On the Asymptotic Capacity of Fading Channels

Amos Lapidoth*

Abstract

We consider a peak-power limited single-antenna flat complex-Gaussian fading channel where the receiver and transmitter, while fully congnizant 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 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 Signal-to-Noise Ratio (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.

KEYWORDS: Fading Channels, Channel Capacity, Asymptotic Expansion, high SNR, multiplexing gain, Rayleigh, Rice, time-selective, non-coherent.

1 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-

*The author is with the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland.

symmetric complex-Gaussian process whose law (i.e, mean and auto-correlation 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 "non-coherent" 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) IID fading. Our emphasis here will, however, be on the case where the fading process has memory (is not IID) 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. 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"[1] 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 non-stationary but it has a more intricate structure. The fading is IID 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 $(T-Q)/T$.

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

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. 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.

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 (22) and lower (36) bounds on the capacity of this channel with a view to an understanding of channel capacity at high SNR (37), (42). With the aid of these bound we shall obtain:

- 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 7 (44).
- An expression for the pre-log when capacity grows logarithmically in the SNR (50).
- 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 9 (72).

The rest of this paper is organized as follows. In the next section we describe the channel model and define its capacity. In the subsequent section we discuss the classical prediction problem and the noisy prediction problem for stationary circularly symmetric Gaussian processes. Section 4 then addresses upper bounds on capacity, while Section 5 addresses lower bounds. An asymptotic analysis of these bounds is performed in Section 6. This analysis is used in Section 7 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 8 and asymptotic behaviors other than logarithmic or double-logarithmic are addressed in Section 9. The paper concludes with a brief summary and some conclusions in Section 10.

2 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 independent and identically distributed (IID) 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

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

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* $F(\lambda)$, $-1/2 \leq \lambda \leq 1/2$. Thus, $F(\cdot)$ is a monotonically non-decreasing function on $[-1/2, 1/2]$ [5, Theorem 3.2, p. 474],

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

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. Typically one considers channel capacity under an energy constraint on the input, but, to simplify the analysis, we have chosen in this paper to consider the peak-power constraint:

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

We define the signal-to-noise ratio SNR by

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

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

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

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

It should be noted that $C(\mathrm{SNR})$ 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*.

3 Noiseless and Noisy Prediction

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

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

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 SNR $\rightarrow \infty$.

Doob [5, XII.2, p. 564] refers to processes for which $\epsilon^2_{\text{pred}} > 0$ as *regular* and to those for which $\epsilon^2_{\text{pred}} = 0$ as *non-regular* or *deterministic*. Note, however, that Ibragimov and Rozanov [6] 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 $\epsilon^2_{\text{pred}} = 0$. 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 IID $\mathcal{N}_{\mathbb{C}}(0, \delta^2)$ 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 $\epsilon^2_{\text{pred}}(\delta^2)$ and note that it is given by

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

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, $\epsilon^2_{\text{pred}}(\delta^2)$ 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

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

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

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

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

Finally, if $\{A_k\}$ is additionally stationary, then the prediction error is monotonically non-increasing in n and

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

(For the latter claim, see [5, p. 562], [5, IV, Theorem 7.4], [5, VII, Theorem 4.3].)

4 An Upper Bound

To upper bound $I(X^n;Y^n)$ we begin by using the chain rule

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

and upper bounding 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 non-negativity of mutual information; the following equality from the absence of feedback, which results in future inputs being independent of the present input given the past inputs and outputs; and the last equality from the expansion of mutual information in terms of differential entropies.

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

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

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:

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

where the second equality follows from the observation that fixing the law of X_k also fixes 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 fixes 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 IID $\mathcal{N}_{\mathbb{C}}(0, \sigma^2/A^2)$, 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

$$
\begin{aligned}
\frac{Y_\nu}{X_\nu} - d &= \frac{H_\nu X_\nu + Z_\nu}{X_\nu} \qquad (19) \\
&\sim H_\nu + W_\nu. \qquad (20)
\end{aligned}
$$

