\mathcal{N}_{\mathbb{C}}(0, \delta^2)$ random variables. The noisy prediction problem is to predict H_0 based on the observations $H_{-1} + W_{-1}, H_{-2} + W_{-2}, \ldots$. We denote the mean-squared error associated with the optimal predictor by $\epsilon^2_{\mathrm{pred}}(\delta^2)$ and note that it is given by

$$\epsilon^2_{\mathrm{pred}}(\delta^2) = \exp\left\{\int_{-1/2}^{1/2} \log\left(F'(\lambda) + \delta^2\right) \mathrm{d}\lambda\right\} - \delta^2. \tag{11}$$

Indeed, the conditional expectation of H_0 given the observations $H_{-1} + W_{-1}, H_{-2} + W_{-2}, \ldots$ is the same as the conditional expectation of $H_0 + W_0$ given those observations. Since W_0 is independent of the observations, $\epsilon^2_{\mathrm{pred}}(\delta^2)$ can be thus written as the prediction error for the process $\{H_k + W_k\}$ but with the variance of W_0 subtracted.

Note that in view of our normalization (4), the fact that H_0 is Gaussian, and the fact that H_0 is also conditionally Gaussian given the noisy past $H_{-1} + W_{-1}, H_{-2} + W_{-2}, \ldots$ it follows that

$$I\left(\{H_\nu + W_\nu\}_{\nu=-\infty}^{-1}; H_0\right) = \log \frac{1}{\epsilon^2_{\mathrm{pred}}(\delta^2)}. \tag{12}$$

We next recall some facts related to the prediction problem for circularly symmetric stationary Gaussian processes. To simplify the exposition, we shall somewhat abuse convention and refer to $\mathcal{N}_{\mathbb{C}}(\mu, \sigma^2)$ complex random variables as circularly symmetric Gaussian even for $\mu \neq 0$. Also, we shall use the notation A_m^n to refer to the random variables $A_m, \ldots, A_n$.

We first note that if a process $\{A_k\}$ is a circularly symmetric Gaussian process, then the conditional distribution of A_0 conditional on $A_{-1}, A_{-2}, \ldots, A_{-n}$ is a Gaussian with a deterministic variance. That is, if

$$\hat{A}_0^{(n)} = \mathsf{E}\left[A_0 | A_{-n}^{-1}\right]$$

then

$$\mathsf{E}\left[\left|A_0 - \hat{A}_0^{(n)}\right|^2 \middle| A_{-n}^{-1}\right] = \mathsf{E}\left[\left|A_0 - \hat{A}_0^{(n)}\right|^2\right], \text{ almost surely.}$$

Moreover, $\hat{A}_0^{(n)}$ has a Gaussian (unconditioned) distribution.

Finally, if $\{A_k\}$ is stationary, then the prediction error is monotonically nonincreasing in n and

$$\lim_{n\to\infty} \mathsf{E}\left[\left|A_0 - \hat{A}_0^{(n)}\right|^2\right] = \mathsf{E}\left[\left|A_0 - \mathsf{E}\left[A_0 | A_{-\infty}^{-1}\right]\right|^2\right]. \tag{13}$$

(For the latter claim, see [9, p. 562], [9, Sec. IV, Theorem 7.4], [9, Sec. VII, Theorem 4.3].)

IV. An Upper Bound

To upper-bound $I(X^n; Y^n)$ we use the chain rule

$$I(X^n; Y^n) = \sum_{k=1}^n I(X^n; Y_k | Y^{k-1}) \tag{14}$$

and upper-bound each of the individual terms in the sum by

$$\begin{aligned} I(X^n; Y_k | Y^{k-1}) &= I(X^n, Y^{k-1}; Y_k) - I(Y_k; Y^{k-1}) \\ &\leq I(X^n, Y^{k-1}; Y_k) \\ &= I(X^k, Y^{k-1}; Y_k) \\ &= h(Y_k) - h(Y_k | X_k, X^{k-1}, Y^{k-1}) \end{aligned} \tag{15}$$

where the first equality follows from the chain rule; the subsequent inequality from the nonnegativity of mutual information; the following equality from the absence of feedback, which results in the future inputs X_{k+1}^n being independent of the present output Y_k given the past and present inputs X^k and the past outputs Y^{k-1}; and the last equality from the expansion of mutual information in terms of differential entropies.

We now consider the maximization of the right-hand side (RHS) of (15) over all joint distributions on X^k satisfying the peak constraint

$$|X_\nu| \leq A, \qquad \nu = 1, \ldots, k, \text{ almost surely.} \tag{16}$$

This maximization can be written as a double maximization over the distribution p_{X_k} of X_k and the conditional law $p_{X^{k-1}|X_k}$ of its past

$$\begin{aligned} &\sup_{p_{X^k}} \left\{h(Y_k) - h(Y_k | X_k, X^{k-1}, Y^{k-1})\right\} \\ &= \sup_{p_{X_k}} \sup_{p_{X^{k-1}|X_k}} \left\{h(Y_k) - h(Y_k | X_k, X^{k-1}, Y^{k-1})\right\} \\ &= \sup_{p_{X_k}} \left\{h(Y_k) - \inf_{p_{X^{k-1}|X_k}} h(Y_k | X_k, X^{k-1}, Y^{k-1})\right\} \end{aligned} \tag{17}$$

where the second equality follows from the observation that specifying the law of X_k also specifies the law of $Y_k = H_k X_k + Z_k$ (because the laws of H_k and Z_k are fixed) and hence also determines its differential entropy $h(Y_k)$.

