

Scalar coherent fading channel: dispersion analysis

Yury Polyanskiy and Sergio Verdú

Abstract—The backoff from capacity due to finite blocklength can be assessed accurately from the channel dispersion. This paper analyzes the dispersion of a single-user, scalar, coherent fading channel with additive Gaussian noise. We obtain a convenient two-term expression for the channel dispersion which shows that, unlike the capacity, it depends crucially on the dynamics of the fading process.

Index Terms—Shannon theory, channel capacity, channel coding, finite blocklength regime, fading channel, Gaussian noise, coherent communication.

I. INTRODUCTION

Given a noisy communication channel, let $M^*(n, \epsilon)$ be the maximal cardinality of a codebook of blocklength n which can be decoded with block error probability no greater than ϵ . The function $M^*(n, \epsilon)$ is the fundamental performance limit in the finite blocklength regime. For non-trivial channel models exact evaluation of $M^*(n, \epsilon)$ is computationally impossible. However, knowledge of Shannon capacity C of the channel enables the approximation¹

$$\log M^*(n, \epsilon) \approx nC \quad (1)$$

for asymptotically large blocklength. It has been shown in [1] that a much tighter approximation can be obtained by defining an additional figure of merit referred to as the channel dispersion:

Definition 1: The dispersion V (measured in squared information units per channel use) of a channel with capacity C is equal to

$$V = \lim_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \frac{(nC - \log M^*(n, \epsilon))^2}{2 \ln \frac{1}{\epsilon}}. \quad (2)$$

Channel capacity and dispersion become important analysis and design tools for systems with delay constraints; see [1] and [2, Chapter 5]. For example, the minimal blocklength required to achieve a given fraction η of capacity with a given error probability ϵ can be estimated as:²

$$n \gtrsim \left(\frac{Q^{-1}(\epsilon)}{1 - \eta} \right)^2 \frac{V}{C^2}. \quad (3)$$

The authors are with the Department of Electrical Engineering, Princeton University, Princeton, NJ, 08544 USA. e-mail: {ypolyans, verdu}@princeton.edu.

The research was supported by the National Science Foundation under grant CCF-10-16625 and by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-09-39370.

¹In this paper, unless explicitly stated all logarithms, log, and exponents, exp, are taken with respect to an arbitrary fixed base, which also determines the information units. Capacity and all rates are measured in information units per channel use.

²As usual, $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$.

The motivation for Definition 1 and estimate (3) is the following expansion for $n \rightarrow \infty$

$$\log M^*(n, \epsilon) = nC - \sqrt{nV} Q^{-1}(\epsilon) + O(\log n). \quad (4)$$

As shown in [1] in the context of memoryless channels, (4) gives an excellent approximation for blocklengths and error probabilities of practical interest.

This paper derives the dispersion of a single-input single-output (SISO), real-valued additive white Gaussian noise (AWGN) channel subject to stationary fading. The receiver is assumed to work in a coherent manner so that a perfect estimate of the fading coefficients is known to the decoder. Under such circumstances, it is well known that the channel capacity is independent of the fading dynamics [3]. On the contrary, we show that the dispersion exhibits an essentially linear behavior with the fading coherence time. In turn, the required blocklength (see (3)) is linear in the dispersion. We have observed [4] a similar effect for the Gilbert-Elliott channel, when the channel state is known at the decoder.

The paper is organized as follows. The channel model and the relevant literature are introduced in Section II. Section III presents a heuristic derivation of the dispersion. Rigorous results are the main content of Section IV. Section V illustrates the application of our results to a first-order auto-regressive Gaussian fading process.

Notation: Vectors and matrices are denoted by bold-face letters (e.g., \mathbf{x} and \mathbf{A}). Components of a random vector \mathbf{x} are denoted by capital letters X_1, X_2 , etc. The standard inner-product and the L_2 norm on \mathbb{R}^n are denoted as (\cdot, \cdot) and $\|\mathbf{x}\|^2 = (\mathbf{x}, \mathbf{x})$, respectively. Entry-wise k -th power of a vector \mathbf{x} is denoted as \mathbf{x}^k . Entry-wise (or Schur) product of two vectors \mathbf{h} and \mathbf{x} is denoted as $\mathbf{h} \odot \mathbf{x}$. The covariance function of a stationary process $\mathbf{X} = \{X_k, k = \dots, -1, 0, 1, \dots\}$ is

$$R_{\mathbf{X}}(k) = \mathbb{E}[(X_k - \mathbb{E}[X_k])(X_0 - \mathbb{E}[X_0])], \quad k \in \mathbb{Z} \quad (5)$$

from which the spectral function $F_{\mathbf{X}}$ is uniquely determined as

$$R_{\mathbf{X}}(k) = \int_{-\pi}^{\pi} \frac{1}{2\pi} e^{i\omega k} dF_{\mathbf{X}}(\omega), \quad F_{\mathbf{X}}(-\pi) = 0. \quad (6)$$