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

$$
\begin{aligned}
&\sup_{p_{X^n}} I(X^n; Y_k|Y^{k-1}) \\
&\quad \leq \sup_{p_{X_k}} I(X_k, H_{k-1} + W_{k-1}, \ldots, H_1 + W_1; Y_k) \\
&\quad \leq \sup_{p_{X_0}} I\left(X_0, \{H_\nu + W_\nu\}_{\nu=-\infty}^{-1}; Y_0\right) \\
&\quad \leq \sup_{p_{X_0}} I(X_0; Y_0) + \sup_{p_{X_0}} I\Big(\{H_\nu + W_\nu\}_{\nu=-\infty}^{-1}; Y_0 \Big| X_0 \Big)
\end{aligned}
$$

$$
\begin{aligned}
&\leq \sup_{p_{X_0}} I(X_0;Y_0) + I\Big(\{H_\nu + W_\nu\}_{\nu=-\infty}^{-1}\,;\,H_0\Big)\\
&= \sup_{p_{X_0}} I(X_0;Y_0) + \log\frac{1}{\epsilon^2_{\text{pred}}(\sigma^2/A^2)}\\
&\leq \log\log\frac{A^2}{\sigma^2} + \log|d|^2 - \mathrm{Ei}(-|d|^2) - 1 + \log\frac{1}{\epsilon^2_{\text{pred}}(\sigma^2/A^2)} + o(1)\\
&= \log\log\mathrm{SNR} + \log|d|^2 - \mathrm{Ei}(-|d|^2) - 1 + \log\frac{1}{\epsilon^2_{\text{pred}}(1/\mathrm{SNR})} + o(1),
\end{aligned}
$$

where the $o(1)$ term follows from the analysis of the capacity of the Ricean channel at high SNR [4]. It depends on d and A^2/σ^2 and tends, for any fixed d, to zero as $A^2/\sigma^2 \to \infty$.

Note: A better bound for the non-asymptotic regime results when the term

$$
\sup_{p_{X_0}} I(X_0;Y_0) \tag{21}
$$

is upper bounded by the tighter firm bound on the capacity of the memoryless Ricean fading channel proposed in [4].

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

$$
C(\mathrm{SNR}) \leq \log\frac{1}{\epsilon^2_{\text{pred}}(1/\mathrm{SNR})} + \log\log\mathrm{SNR} + \log|d|^2 - \mathrm{Ei}(-|d|^2) - 1 + o(1), \tag{22}
$$

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].

5 A Lower Bound

To derive a lower bound on channel capacity we shall consider inputs $\{X_k\}$ that are IID and uniformly distributed over the set $\{z \in \mathbb{C} : A/2 \leq |z| \leq A\}$. Using the chain rule [7, Thm. 2.5.2]

$$
\frac{1}{n} I\big(X^n;Y^n\big) = \frac{1}{n}\sum_{k=1}^{n} I\big(X_k;Y^n \,\big|\, X^{k-1}\big) \tag{23}
$$

and a Cesáro-type theorem [7, Thm. 4.2.3] we obtain that the capacity C can be lower bounded by:

$$
C \geq \liminf_{k\to\infty} I\big(X_k;Y^n \,\big|\, X^{k-1}\big). \tag{24}
$$

We now proceed to lower bound the term on the RHS of (24) using the fact that we have chosen $\{X_k\}$ to be IID and satisfying $|X_k| \geq A/2$, almost surely:

$$
\begin{aligned}
I\big(X_k; Y^n \,\big|\, X^{k-1}\big) &\geq I\big(X_k; Y^k \,\big|\, X^{k-1}\big) \\
&= I\big(X_k; X^{k-1}, Y^{k-1}, Y_k\big) \\
&= I\left(X_k\,;\, Y_k, \left\{\frac{Y_\nu}{X_\nu} - d\right\}_{\nu=1}^{k-1}, X^{k-1}\right) \\
&= I\left(X_k\,;\, Y_k, \left\{\frac{Y_\nu}{X_\nu} - d\right\}_{\nu=1}^{k-1} \,\middle|\, X^{k-1}\right) \\
&= I\left(X_k\,;\, Y_k, \left\{H_\nu + \frac{Z_\nu}{X_\nu}\right\}_{\nu=1}^{k-1} \,\middle|\, X^{k-1}\right) \\
&\geq I\left(X_k\,;\, Y_k, \{H_\nu + W'_\nu\}_{\nu=1}^{k-1}\right) \\
&= I\left(X_k\,;\, Y_k \,\middle|\, \{H_\nu + W'_\nu\}_{\nu=1}^{k-1}\right)
\end{aligned}
\tag{25}
$$

where

$$
\{W'_\nu\} \sim \text{IID} \quad \mathcal{N}_{\mathbb{C}}\left(0, \frac{\sigma^2}{(A/2)^2}\right). \tag{26}
$$