We next note that

$$\inf_{p_{X^{k-1}|X_k}} h(Y_k|X_k, X^{k-1}, Y^{k-1}) = h(Y_k|X_k, H_{k-1}+W_{k-1}, \ldots, H_1+W_1) \quad (18)$$

where $W_1, \ldots W_{k-1}$ are i.i.d. $\mathcal{N}_{\mathbb{C}}(0, \sigma^2/A^2)$, and where the infimum is achieved by any conditional law $p_{X^{k-1}|X_k}$ under which $X_1, \ldots, X_{k-1}$ are almost surely of magnitude A. This follows because once the value of X_k has been fixed, the variables $X_1, Y_1, \ldots, X_{k-1}, Y_{k-1}$ influence the conditional differential entropy of Y_k only through the information they convey on $H_1, \ldots, H_{k-1}$ and hence on H_k. These variables convey information about $H_1, \ldots, H_{k-1}$ through the ratios $Y_1/X_1, \ldots, Y_{k-1}/X_{k-1}$, and this information is maximized when the inputs are of maximum magnitude A, in which case

$$\frac{Y_\nu}{X_\nu} - d = \frac{H_\nu X_\nu + Z_\nu}{X_\nu} \quad (19)$$

$$\sim H_\nu + W_\nu. \quad (20)$$

Combining (18) with (17) and (15) we obtain

$$\begin{aligned}
&\sup_{p_{X^n}} I(X^n; Y_k|Y^{k-1}) \\
&\le \sup_{p_{X^k}} \left\{h(Y_k) - h(Y_k|X_k, X^{k-1}, Y^{k-1})\right\} && (21)\\
&= \sup_{p_{X_k}} \left\{h(Y_k) - \inf_{p_{X^{k-1}|X_k}} h(Y_k|X_k, X^{k-1}, Y^{k-1})\right\} && (22)\\
&= \sup_{p_{X_k}} \{h(Y_k) \\
&\qquad - h(Y_k|X_k, H_{k-1}+W_{k-1}, \ldots, H_1+W_1)\} && (23)\\
&= \sup_{p_{X_k}} I(X_k, H_{k-1}+W_{k-1}, \ldots, H_1+W_1; Y_k) && (24)\\
&= \sup_{p_{X_0}} I\left(X_0, \{H_\nu+W_\nu\}_{\nu=-(k-1)}^{-1}; Y_0\right) && (25)\\
&\le \sup_{p_{X_0}} I\left(X_0, \{H_\nu+W_\nu\}_{\nu=-\infty}^{-1}; Y_0\right) && (26)\\
&= \sup_{p_{X_0}} \left\{I(X_0;Y_0) + I\left(\{H_\nu+W_\nu\}_{\nu=-\infty}^{-1}; Y_0|X_0\right)\right\} && (27)\\
&\le \sup_{p_{X_0}} I(X_0;Y_0) + \sup_{p_{X_0}} I\left(\{H_\nu+W_\nu\}_{\nu=-\infty}^{-1}; Y_0|X_0\right) && (28)\\
&\le \sup_{p_{X_0}} I(X_0;Y_0) + I\left(\{H_\nu+W_\nu\}_{\nu=-\infty}^{-1}; H_0\right) && (29)\\
&= \sup_{p_{X_0}} I(X_0;Y_0) + \log \frac{1}{\epsilon_{\text{pred}}^2(\sigma^2/A^2)} && (30)\\
&= \log\log\frac{A^2}{\sigma^2} + \log|d|^2 - \text{Ei}(-|d|^2) - 1 \\
&\qquad + \log\frac{1}{\epsilon_{\text{pred}}^2(\sigma^2/A^2)} + o(1) && (31)\\
&= \log\log \text{SNR} + \log|d|^2 - \text{Ei}(-|d|^2) - 1 \\
&\qquad + \log\frac{1}{\epsilon_{\text{pred}}^2(1/\text{SNR})} + o(1) && (32)
\end{aligned}$$

where the $o(1)$ term depends on d and A^2/σ^2 and for any fixed d, converges to zero as $A^2/\sigma^2 \to \infty$. Here (21) follows from (15); (22) follows from (17); (23) follows from (18); (24) follows from the definition of mutual information; (25) follows from the stationarity of the fading process; (26) follows by revealing the distant past; (27) follows by the chain rule; (28) by maximizing each of the two terms individually; (29) follows from the data-processing inequality, which applies because conditional on (X_0, H_0), the random variable Y_0 is independent of $\{H_\nu + W_\nu\}_{\nu=-\infty}^{-1}$; (30) follows from (12); (31) follows from the analysis of the high-SNR behavior of the capacity of the peak-limited *memoryless* Ricean fading channel [4, Corollary 4.19]; and (32) follows from the definition the SNR (6).

Notice that (32) could be turned into a nonasymptotic bound by upper bounding the term $\sup_{p_{X_0}} I(X_0;Y_0)$ with an appropriate nonasymptotic upper bound on the capacity of the peak-limited memoryless Ricean fading channel, e.g., [4, Sec. IV.F.3, eqs. (166) and (174)]. We shall not pursue this here because our focus is on the high-SNR regime.

Combining (32) with (14) and (7) we obtain the upper bound

$$C(\text{SNR}) \le \log\frac{1}{\epsilon_{\text{pred}}^2(1/\text{SNR})} + \log\log\text{SNR} + \log|d|^2 - \text{Ei}(-|d|^2) - 1 + o(1) \quad (33)$$

where the $o(1)$ is as above. Note that this $o(1)$ term can be upper-bounded firmly as in the analysis of Ricean fading [4].

V. A Lower Bound

To derive a lower bound on channel capacity we shall consider inputs $\{X_k\}$ that are i.i.d. and uniformly distributed over the set $\{z \in \mathbb{C} : A/2 \le |z| \le A\}$. Using the chain rule

$$\frac{1}{n} I(X^n; Y^n) = \frac{1}{n}\sum_{k=1}^{n} I\left(X_k; Y^n|X^{k-1}\right) \quad (34)$$

and a Cesáro-type theorem [11, Theorem 4.2.3] we obtain after discarding future outputs that the capacity C can be lower-bounded by

$$C \ge \varliminf_{k\to\infty} I\left(X_k; Y^k|X^{k-1}\right). \quad (35)$$

We now proceed to lower-bound the term on the RHS of (35) using the fact that we have chosen $\{X_k\}$ to be i.i.d. and satisfying $|X_k| \ge A/2$, almost surely

$$\begin{aligned}
&I\left(X_k; Y^k|X^{k-1}\right) \\
&= I\left(X_k; X^{k-1}, Y^{k-1}, Y_k\right) \\
&= I\left(X_k; Y_k, \left\{\frac{Y_\nu}{X_\nu} - d\right\}_{\nu=1}^{k-1}, X^{k-1}\right) \\
&= I\left(X_k; Y_k, \left\{\frac{Y_\nu}{X_\nu} - d\right\}_{\nu=1}^{k-1} \middle| X^{k-1}\right) \\
&= I\left(X_k; Y_k, \left\{H_\nu + \frac{Z_\nu}{X_\nu}\right\}_{\nu=1}^{k-1} \middle| X^{k-1}\right) \\
&\ge I\left(X_k; Y_k, \{H_\nu + W'_\nu\}_{\nu=1}^{k-1}\right) \\
&= I\left(X_k; Y_k| \{H_\nu + W'_\nu\}_{\nu=1}^{k-1}\right) && (36)
\end{aligned}$$

where

$$\{W'_\nu\} \sim \text{i.i.d.} \quad \mathcal{N}_\mathbb{C}\left(0, \frac{\sigma^2}{(A/2)^2}\right). \tag{37}$$

Notice that it was only in the last inequality that we used the fact that under the input distribution that we have chosen all inputs are of magnitude no smaller than $A/2$.