When $F_{\mathbf{X}}$ is absolutely continuous, its derivative is the spectral density $S_{\mathbf{X}}$ which (under certain conditions) can be found as

$$S_{\mathbf{X}}(\omega) = \sum_{k=-\infty}^{\infty} R_{\mathbf{X}}(k) e^{-i\omega k}. \quad (7)$$

If $S_{\mathbf{X}}$ exists and is continuous at zero, the long-term variance of \mathbf{X} is defined as

$$\mathbb{L}[\mathbf{X}] \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} \text{Var} \left[\sum_{i=1}^n X_i \right], \quad (8)$$

where the limit is guaranteed to exist and is equal to [5, Section 1.3]

$$\mathbb{L}[\mathbf{X}] = S_{\mathbf{X}}(0) \quad (9)$$

$$= R_{\mathbf{X}}(0) + 2 \sum_{k=1}^{\infty} R_{\mathbf{X}}(k). \quad (10)$$

Given two σ -algebras \mathcal{F} and \mathcal{G} we define the α -mixing coefficient as

$$\alpha(\mathcal{F}, \mathcal{G}) \triangleq \sup_{A \in \mathcal{F}, B \in \mathcal{G}} |\mathbb{P}[A, B] - \mathbb{P}[A]\mathbb{P}[B]|. \quad (11)$$

The sequence of α -mixing coefficients of a stationary process \mathbf{X} is

$$\alpha_{\mathbf{X}}(k) = \alpha(\sigma\{X_j, j \leq 0\}, \sigma\{X_j, j \geq k\}). \quad (12)$$

II. CHANNEL MODEL

In accordance with [6], a channel is a sequence of random transformations (transition probability kernels) parametrized by the blocklength n . The scalar, frequency-flat, coherent fading channel with SNR P is defined as follows. Consider a stationary and ergodic real-valued process $\mathbf{H} = \{H_i\}$ satisfying

$$\mathbb{E}[|H_i|^2] = 1. \quad (13)$$

For each blocklength $n \geq 1$ we have:

- the input space is a subset of \mathbb{R}^n satisfying the power constraint:

$$\|\mathbf{x}\|^2 \leq nP. \quad (14)$$

- the output space is $\mathbb{R}^n \times \mathbb{R}^n$ consisting of two vectors known to the receiver

$$\mathbf{h} = (H_1, \dots, H_n) \quad (15)$$

$$\mathbf{y} = (Y_1, \dots, Y_n). \quad (16)$$

- the input-output relation is given by

$$Y_i = H_i X_i + Z_i, \quad i = 1, \dots, n, \quad (17)$$

where Z_i are i.i.d. standard normal random variables $\sim \mathcal{N}(0, 1)$ independent of \mathbf{H} and \mathbf{x} .

The capacity of such a channel is given by [3, (3.3.10)]:

$$C(P) = \mathbb{E}[C(PH^2)], \quad (18)$$

where $C(P)$ is the capacity of the AWGN channel with SNR P :

$$C(P) = \frac{1}{2} \log(1 + P). \quad (19)$$

Traditionally, the dependence of the optimal coding rate on blocklength has been associated with the question of computing the channel reliability function. However, predictions on required blocklength obtained from error exponents may be far inferior compared to those obtained from (3) (e.g. [1, Table I]). Despite considerable efforts surveyed in [3, Section III.C7] the reliability function of the channel treated in this paper remains unknown even at rates near capacity. Error-exponent results are available for the block-fading channel [7, Section 3.4], [8] and

[9]. See also [10, Section 4] for discussion of key differences between block-fading and stationary fading.

Typically channel dispersion equals the reciprocal of the second derivative of the reliability function at capacity. Thus, in the absence of analytically tractable expressions in the realm of error exponents, the simplicity of the channel dispersion formula (27) is illuminating.

III. HEURISTIC DERIVATION

Before presenting the result we motivate it with a simple heuristic. Let us replace a stationary process $\{H_i\}$ via a block-stationary process $\{\hat{H}_i\}$ with block size T . In other words, $(\hat{H}_1, \dots, \hat{H}_T)$ are distributed as (H_1, \dots, H_T) and different T -blocks of $\{\hat{H}_i\}$ are independent. The key (and, at this point, unjustified) assumption is that the resulting channel dispersion converges to the sought-after one as $T \rightarrow \infty$.