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

Expressing the present fading H_k as

$$
H_k = \hat{D} + \tilde{D}, \tag{27}
$$

where

$$
\hat{D} = \mathsf{E}\left[H_k \middle| \{H_\nu + W'_\nu\}_{\nu=1}^{k-1}\right] \tag{28}
$$

we obtain from (25) that

$$
I\big(X_k; Y^n \,\big|\, X^{k-1}\big) \geq I\big(X; (d + \hat{D} + \tilde{D})X + Z \,\big|\, \hat{D}\big), \tag{29}
$$

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 $\mathcal{N}_{\mathbb{C}}(0, \tilde{\epsilon}_k^2)$ 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)

$$\lim_{k\to\infty} \tilde{\epsilon}_k^2 = \epsilon_{\text{pred}}^2(\delta^2)\Big|_{\delta^2 = \frac{\sigma^2}{(A/2)^2}}. \tag{30}$$

To lower bound the RHS of (29) we derive in Appendix A the lower bound

$$I\big(X; (\hat{d} + \tilde{D})X + Z\big) \geq \log\left(1 + \frac{\mathcal{E}_\text{s}|\hat{d}|^2}{\mathcal{E}_\text{s}\tilde{\epsilon}_k^2 + \sigma^2}\right) - \big(\log(\pi e \mathcal{E}_\text{s}) - h(X)\big) \tag{31}$$

$$> \log|\hat{d}|^2 + \log\frac{1}{\tilde{\epsilon}_k^2 + \sigma^2/\mathcal{E}_\text{s}} - \Delta_h, \tag{32}$$

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

$$\mathcal{E}_\text{s} = \mathsf{E}\left[|X|^2\right] \qquad \text{and} \qquad \Delta_h = \log(\pi e \mathcal{E}_\text{s}) - h(X). \tag{33}$$

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 (uniform over $\{z \in \mathbb{C} : A/2 \leq |z| \leq A\}$) we have

$$\mathsf{E}\left[|X|^2\right] = \frac{5}{8}A^2, \qquad \Delta_h = \log\frac{5e}{6}, \tag{34}$$

so that (32) implies:

$$I\big(X; (\hat{d} + \tilde{D})X + Z\big) > \log\frac{1}{\tilde{\epsilon}_k^2 + 8/(5\text{SNR})} + \log|\hat{d}|^2 - \log\frac{5e}{6}. \tag{35}$$

To use this bound in order to lower bound the RHS of (29) we note that the RHS of (29) is just the expectation of the LHS of (35) over $\hat{d}$ with respect to the distribution of $d + \hat{D}$. Thus, from (35) and the expectation of the logarithm of a non-central chi-square random of two degrees of freedom [4]

$$\mathsf{E}\left[\log|d + \hat{D}|^2\right] = \log|d|^2 + \text{Ei}\left(-\frac{|d|^2}{\mathsf{E}\left[|\hat{D}|^2\right]}\right)$$

$$= \log(|d|^2) - \mathrm{Ei}\left(-\frac{|d|^2}{1-\tilde{\epsilon}_k^2}\right),$$

we now obtain using (24), (29), and (30)

$$C(\mathrm{SNR}) \geq \log \frac{1}{\epsilon^2_{\mathrm{pred}}(\delta^2) + 8/(5\mathrm{SNR})}\Bigg|_{\delta^2=\frac{4}{\mathrm{SNR}}} + \log|d|^2 - \mathrm{Ei}\left(-\frac{|d|^2}{1-\epsilon^2_{\mathrm{pred}}(\delta^2)}\right)\Bigg|_{\delta^2=\frac{4}{\mathrm{SNR}}} - \log\frac{5e}{6}. \quad (36)$$

6 Asymptotic Analysis

To simplify the asymptotic analysis we shall simplify the bounds at some cost in accuracy. We begin by writing the upper bound (22) as:

$$C(\mathrm{SNR}) \leq \log \frac{1}{\epsilon^2_{\mathrm{pred}}(1/\mathrm{SNR})} + \log\log\mathrm{SNR} + O(1), \quad (37)$$