Expressing the present fading H_k as

$$H_k = \hat{D} + \tilde{D} \tag{38}$$

where

$$\hat{D} = \mathsf{E}\left[H_k | \{H_\nu + W'_\nu\}_{\nu=1}^{k-1}\right] \tag{39}$$

we obtain from (36) that

$$I\left(X_k; Y^k | X^{k-1}\right) \geq I\left(X; (d + \hat{D} + \tilde{D})X + Z \middle| \hat{D}\right) \tag{40}$$

where X, Z, $\tilde{D}$, $\hat{D}$, are independent random variables of the following laws: X is uniformly distributed over the set $\{z \in \mathbb{C} : A/2 \leq |z| \leq A\}$; the additive noise Z is $\mathcal{N}_\mathbb{C}(0, \sigma^2)$ distributed; the prediction error $\tilde{D}$ in predicting H_k from $\{H_\nu + W'_\nu\}_{\nu=1}^{k-1}$ is $\mathcal{N}_\mathbb{C}(0, \tilde{\epsilon}_k^2)$ distributed where $\tilde{\epsilon}_k^2$ is the mean-squared prediction error; and $\hat{D} \sim \mathcal{N}_\mathbb{C}(0, 1 - \tilde{\epsilon}_k^2)$. Notice that by (13)

$$\lim_{k\to\infty} \tilde{\epsilon}_k^2 = \epsilon_{\text{pred}}^2(\delta^2)|_{\delta^2 = \frac{\sigma^2}{(A/2)^2}}. \tag{41}$$

To lower-bound the RHS of (40) we derive in Appendix I the lower bound

$$\begin{aligned} I\left(X; (\hat{d} + \tilde{D})X + Z\right) &\geq \log\left(1 + \frac{\mathcal{E}_\text{s}|\hat{d}|^2}{\mathcal{E}_\text{s}\tilde{\epsilon}_k^2 + \sigma^2}\right) \\ &\quad - (\log(\pi e \mathcal{E}_\text{s}) - h(X)) &(42) \\ &> \log|\hat{d}|^2 + \log\frac{1}{\tilde{\epsilon}_k^2 + \sigma^2/\mathcal{E}_\text{s}} - \Delta_h &(43) \end{aligned}$$

for $X, \tilde{D}, Z$ as above and for $\hat{d} \in \mathbb{C}$ deterministic. Here

$$\mathcal{E}_\text{s} = \mathsf{E}\left[|X|^2\right] \text{ and } \Delta_h = \log(\pi e \mathcal{E}_\text{s}) - h(X). \tag{44}$$

This lower bound actually holds for any law on X and has the following interpretation: it is the relative entropy distance between the law on X and a Gaussian law of equal power, subtracted from the Gaussian capacity corresponding to output power $|\hat{d}|^2\mathcal{E}_\text{s}$ and noise $\mathcal{E}_\text{s}\tilde{\epsilon}_k^2 + \sigma^2$.

For the distribution on X in which we are interested (X uniform over $\{z \in \mathbb{C} : A/2 \leq |z| \leq A\}$) we have

$$\mathsf{E}\left[|X|^2\right] = \frac{5}{8}A^2, \quad \Delta_h = \log\frac{5e}{6} \tag{45}$$

so that (43) implies

$$I\left(X; (\hat{d} + \tilde{D})X + Z\right) > \log\frac{1}{\tilde{\epsilon}_k^2 + 8/(5\text{SNR})} + \log|\hat{d}|^2 - \log\frac{5e}{6}. \tag{46}$$

To use this bound in order to lower-bound the RHS of (40) we note that the RHS of (40) is just the expectation of the left-hand side (LHS) of (46) over $\hat{d}$ with respect to the distribution of $d + \hat{D}$. Thus, from (46) and the expectation of the logarithm of a noncentral chi-square random variable of two degrees of freedom [4, Appendix X]

$$\begin{aligned} \mathsf{E}\left[\log|d + \hat{D}|^2\right] &= \log|d|^2 - \text{Ei}\left(-\frac{|d|^2}{\mathsf{E}\left[|\hat{D}|^2\right]}\right) \\ &= \log|d|^2 - \text{Ei}\left(-\frac{|d|^2}{1 - \tilde{\epsilon}_k^2}\right) \end{aligned}$$

we now obtain using (35), (40), and (41)

$$\begin{aligned} C(\text{SNR}) \geq{}& \log\frac{1}{\epsilon_{\text{pred}}^2(\delta^2) + 8/(5\text{SNR})}\bigg|_{\delta^2 = \frac{4}{\text{SNR}}} + \log|d|^2 \\ &- \text{Ei}\left(-\frac{|d|^2}{1 - \epsilon_{\text{pred}}^2(\delta^2)}\right)\bigg|_{\delta^2 = \frac{4}{\text{SNR}}} - \log\frac{5e}{6}. \end{aligned} \tag{47}$$

Here for the above to hold also in the case $d = 0$ (corresponding to a *central* chi-square random variable) we define the value of the function $\log(\xi) - \text{Ei}(-\xi)$ at $\xi = 0$ as $-\gamma$, where $\gamma \approx 0.577$ denotes Euler's constant. With this definition, the function $\log(\xi) - \text{Ei}(-\xi)$ is continuous at $\xi = 0$. In fact, it is continuous, monotonically increasing, and concave in the interval $[0, \infty)$. Its value at $\xi = 0$ is $-\gamma$ and as $\xi \uparrow \infty$ it tends to infinity logarithmically.