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

$$\begin{aligned} C(\mathrm{SNR}) &\leq C_{\mathrm{PSI}}(\mathrm{SNR}) \\ &= \mathsf{E}\left[\log\left(1 + \frac{\mathsf{E}\left[|X|^2\right]\cdot|d+H_k|^2}{\sigma^2}\right)\right] \\ &\leq \log\left(1 + \frac{\mathsf{E}\left[|X|^2\right]\cdot\left(|d|^2+1\right)}{\sigma^2}\right) \\ &\leq \log\Big(1 + \mathrm{SNR}\cdot\left(|d|^2+1\right)\Big) \\ &= \log\mathrm{SNR} + \log\left(|d|^2+1\right) + o(1), \end{aligned} \quad (38)$$

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

As to the lower bound, we rewrite (36) as:

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

$$\geq \log \frac{1}{\epsilon^2_{\text{pred}}(4/\text{SNR}) + 4/\text{SNR}} + O(1). \tag{39}$$

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 (38) and (39) we obtain

$$\lim_{\delta^2 \downarrow 0} \frac{\epsilon^2_{\text{pred}}(\delta^2)}{\delta^2} < \infty \implies \lim_{\text{SNR} \to \infty} \frac{C(\text{SNR})}{\log \text{SNR}} = 1. \tag{40}$$

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

$$\lim_{\delta^2 \downarrow 0} \frac{\epsilon^2_{\text{pred}}(\delta^2)}{\delta^2} = \infty \tag{41}$$

then the lower bound (39) can be simplified to yield

$$C(\text{SNR}) \geq \log \frac{1}{\epsilon^2_{\text{pred}}(4/\text{SNR})} + O(1), \quad \text{if (41) holds.} \tag{42}$$

Compare (42) and (37).

In view of the form of the noisy prediction error (11) it is convenient to express the bounds in terms of $\epsilon^2_{\text{pred}}(\delta^2) + \delta^2$ rather than in terms of $\epsilon^2_{\text{pred}}(\delta^2)$ only. To this end we note that if (41) holds, then we can simplify (37) to:

$$C(\text{SNR}) \leq \log \frac{1}{\epsilon^2_{\text{pred}}(1/\text{SNR}) + 1/\text{SNR}} + \log \log \text{SNR} + O(1), \quad \text{if (41) holds.} \tag{43}$$

Compare (43) and (39).

7 The Log-Log

In this section we shall use the asymptotic results of Section 6 to characterize the fading processes that yield a double-logarithmic dependence of channel capacity on the SNR. We will show

$$\boxed{\overline{\lim_{\text{SNR} \to \infty}} \frac{C(\text{SNR})}{\log \log \text{SNR}} < \infty \iff \overline{\lim_{\delta^2 \downarrow 0}} \frac{-\int_{-1/2}^{1/2} \log\bigl(F'(\lambda) + \delta^2\bigr)\, \mathrm{d}\lambda}{\log \log \frac{1}{\delta^2}} < \infty,} \tag{44}$$

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

$$\overline{\lim_{\mathrm{SNR}\to\infty}} \frac{C(\mathrm{SNR})}{\log\log \mathrm{SNR}} < \infty \Longleftrightarrow \overline{\lim_{\delta^2 \downarrow 0}} \frac{\log \frac{1}{\epsilon^2_{\mathrm{pred}}(\delta^2)+\delta^2}}{\log\log \frac{1}{\delta^2}} < \infty. \tag{45}$$

Notice that the above condition is satisfied whenever $\epsilon^2_{\mathrm{pred}}(\delta^2)$ is bounded away from zero, i.e., whenever $\epsilon^2_{\mathrm{pred}} > 0$. It can, however, be satisfied also by non regular fading processes. An example of a process for which $\epsilon^2_{\mathrm{pred}}(0) = 0$ and yet (44) is satisfied is one of spectral density:

$$f(\lambda) = \begin{cases} K \cdot \exp\left\{1 - \left(\frac{\omega}{|\lambda|}\right)\right\} & \text{if } |\lambda| \leq \omega \\ K & \text{if } \omega \leq |\lambda| \leq 1/2 \end{cases}, \tag{46}$$

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

To prove (45) we first show that the RHS implies the LHS using the upper bound (37). To this end we begin by noting that the RHS implies that

$$\overline{\lim_{\delta^2 \downarrow 0}} \frac{\epsilon^2_{\mathrm{pred}}(\delta^2)}{\delta^2} = \infty, \tag{47}$$

for otherwise we could find a sequence $\delta_n^2 \downarrow 0$ and some M such that

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

so that

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

in contradiction to the RHS of (45).

Having established that the RHS of (45) implies (47), we now note that the two combine to imply

$$\overline{\lim_{\delta^2 \downarrow 0}} \frac{\log \frac{1}{\epsilon^2_{\mathrm{pred}}(\delta_n^2)}}{\log\log \frac{1}{\delta_n^2}} < \infty,$$

which combines with the upper bound (37) to imply the LHS.

Having proved that the RHS of (45) implies the LHS, we next turn to prove the reverse. In fact, we will show that

$$\overline{\lim_{\delta^2 \downarrow 0}} \frac{\log \frac{1}{\epsilon^2_{\text{pred}}(\delta^2)+\delta^2}}{\log \log \frac{1}{\delta^2}} = \infty \Longrightarrow \overline{\lim_{\text{SNR}\to\infty}} \frac{C(\text{SNR})}{\log \log \text{SNR}} = \infty. \tag{48}$$

This actually follows quite easily from the lower bound (39). Assume the LHS of the above, and let $\delta_n^2 \downarrow 0$ be such that

$$\lim_{n\to\infty} \frac{\log \frac{1}{\epsilon^2_{\text{pred}}(\delta_n^2)+\delta_n^2}}{\log \log \frac{1}{\delta_n^2}} = \infty \tag{49}$$

and define the sequence

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

Then

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

where the first equality follows from the behavior of the $\log\log(\cdot)$ function, the subsequent inequality from the lower bound (39), and the final equality from (49).

8 The Pre-Log

In this section we shall determine the asymptotic "pre-log" term. In the multi-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 $\log \text{SNR}$ 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:

$$\boxed{\lim_{\mathrm{SNR}\to\infty}\frac{C(\mathrm{SNR})}{\log\mathrm{SNR}}=\mu\Big(\Big\{\lambda : F'(\lambda)=0\Big\}\Big),} \tag{50}$$

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

To prove (50) we begin by noting that if its RHS is 1, i.e., if $F'(\lambda)$ is almost everywhere zero, then by (11) $\epsilon^2_{\text{pred}}(\delta^2)=0$ for any $\delta^2\geq 0$. Consequently the claim in this case follows from (40).

As to the case where the RHS of (50) is strictly smaller than 1, we note that in this case it suffices to show

$$\lim_{\delta^2\downarrow 0}\frac{\log\big(\epsilon^2_{\text{pred}}(\delta^2)+\delta^2\big)}{\log\delta^2}=\mu\Big(\Big\{\lambda : F'(\lambda)=0\Big\}\Big), \tag{51}$$

because this would imply that (41) holds, and the result would then follows from (39) and (43).

We thus proceed to prove (51) or equivalently (in view of (11))

$$\lim_{\delta^2\downarrow 0}\frac{\int_{-1/2}^{1/2}\log\big(F'(\lambda)+\delta^2\big)\,\mathrm{d}\lambda}{\log\delta^2}=\mu\Big(\Big\{\lambda : F'(\lambda)=0\Big\}\Big). \tag{52}$$

To this end we divide up the integration in (52) into three different regions[2], depending on whether $F'(\lambda)$ is zero, it is in the interval $(0,1)$, or it is in the interval $[1,\infty)$:

$$\frac{\int_{-1/2}^{1/2}\log\big(F'(\lambda)+\delta^2\big)\,\mathrm{d}\lambda}{\log\delta^2}= \\ \int_{F'(\lambda)=0}+\int_{0<F'(\lambda)<1}+\int_{F'(\lambda)\geq 1}\frac{\log\big(F'(\lambda)+\delta^2\big)}{\log\delta^2}\,\mathrm{d}\lambda. \tag{53}$$

The easiest term to deal with is the first term because the integrand does not depend on $\delta^2>0$:

$$\int_{F'(\lambda)=0}\frac{\log\big(F'(\lambda)+\delta^2\big)}{\log\delta^2}\,\mathrm{d}\lambda=\mu\Big(\Big\{\lambda : F'(\lambda)=0\Big\}\Big),\quad \delta^2>0. \tag{54}$$

[2]The set of λ's for which the derivative $F'(\lambda)$ is undefined is of Lebesgue measure zero.