VI. Asymptotic Analysis

To simplify the asymptotic analysis, we shall relax the bounds at the expense of some accuracy. We begin by writing the upper bound (33) as

$$C(\text{SNR}) \leq \log\frac{1}{\epsilon_{\text{pred}}^2(1/\text{SNR})} + \log\log\text{SNR} + O(1) \tag{48}$$

where the $O(1)$ term depends on d only. We also note that the capacity is always upper bounded by the capacity $C_{\text{PSI}}(\text{SNR})$ corresponding to the case where the receiver has perfect side information, i.e., has access to the realization of the fading process. Thus,

$$\begin{aligned} C(\text{SNR}) &\leq C_{\text{PSI}}(\text{SNR}) \\ &= \mathsf{E}\left[\log\left(1 + \frac{\mathsf{E}\left[|X|^2\right] \cdot |d + H_k|^2}{\sigma^2}\right)\right] \\ &\leq \log\left(1 + \frac{\mathsf{E}\left[|X|^2\right] \cdot \left(|d|^2 + 1\right)}{\sigma^2}\right) \\ &\leq \log\left(1 + \text{SNR} \cdot \left(|d|^2 + 1\right)\right) \\ &= \log\text{SNR} + \log\left(|d|^2 + 1\right) + o(1) \end{aligned} \tag{49}$$

where the $o(1)$ term tends to zero as the SNR tends to infinity.

Here, the first inequality follows because side information cannot hurt; the second inequality follows from Jensen's inequality; and the third inequality follows because the second moment of a random variable is always upper bounded by the square of its maximal magnitude.

We next consider the lower bound (47). Since $-\text{Ei}(-\xi)$ is monotonically decreasing in ξ in the interval $(0, \infty)$ (see (10)) it follows that for $\delta^2 \leq 1$

$$-\text{Ei}\left(-\frac{|d|^2}{1 - \epsilon_{\text{pred}}^2(\delta^2)}\right) \geq -\text{Ei}\left(-\frac{|d|^2}{1 - \epsilon_{\text{pred}}^2(1)}\right).$$

Collecting the RHS of the above with the $\log|d|^2 - \log(5e/6)$ terms in the RHS of (47) into an $O(1)$ term we thus obtain for $\mathrm{SNR} \geq 4$

$$C(\mathrm{SNR}) \geq \log \frac{1}{\epsilon^2_{\mathrm{pred}}(4/\mathrm{SNR}) + \frac{2}{5}\cdot(4/\mathrm{SNR})} + O(1)$$
$$\geq \log \frac{1}{\epsilon^2_{\mathrm{pred}}(4/\mathrm{SNR}) + 4/\mathrm{SNR}} + O(1) \qquad (50)$$

where the $O(1)$ term is finite and does not depend on the SNR. (The validity of (50) in the case where $\epsilon^2_{\mathrm{pred}} = 1$, i.e., for memoryless fading, is verified separately.)

The upper bounds (48) and (49) and the lower bound (50) will be the main tools in our asymptotic analysis.

To continue with the asymptotic analysis we now distinguish between two cases depending on whether the noisy prediction error is small or large compared with the noise variance.

Small Prediction Error: By (49) and (50) we obtain

$$\lim_{\delta^2\downarrow 0} \frac{\epsilon^2_{\mathrm{pred}}(\delta^2)}{\delta^2} < \infty \Longrightarrow \lim_{\mathrm{SNR}\to\infty} \frac{C(\mathrm{SNR})}{\log \mathrm{SNR}} = 1. \qquad (51)$$

Large Prediction Error: In the other extreme we note that if

$$\lim_{\delta^2\downarrow 0} \frac{\epsilon^2_{\mathrm{pred}}(\delta^2)}{\delta^2} = \infty \qquad (52)$$

then the lower bound (50) can be simplified to yield

$$C(\mathrm{SNR}) \geq \log \frac{1}{\epsilon^2_{\mathrm{pred}}(4/\mathrm{SNR})} + O(1), \quad \text{if (52) holds.} \quad (53)$$

As we shall later see, when (52) holds, the bounds (53) and (48) will in most cases of interest suffice to capture the high-SNR behavior of channel capacity. There is, however, one more cosmetic change we would like to introduce. In view of the form of the noisy prediction error (11), it is convenient to express the bounds in terms of $\epsilon^2_{\mathrm{pred}}(\delta^2) + \delta^2$ rather than in terms of $\epsilon^2_{\mathrm{pred}}(\delta^2)$ only. To this end, we note that if (52) holds, then we can simplify (48) to

$$C(\mathrm{SNR}) \leq \log \frac{1}{\epsilon^2_{\mathrm{pred}}(1/\mathrm{SNR}) + 1/\mathrm{SNR}} + \log\log \mathrm{SNR} + O(1), \quad \text{if (52) holds.} \quad (54)$$

The bounds (54) and (50) are now both in this more convenient form.

VII. The Log-Log

In this section, we shall use the asymptotic results of Section VI to characterize the fading processes that yield a double-logarithmic dependence of channel capacity on the SNR. We will show

$$\overline{\lim_{\mathrm{SNR}\to\infty}} \frac{C(\mathrm{SNR})}{\log\log \mathrm{SNR}} < \infty \iff \overline{\lim_{\delta^2\downarrow 0}} \frac{-\int_{-1/2}^{1/2} \log\left(F'(\lambda)+\delta^2\right) \mathrm{d}\lambda}{\log\log \frac{1}{\delta^2}} < \infty \qquad (55)$$

which, in view of (11), can also be stated as

$$\overline{\lim_{\mathrm{SNR}\to\infty}} \frac{C(\mathrm{SNR})}{\log\log \mathrm{SNR}} < \infty \iff \overline{\lim_{\delta^2\downarrow 0}} \frac{\log \frac{1}{\epsilon^2_{\mathrm{pred}}(\delta^2)+\delta^2}}{\log\log \frac{1}{\delta^2}} < \infty. \qquad (56)$$

Notice that the relation on the RHS (and hence also on the LHS) of (56) is satisfied whenever $\epsilon^2_{\mathrm{pred}}(\delta^2)$ is bounded away from zero, i.e., whenever $\epsilon^2_{\mathrm{pred}} > 0$ (i.e., the fading is regular). It can, however, be satisfied also by nonregular fading processes. An example of a process for which $\epsilon^2_{\mathrm{pred}}(0) = 0$ and yet both sides of (55) are satisfied is one of spectral density

$$f(\lambda) = \begin{cases} K\cdot\exp\left\{1 - \left(\frac{\omega}{|\lambda|}\right)\right\}, & \text{if } |\lambda| \leq \omega \\ K, & \text{if } \omega \leq |\lambda| \leq 1/2 \end{cases} \qquad (57)$$

where $0 < \omega < 1/2$ is arbitrary, and where K is chosen so that the variance of the fading be 1.

To prove (56), we begin by showing that its RHS implies its LHS. We do so by showing that its RHS implies (52) so that its LHS follows from the upper bound (54) upon dividing both sides of the bound by $\log\log \mathrm{SNR}$. To see that the RHS of (56) implies (52) we assume that (52) does not hold and show that this contradicts the RHS of (56). Indeed, were (52) not to hold, it would imply that there is a sequence $\{\delta_n^2\} \downarrow 0$ and some real number M such that

$$\frac{\epsilon^2_{\mathrm{pred}}(\delta_n^2)}{\delta_n^2} < M, \qquad n = 1, 2, 3, \ldots$$

but this would imply

$$\frac{\log \frac{1}{\epsilon^2_{\mathrm{pred}}(\delta_n^2)+\delta_n^2}}{\log\log \frac{1}{\delta_n^2}} > \frac{\log \frac{1}{(M+1)\delta_n^2}}{\log\log \frac{1}{\delta_n^2}} \to \infty$$

in contradiction to the RHS of (56).

Having proved that the RHS of (56) implies the LHS, we next turn to prove the reverse. In fact, we will show that

$$\overline{\lim_{\delta^2\downarrow 0}} \frac{\log \frac{1}{\epsilon^2_{\mathrm{pred}}(\delta^2)+\delta^2}}{\log\log \frac{1}{\delta^2}} = \infty \Longrightarrow \overline{\lim_{\mathrm{SNR}\to\infty}} \frac{C(\mathrm{SNR})}{\log\log \mathrm{SNR}} = \infty. \qquad (58)$$

This actually follows quite easily from the lower bound (50). Assume the LHS of the above, and let $\delta_n^2 \downarrow 0$ be such that

$$\lim_{n\to\infty} \frac{\log \frac{1}{\epsilon^2_{\mathrm{pred}}(\delta_n^2)+\delta_n^2}}{\log\log \frac{1}{\delta_n^2}} = \infty \qquad (59)$$

and define the sequence

$$\mathrm{SNR}_n = \frac{4}{\delta_n^2}.$$

Then

$$\lim_{n\to\infty} \frac{C(\mathrm{SNR}_n)}{\log\log \mathrm{SNR}_n} = \lim_{n\to\infty} \frac{C(\mathrm{SNR}_n)}{\log\log(\mathrm{SNR}_n/4)}$$
$$\geq \lim_{n\to\infty} \frac{\log \frac{1}{\epsilon^2_{\mathrm{pred}}(\delta_n^2)+\delta_n^2}}{\log\log \frac{1}{\delta_n^2}}$$
$$= \infty$$

where the first equality follows from the behavior of the $\log\log(\cdot)$ function, the subsequent inequality from the lower bound (50), and the final equality from (59).

In subsequent work [8], it has been recently shown that when the limit on the RHS of (55) exists then one can express the "pre-loglog" as

$$\lim_{\mathrm{SNR}\to\infty} \frac{C(\mathrm{SNR})}{\log\log \mathrm{SNR}} = 1 + \lim_{\delta^2\downarrow 0} \frac{-\int_{-1/2}^{1/2} \log\left(F'(\lambda)+\delta^2\right) \mathrm{d}\lambda}{\log\log \frac{1}{\delta^2}}. \quad (60)$$

To prove this result one needs an improved lower bound on channel capacity for cases where capacity grows double-logarithmically in the SNR. Such a bound is derived in [8] by considering i.i.d. inputs where rather than having the inputs be uniformly distributed over the disc $\{z : A/2 \leq |z| \leq A\}$, the inputs are chosen to be still circularly symmetric but with the logarithm of their squared magnitudes being uniformly distributed between $\alpha \log A^2$ and $\log A^2$ for some constant $0 < \alpha < 1$.

VIII. The Pre-Log

In this section, we shall determine the asymptotic "pre-log" term. In the multiple-antenna literature this is sometimes called the "multiplexing gain," but this term does not seem very appropriate in our single-antenna context, especially since this ratio cannot exceed one, so that, if anything, it is not a "gain" but rather a "loss." We will show that the limiting ratio of channel capacity to $\log \mathrm{SNR}$ is determined by the nulls of the spectral density. It is the ratio of the total length of the frequency bands where the spectral density is null to the total frequencies

$$\boxed{\lim_{\mathrm{SNR}\to\infty} \frac{C(\mathrm{SNR})}{\log \mathrm{SNR}} = \mu\left(\{\lambda : F'(\lambda) = 0\}\right)} \quad (61)$$

where $\mu(\cdot)$ denotes the Lebesgue measure on the interval $[-1/2, 1/2]$.

To prove (61), we begin by noting that if its RHS is 1, i.e., if $F'(\lambda)$ is almost everywhere zero, then by (11), $\epsilon^2_{\mathrm{pred}}(\delta^2) = 0$ for any $\delta^2 \geq 0$. Consequently, the claim in this case follows from (51).

As to the case where the RHS of (61) is strictly smaller than 1, we note that in this case it suffices to show that

$$\lim_{\delta^2\downarrow 0} \frac{\log\left(\epsilon^2_{\mathrm{pred}}(\delta^2)+\delta^2\right)}{\log \delta^2} = \mu\left(\{\lambda : F'(\lambda) = 0\}\right). \quad (62)$$

Indeed, by rewriting (62) as

$$\lim_{\delta^2\downarrow 0} \frac{\log \frac{\epsilon^2_{\mathrm{pred}}(\delta^2)+\delta^2}{\delta^2}}{\log \frac{1}{\delta^2}} = 1 - \mu\left(\{\lambda : F'(\lambda) = 0\}\right)$$

it is readily verified that when the RHS of (61) is strictly smaller than 1, (62) implies (52), and the result then follows from (50), (52), (54), and (62).