The third term is easily handled using the Monotone Convergence Theorem by noting that for any $a > 0$ the function

$$\delta^2 \mapsto \frac{\log\bigl(a+\delta^2\bigr)}{\log \delta^2} \tag{55}$$

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, 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. By setting its derivative to zero, we find that for $0 < a < 1$ the function (55) 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

$$\frac{\xi}{a+\xi} < 1.$$

Consequently, for small δ^2, e.g., $\delta^2 < 1/2$ the maximum value of the magnitude of the function (55) is either achieved at $\delta^2 = 1/2$ or else inside the interval $(0,1/2)$ whence it is upper bounded by 1. Thus

$$\left|\frac{\log\bigl(F'(\lambda)+\delta^2\bigr)}{\log \delta^2}\right| \leq \max\left\{1, \left|\frac{\log\bigl(F'(\lambda)+1/2\bigr)}{\log 1/2}\right|\right\}, \qquad 0 < F'(\lambda) < 1, 0 \leq \delta^2 \leq 1/2. \tag{56}$$

Since the RHS 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 (53) converges to zero.

9 Other Asymptotic Behaviors

In this section we consider a family of spectra that will give rise to new asymptotic behaviors of channel capacity.

The spectra are parametrized by two parameters: $\alpha > 1$ and $0 < \omega < 1/2$. They are given by

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

where the constant K normalizes the spectrum so that

$$\int_{-1/2}^{1/2} f(\lambda)\, \mathrm{d}\lambda = 1. \tag{58}$$

Notice that since the RHS of (57) never exceeds K, the normalizing constant K must satisfy $K \geq 1$. However, since the RHS of (57) is equal to K for $\omega \leq |\lambda| \leq 1/2$ we must also have $2K(1/2 - \omega) \leq 1$. Thus,

$$1 \leq K \leq \frac{1}{1 - 2\omega}. \tag{59}$$

Note also that for processes of these spectra, the prediction error ϵ^2_{pred} in the absence of noise is zero:

$$\int_{-1/2}^{1/2} \log f(\lambda)\, \mathrm{d}\lambda = -\infty, \quad \alpha > 1, \omega > 0. \tag{60}$$

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, and we shall therefore focus on the integral

$$\int_{-1/2}^{1/2} \log\big(f(\lambda) + \delta^2\big)\, \mathrm{d}\lambda \tag{61}$$

without paying attention the the term δ^2, which must be subtracted to obtain $\epsilon^2_{\text{pred}}(\delta^2)$.

We shall next proceed to estimate (61) for small δ^2. In particular, we shall assume $0 < \delta^2 \ll K$. To this end, we define $\eta(\delta^2)$ as the solution in $(0, \omega)$ of the equation[3]:

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

[3] To simplify notation we make the dependence of η on δ^2 implicit and write η rather than $\eta(\delta^2)$.

or explicitly as

$$\eta = \frac{\omega}{\left(1 + \log \frac{K}{\delta^2}\right)^{1/\alpha}}. \tag{62}$$

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

$$f(\lambda) \le \delta^2, \quad 0 \le \lambda \le \eta, \tag{63}$$

$$f(\lambda) \ge \delta^2, \quad \eta \le \lambda \le 1/2. \tag{64}$$

By symmetry,

$$\int_{-1/2}^{1/2} \log\bigl(f(\lambda) + \delta^2\bigr)\, \mathrm{d}\lambda = 2 \int_0^{1/2} \log\bigl(f(\lambda) + \delta^2\bigr)\, \mathrm{d}\lambda, \tag{65}$$

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 (63) we can bound the integrand in the first integral by

$$\log \delta^2 \le \log\bigl(f(\lambda) + \delta^2\bigr) \le \log\bigl(2\delta^2\bigr), \quad 0 \le \lambda \le \eta,$$

to conclude that

$$\int_0^{\eta} \log\bigl(f(\lambda) + \delta^2\bigr)\, \mathrm{d}\lambda = \eta \log \delta^2 + o(1), \tag{66}$$

where the $o(1)$ term is between 0 and $\eta \log 2$ and thus tends to zero as δ^2 approaches zero. Using (64) we obtain

$$\log f(\lambda) \le \log\bigl(f(\lambda) + \delta^2\bigr) \le \log\bigl(2 f(\lambda)\bigr), \quad \eta \le \lambda \le \omega,$$

to conclude that

$$\begin{aligned} \int_{\eta}^{\omega} \log\bigl(f(\lambda) + \delta^2\bigr) d\lambda &= \int_{\eta}^{\omega} \log f(\lambda)\, \mathrm{d}\lambda + O(1) \\ &= (\omega - \eta) \log(Ke) + \frac{\omega}{\alpha - 1} - \frac{\omega^{\alpha}}{\alpha - 1} \cdot \frac{1}{\eta^{\alpha-1}} + O(1), \end{aligned} \tag{67}$$