We thus proceed to prove (62) or equivalently (in view of (11))

$$\lim_{\delta^2\downarrow 0} \frac{\int_{-1/2}^{1/2} \log\left(F'(\lambda)+\delta^2\right) \mathrm{d}\lambda}{\log \delta^2} = \mu\left(\{\lambda : F'(\lambda) = 0\}\right). \quad (63)$$

To this end, we divide up the integration in (63) into three different regions, depending on whether $F'(\lambda)$ is zero, it is in the interval $(0,1)$, or it is in the interval $[1,\infty)$. (The set of λ's for which the derivative $F'(\lambda)$ is undefined is of Lebesgue measure zero.)

$$\frac{\int_{-1/2}^{1/2} \log\left(F'(\lambda)+\delta^2\right) \mathrm{d}\lambda}{\log \delta^2} = \int_{F'(\lambda)=0} + \int_{0<F'(\lambda)<1} + \int_{F'(\lambda)\geq 1} \frac{\log\left(F'(\lambda)+\delta^2\right)}{\log \delta^2} \mathrm{d}\lambda. \quad (64)$$

The easiest term to deal with is the first term because for λ such that $F'(\lambda) = 0$, the integrand is 1, irrespective of $\delta^2 > 0$

$$\int_{F'(\lambda)=0} \frac{\log\left(F'(\lambda)+\delta^2\right)}{\log \delta^2} \mathrm{d}\lambda = \mu\left(\{\lambda : F'(\lambda) = 0\}\right), \quad \delta^2 > 0. \quad (65)$$

The third term converges to zero as $\delta^2 \downarrow 0$. This can be shown using the Monotone Convergence Theorem by noting that for any $a > 0$ the function

$$\delta^2 \mapsto \frac{\log\left(a+\delta^2\right)}{\log \delta^2} \quad (66)$$

approaches zero as $\delta^2 \downarrow 0$, and that if $a \geq 1$, then this function is monotonically decreasing in δ^2 in the interval $(0,1)$.

To demonstrate that the second integral—the one corresponding to $0 < F'(\lambda) < 1$—approaches zero as $\delta^2 \downarrow 0$, we must exercise a little more care, since the above function is no longer monotonic in $(0,1)$. Thus, rather than relying on the Monotone Convergence Theorem, we shall rely on the Dominated Convergence Theorem. Consider thus the function (66) for $0 < a < 1$. It is nonnegative for $0 < \delta^2 \leq 1-a$ and negative for $1-a < \delta^2 < 1$. It is zero at $\delta^2 = 1-a$ and converges to zero as $\delta^2 \downarrow 0$. By setting its derivative to zero, we find that for $0 < a < 1$ the function (66) has a maximum in the interval $(0,1)$ at $\delta^2 = \xi$ where ξ satisfies

$$\xi \log \xi = (a+\xi)\log(a+\xi)$$

whence the function takes on the value $\xi/(a+\xi) < 1$. Thus,

$$\sup_{0<\delta^2<1} \frac{\log\left(a+\delta^2\right)}{\log \delta^2} < 1, \qquad 0 < a < 1.$$

As to the negative of this function, we note that it is positive in the interval $(0,1)$ only for $1-a < \delta^2 < 1$ whence it is monotonically increasing. Consequently, if we limit ourselves to $0 \leq \delta^2 \leq 1/2$ we obtain

$$\max_{0\leq\delta^2\leq 1/2} -\frac{\log\left(a+\delta^2\right)}{\log \delta^2} = \begin{cases} 0, & \text{if } 1/2 \leq 1-a \\ -\frac{\log(a+1/2)}{\log 1/2}, & \text{otherwise,} \end{cases} \quad 0 < a < 1. \quad (67)$$

Consequently

$$\left|\frac{\log\left(F'(\lambda)+\delta^2\right)}{\log \delta^2}\right| \le \max\left\{1, \left|\frac{\log\left(F'(\lambda)+1/2\right)}{\log 1/2}\right|\right\},\quad 0 < F'(\lambda) < 1,\ 0 < \delta^2 \le 1/2. \tag{68}$$

Since the RHS of the above is integrable over

$$\{\lambda \in [-1/2, +1/2] : 0 < F'(\lambda) < 1\}$$

we obtain from the Dominated Convergence Theorem that the second term in (64) also converges to zero.

In subsequent work [12], it has been shown that (61) continues to hold even if in the LHS we replace channel capacity $C(\text{SNR})$ with the feedback capacity $C_{\text{FB}}(\text{SNR})$. Thus, even though the fading channels we are considering are channels with memory for which feedback *can* increase channel capacity, the increase, if any, does not change the pre-log

$$\lim_{\text{SNR}\to\infty} \frac{C_{\text{FB}}(\text{SNR})}{\log \text{SNR}} = \mu\left(\{\lambda : F'(\lambda) = 0\}\right). \tag{69}$$

Feedback also does not increase the fading number when the fading process is regular [12].

IX. Other Asymptotic Behaviors

In this section, we consider a family of spectra that gives rise to new asymptotic behaviors of channel capacity. Gaussian fading processes of these spectra result in channel capacity growing as a fractional power of $\log \text{SNR}$. Note that since these Gaussian processes have a spectral density, they are ergodic, so that there is a coding theorem for the channels they induce.

The spectra are parameterized by two parameters: $\alpha > 1$ and $0 < \omega < 1/2$. The spectral densities $f(\cdot)$ are given by

$$f(\lambda) = \begin{cases} K \cdot \exp\left\{1 - \left(\frac{\omega}{|\lambda|}\right)^{\alpha}\right\}, & \text{if } |\lambda| \le \omega \\ K, & \text{if } \omega \le |\lambda| \le 1/2 \end{cases} \tag{70}$$

where the constant K normalizes the spectrum so that

$$\int_{-1/2}^{1/2} f(\lambda)\mathrm{d}\lambda = 1. \tag{71}$$

Processes of such spectral densities are nonregular. Indeed, by (8), the prediction error ϵ^2_{pred} in the absence of noise is zero because

$$\int_{-1/2}^{1/2} \log f(\lambda)\mathrm{d}\lambda = -\infty, \qquad \alpha > 1,\ \omega > 0. \tag{72}$$

In Appendix II, it is shown using the bounds (53), and (48), that

$$\lim_{\text{SNR}\to\infty} \frac{C(\text{SNR})}{(\log \text{SNR})^{1-1/\alpha}} = \frac{2\omega\alpha}{\alpha - 1} \tag{73}$$

or, upon substituting $\beta = (\alpha - 1)/\alpha$

$$\boxed{\lim_{\text{SNR}\to\infty} \frac{C(\text{SNR})}{(\log \text{SNR})^{\beta}} = \frac{2\omega}{\beta}, \quad 0 < \beta < 1,\ 0 < \omega < 1/2.} \tag{74}$$

X. Applications to Regular Fading

While this paper has so far focused on *nonregular* Gaussian fading, it should be emphasized that the upper and lower bounds of Sections IV and V are also applicable for *regular* Gaussian fading. In fact, for regular Gaussian fading channels they can be used to assess the SNRs at which the expansion (9) is useful. A figure of merit for where this "log-log regime" begins was given in [4] in terms of communication *rates*. It was suggested that the double-logarithmic behavior begins when the rate of communication significantly exceeds the fading number χ.

With the aid of the bounds of Sections IV and V, we can obtain an indication of where the log-log regime begins in terms of the SNR. For example, the above bounds suggest the rule of thumb that the log-log regime roughly begins when the SNR is so high that

$$\epsilon^2_{\text{pred}}(1/\text{SNR}) \le 2\epsilon^2_{\text{pred}}(0) \tag{75}$$

i.e., when the prediction of the present fading based on noisy observations of the past fading corrupted by Gaussian noise of variance $1/\text{SNR}$ is almost as effective as the prediction based on noiseless observations of the past. To this one should, of course, add that the SNR has to be high enough so that the i.i.d. channel capacity be in the double-logarithmic regime, i.e., that

$$\log\left(1 + \frac{\text{SNR}\cdot|d|^2}{1+\text{SNR}}\right) \approx \log\left(1 + |d|^2\right). \tag{76}$$

XI. Summary and Conclusion

In this paper, we studied the capacity of a stationary discrete-time Gaussian fading channel with memory, where both transmitter and receiver are cognizant of the fading law (mean and autocorrelation), but neither has access to the realizations of the fading process. While previous studies [4] focused on the case where the fading process is *regular* (i.e., one where the present fading cannot be predicted precisely from past fading values), here we extended the analysis to nonregular processes as well.

It was demonstrated that while regular fading processes result in capacity growing only double-logarithmically in the SNR, nonregular fading can result in very diverse asymptotic behaviors. Capacity may grow logarithmically in the SNR, double-logarithmically, or in between, e.g., as a fractional power of the logarithm of the SNR.

When capacity grows logarithmically, it was demonstrated that the "pre-log" can be very easily determined from the spectrum of the fading process. For fading processes having a power spectral density, it is simply the Lebesgue measure of the set of harmonics in $[-1/2, 1/2]$ where the power spectral density is zero. It is interesting to compare this result to the one obtained via the block-constant fading model ($(T-1)/T$ where T is the block duration [2]) or the more general model proposed in [3] ($(T-Q)/Q$ where Q is the rank of the covariance matrix of the fading inside the block).

Note, however, that our results on stationary fading channels cannot be directly compared with those relating to the block-constant fading model and the more general model of [3] because the latter two models are not stationary. Nonstationary models are sometimes appropriate for modeling slow frequency hopping systems with random frequency hopping. For such systems, the block-constant fading model may be appropriate if we interpret T as the number of symbols transmitted per hop, but only if we are also willing to model the channel at each frequency as being a nonergodic channel with a fading level that does not vary in time.[5]

It should be pointed out that in this paper we considered, for mathematical convenience, a peak-power constraint rather than the more common average-power constraint. Clearly, the pre-log of the former cannot exceed that of the latter. We conjecture that the two pre-logs are in fact identical. Indeed, for regular processes, a peak-power constraint and an average power constraint lead to identical fading numbers [4]. From an engineering point of view, the pre-log of the peak-constrained channel is arguably more interesting than that of the average-power constrained channel. Indeed, if the latter were strictly greater than the former, it would indicate that it can only be achieved by inputs whose peak-to-average power ratios tend to infinity.

More critical, however, is the assumption that time is discrete. We suspect that the results may change once a continuous-time model is addressed. Nevertheless, the discrete-time model is of interest not only because it is tractable, but because it is relevant in practice in all systems that base their receiver on samples at the output of the matched filter, even if those do not form a sufficient statistic.

Our results indicate that the asymptotic behavior of channel capacity depends critically on the question of whether the fading process is regular or not. The difficulties of answering this question are discussed in [13]. For the reasons outlined in [13], we suspect that there may not be a definitive answer to this question. This is not to say that all channel modeling and capacity calculations are pointless. From a practical point of view, one can and should pick a model that is reasonable for the range of SNRs of interest. One must, however, exercise great caution in studying the asymptotes of channel capacity in limiting SNRs that are beyond the range of applicability of the channel model. Additionally, every asymptotic capacity calculation must be accompanied by conditions on the range of SNRs for which it is useful. For regular fading channels, the fading number [4] and the rule (75) may be indications of the range of rates and SNRs for which the expansion (9) is useful.

Appendix I
A Lower Bound on the Ricean Mutual Information

In this appendix, we prove the lower bound (42) on the mutual information across the terminals of a Ricean channel. We define $Y = (\hat{d} + \tilde{D})X + Z$ and lower-bound the mutual information $I(X;Y)$ by

$$\begin{aligned} I(X;Y) &= h(X) - h(X|Y) \\ &= h(X) - h(X - \alpha Y|Y) \\ &\geq h(X) - h(X - \alpha Y) \\ &\geq h(X) - \log\left(\pi e \mathsf{E}\left[|X - \alpha Y|^2\right]\right) \end{aligned}$$

where $\alpha \in \mathbb{C}$ is arbitrary. Here, the first inequality follows because conditioning cannot increase differential entropy, and the subsequent inequality follows because the Gaussian distribution maximizes differential entropy for a given second moment. Inequality (42) now follows by optimizing over α, i.e., by choosing α to minimize $\mathsf{E}\left[|X - \alpha Y|^2\right]$ namely

$$\begin{aligned} \alpha &= \frac{\mathsf{E}[XY^*]}{\mathsf{E}[|Y|^2]} \\ &= \frac{\mathsf{E}\left[|X|^2\right] \hat{d}^*}{\mathsf{E}[|X|^2]\left(|\hat{d}|^2 + \tilde{\epsilon}_k^2\right) + \sigma^2}. \end{aligned}$$

Appendix II
Analysis of the Family of Spectra

In this section we analyze the family of spectral densities (70) and (71).