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

$$\int_{\omega}^{1/2} \log\bigl(f(\lambda) + \delta^2\bigr)\, \mathrm{d}\lambda = (1/2 - \omega) \log(K + \delta^2). \tag{68}$$

It thus follows from (65), (66), (67) and (68) that

$$\int_{-1/2}^{1/2} \log\bigl(f(\lambda) + \delta^2\bigr)\,\mathrm{d}\lambda = 2\eta \log \delta^2 - 2\frac{\omega^\alpha}{\alpha - 1}\frac{1}{\eta^{\alpha-1}} + O(1), \tag{69}$$

where the $O(1)$ is bounded in δ^2.

We now note that by (62)

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

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 (69)

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

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 (41) holds; and that

$$\log \frac{1}{\epsilon^2_{\text{pred}}(\delta^2)} = \frac{2\omega\alpha}{\alpha - 1}\left(\log \frac{1}{\delta^2}\right)^{1-1/\alpha} + O(1). \tag{70}$$

The asymptotic behavior of the capacity can be deduced from (70), (42), and (37). For example,

$$\lim_{\mathrm{SNR}\to\infty} \frac{C(\mathrm{SNR})}{\bigl(\log \mathrm{SNR}\bigr)^{1-1/\alpha}} = \frac{2\omega\alpha}{\alpha - 1}, \tag{71}$$

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

$$\boxed{\lim_{\mathrm{SNR}\to\infty} \frac{C(\mathrm{SNR})}{\bigl(\log \mathrm{SNR}\bigr)^{\beta}} = \frac{2\omega}{\beta}, \quad 0 < \beta < 1, \quad 0 < \omega < 1/2.} \tag{72}$$

10 Summary and Conclusions

In this paper we studied the capacity of a discrete-time Gaussian fading channel with memory, where both transmitter and receiver are cognizant of the fading law (mean and auto-correlation), 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 non-regular processes too.

It was demonstrated that while regular fading processes result in capacity growing only double-logarithmically in the SNR, non-regular 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 merely 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).

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. We suspect, however, that this makes little difference in the asymptotic high SNR regime. Indeed, for regular process, a peak power constraint and an average power constraint lead to identical fading numbers [4].

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 condition.

Finally we should comment on our assumption of a single-antenna. We first note that most of the analysis can be extended to channel where the receiver (but not transmitter) uses more than one antenna. Moreover, while multi-antenna systems are of great practical value, it is felt that there is still much that needs to be understood about single-antenna fading channels. The asymptotics addressed in this paper are but a step in that direction.

A A Lower Bound on the Ricean Mutual Information

In this appendix we prove the lower bound (31) 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 (31) now follows by optimizing over α, i.e., by choosing α to minimize $\mathsf{E}\left[|X - \alpha Y|^2\right]$ namely

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

Acknowledgment

The author wishes to thank Sergio Verdú whose comments were an impetus to this research.

References

[1] T. Marzetta and B. Hochwald, "Capacity of a multiple-antenna communication link in Rayleigh flat fading environment," *IEEE Trans. on Inform. Theory*, vol. 45, pp. 139–157, Jan. 1999.

[2] L. Zheng and D. Tse, "Communicating on the Grassmann manifold: A geometric approach to the non-coherent multiple antenna channel," *IEEE Trans. on Inform. Theory*, vol. 48, pp. 359–383, Feb. 2002.

[3] Y. Liang and V. V. Veeravalli, "Capacity of noncoherent time-selective Rayleigh fading channels," `submitted to IEEE Trans. on Inform. Theory`, 2003.

[4] A. Lapidoth and S. Moser, "Capacity bounds via duality with applications to multi-antenna systems on flat fading channels," *to appear in IEEE Trans. on Inform. Theory Special Issue on Space-Time Transmission, Reception, Coding and Signal Design*, 2003.

[5] J. Doob, *Stochastic Processes.* John Wiley & Sons, 1990.

[6] I. Ibragimov and Y. Rozanov, *Gaussian Random Processes.* Springer-Verlag, 1978.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: J. Wiley & Sons, 1991.