Notice that since the RHS of (70) never exceeds K, the normalizing constant K must satisfy $K \geq 1$. However, since the RHS of (70) is equal to K for $\omega \leq |\lambda| \leq 1/2$, we must also have $2K(1/2 - \omega) \leq 1$. Thus,

$$1 \leq K \leq \frac{1}{1 - 2\omega}. \tag{77}$$

To study the prediction error in the presence of noise $\epsilon^2_{\text{pred}}(\delta^2)$, we need to study (11). As we shall see, for processes with these spectra, $\epsilon^2_{\text{pred}}(\delta^2)/\delta^2$ tends to infinity as $\delta^2 \downarrow 0$, and we shall therefore focus on the integral

$$\int_{-1/2}^{1/2} \log\left(f(\lambda) + \delta^2\right) \mathrm{d}\lambda. \tag{78}$$

We shall next proceed to estimate (78) for small δ^2. In particular, we shall assume $0 < \delta^2 \ll K$. To this end, we define $\eta = \eta(\delta^2)$ as the solution in $(0, \omega)$ of the equation

$$f(\eta) = \delta^2$$

or explicitly as

$$\eta = \frac{\omega}{\left(1 + \log\frac{K}{\delta^2}\right)^{1/\alpha}}. \tag{79}$$

Notice that since $f(\lambda)$ is monotonic on $[0, 1/2]$ it follows that

$$f(\lambda) \leq \delta^2, \qquad 0 \leq \lambda \leq \eta \tag{80}$$

$$f(\lambda) \geq \delta^2, \qquad \eta \leq \lambda \leq 1/2. \tag{81}$$

By symmetry

$$\int_{-1/2}^{1/2} \log\left(f(\lambda) + \delta^2\right) \mathrm{d}\lambda = 2\int_{0}^{1/2} \log\left(f(\lambda) + \delta^2\right) \mathrm{d}\lambda \tag{82}$$

and we thus proceed to estimate the integral over $\lambda \in [0, 1/2]$. We break this integral into three integrals over the intervals $[0, \eta]$, $[\eta, \omega]$, and $[\omega, 1/2]$. Using (80), we can bound the integrand in the first integral by

$$\log \delta^2 \leq \log\left(f(\lambda) + \delta^2\right) \leq \log\left(2\delta^2\right), \qquad 0 \leq \lambda \leq \eta$$

to conclude that

$$\int_0^{\eta} \log\left(f(\lambda) + \delta^2\right) \mathrm{d}\lambda = \eta \log \delta^2 + o(1) \tag{83}$$

[5]Since such a nonergodic channel has a pre-log of 1, one would be asymptotically better off not hopping at all.

where the $o(1)$ term is between 0 and $\eta \log 2$ and thus tends to zero as δ^2 approaches zero. Using (81) we obtain

$$\log f(\lambda) \le \log\left(f(\lambda)+\delta^2\right) \le \log\left(2f(\lambda)\right), \qquad \eta \le \lambda \le \omega$$

to conclude that

$$\int_\eta^\omega \log\left(f(\lambda)+\delta^2\right) d\lambda = \int_\eta^\omega \log f(\lambda) \mathrm{d}\lambda + O(1) = (\omega-\eta)\log(Ke) + \frac{\omega}{\alpha-1} - \frac{\omega^\alpha}{\alpha-1}\cdot\frac{1}{\eta^{\alpha-1}} + O(1) \tag{84}$$

where the $O(1)$ terms are between 0 and $\omega \log 2$. Finally, the integral over $[\omega, 1/2]$ can be precisely computed as

$$\int_\omega^{1/2} \log\left(f(\lambda)+\delta^2\right) \mathrm{d}\lambda = (1/2-\omega)\log(K+\delta^2). \tag{85}$$

It thus follows from (82)–(85) that

$$\int_{-1/2}^{1/2} \log\left(f(\lambda)+\delta^2\right) \mathrm{d}\lambda = 2\eta\log\delta^2 - 2\frac{\omega^\alpha}{\alpha-1}\frac{1}{\eta^{\alpha-1}} + O(1) \tag{86}$$

where the $O(1)$ is bounded in δ^2.

We now note that by (79)

$$\eta \log\frac{1}{\delta^2} = \omega\left(\log\frac{1}{\delta^2}\right)^{1-1/\alpha} + o(1)$$

and

$$\frac{\omega^\alpha}{\alpha-1}\frac{1}{\eta^{\alpha-1}} = \frac{\omega}{\alpha-1}\left(\log\frac{1}{\delta^2}\right)^{1-1/\alpha} + o(1)$$

so that by (86)

$$\int_{-1/2}^{1/2} \log\left(f(\lambda)+\delta^2\right) \mathrm{d}\lambda = -\frac{2\omega\alpha}{\alpha-1}\left(\log\frac{1}{\delta^2}\right)^{1-1/\alpha} + O(1).$$

Since $\log\delta^2$ is much more negative than the RHS of the above, we conclude that in (11) the integral is, indeed, the dominant term; that (52) holds; and that

$$\log\frac{1}{\epsilon^2_{\text{pred}}(\delta^2)} = \frac{2\omega\alpha}{\alpha-1}\left(\log\frac{1}{\delta^2}\right)^{1-1/\alpha} + O(1). \tag{87}$$

The asymptotic behavior of the capacity (73) can be now deduced from (87), (53), and (48).

ACKNOWLEDGMENT

The author wishes to thank Giuseppe Caire for interesting discussions on the modeling of fading channels, Sergio Verdú for his comments that stimulated this research, and İ. Emre Telatar for his constructive editorial comments.

REFERENCES

[1] T. Marzetta and B. Hochwald, “Capacity of a multiple-antenna communication link in Rayleigh flat fading environment,” *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.
[2] L. Zheng and D. N. C. Tse, “Communicating 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] 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.
[4] A. Lapidoth and S. M. Moser, “Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels,” *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.
[5] R.-R. Chen, B. Hajek, R. Koetter, and U. Madhow, “On fixed input distributions for noncoherent communication over high-SNR Rayleigh-fading channels,” *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3390–3396, Dec. 2004.
[6] R. Etkin and D. N. C. Tse, “Degrees of freedom in underspread MIMO fading channels,” preprint, 2003.
[7] A. Lapidoth and S. M. Moser, “The fading number of SIMO fading channels with memory,” in *Proc. 2004 Int. Symp. Information Theory and Its Applications (ISITA’04)*, Parma, Italy, Oct. 10–13, 2004, pp. 287–292.
[8] T. Koch, “On the asymptotic capacity of multiple-input single-output fading channels with memory,” Master’s thesis, ETH, Zurich, Switzerland, 2004.
[9] J. Doob, *Stochastic Processes*. New York: Wiley, 1990.
[10] I. Ibragimov and Y. Rozanov, *Gaussian Random Processes*. Berlin, Germany: Springer-Verlag, 1978.
[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[12] A. Lapidoth and S. M. Moser, “Feedback increases neither the fading number nor the pre-log,” in *Proc. 23rd IEEE Conv. Electrical and Electronics Engineers in Israel (IEEEI’04)*, Herzlia, Israel, Sep. 6–7, 2004, pp. 213–215.
[13] 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. 1–3, 2003.