Considering $\mathbf{x} = (X_1, \dots, X_T)$, $\mathbf{y} = (Y_1, \dots, Y_T)$ and $\hat{\mathbf{h}} = (\hat{H}_1, \dots, \hat{H}_T)$ as single super-letters, the channel model becomes memoryless:

$$\mathbf{y}_k = \hat{\mathbf{h}}_k \odot \mathbf{x}_k + \mathbf{z}_k, \quad k = 1, \dots, n, \quad (20)$$

with n equal to the coding blocklength, $\mathbf{z}_k \sim \mathcal{N}(0, \mathbf{I}_T)$, and inputs subject to a power constraint

$$\sum_{k=1}^n \|\mathbf{x}_k\|^2 \leq nTP. \quad (21)$$

By [7] the capacity of such a channel is $TC(P)$ (per T -block) achieved by taking $\mathbf{x} \sim \mathcal{N}(0, P\mathbf{I}_T)$.

For both the AWGN and the DMC with cost constraints (see [1, Section IV.B] and [2, Section 3.4.6], resp.), the channel dispersion is given by

$$V = \text{Var}[\iota_{X;Y}(X; Y)|X], \quad (22)$$

where X is distributed according to the capacity achieving distribution and

$$\iota_{X;Y}(a; b) = \log \frac{dP_{Y|X}}{dP_Y}(b|a). \quad (23)$$

Thus, the extension of (22) to the channel model (20) merely involves replacing Y with $(\mathbf{y}, \hat{\mathbf{H}})$ and taking $X \sim \mathcal{N}(0, P\mathbf{I}_T)$. Doing so one gets for $\iota_{X;Y}$ the following

$$\frac{1}{2} \sum_{i=1}^T \log(1 + P\hat{H}_i^2) + \frac{\hat{H}_i^2 X_i^2 + 2\hat{H}_i X_i Z_i - P\hat{H}_i^2 Z_i^2}{1 + P\hat{H}_i^2} \log e \quad (24)$$

The expectation of (24) equals, of course, $TC(P)$, while for the conditional variance of (24) we find

$$V_T = \text{Var} \left[\sum_{i=1}^T \frac{1}{2} \log(1 + P\hat{H}_i^2) \right] + T \frac{\log^2 e}{2} \left(1 - \mathbb{E}^2 \left[\frac{1}{1 + P\hat{H}_i^2} \right] \right). \quad (25)$$

Thus, the limiting dispersion (per channel use) is

$$\begin{aligned} \mathbb{V}(P) &= \lim_{T \rightarrow \infty} \frac{V_T}{T} \\ &= \mathbb{L} [C(PH^2)] + \frac{\log^2 e}{2} \left(1 - \mathbb{E}^2 \left[\frac{1}{1 + PH^2} \right] \right) \end{aligned} \quad (26)$$

$$(27)$$

It is interesting to contrast (27) with the result for the *discrete* additive-noise channel, whose instantaneous noise distribution is governed by a stationary and ergodic state process $\mathbf{S} = \{S_j, j \in \mathbb{Z}\}$. The capacity-dispersion pair can be shown to satisfy

$$\bar{C} = \mathbb{E} [C(S_0)], \quad (28)$$

$$\bar{V} = \mathbb{L} [C(S)] + \mathbb{E} [V(S_0)], \quad (29)$$

where $C(s)$ and $V(s)$ are the capacity and the dispersion of the DMC corresponding to the state s ; see [4] and also [11] for the case of a memoryless state process.

While (28) is the counterpart of (18), (29) is not the counterpart of (27). In fact, (27) can be written as

$$\mathbb{V}(P) = \mathbb{L} [C(PH^2)] + \mathbb{E} [V(PH^2)] + \frac{\log^2 e}{2} \text{Var} \left[\frac{1}{1 + PH^2} \right] \quad (30)$$

where $V(P)$ is the dispersion of the AWGN channel with SNR P [1, (293)]:

$$V(P) = \frac{\log^2 e}{2} \left(1 - \left(\frac{1}{1 + P} \right)^2 \right). \quad (31)$$

Comparing (29) and (30) we see that in both cases the dynamics of the fading (or state) process affects the dispersion through the spectral density at zero of the corresponding capacity process. However, from (30) we see that the cost constraint introduces an additional dynamics-independent term in the expression for dispersion.

IV. MAIN RESULT

Theorem 1: Assume that the stationary process $\mathbf{H} = \{H_i, i \in \mathbb{Z}\}$ satisfies the following assumptions:

- 1) Condition (13) on the second moment holds.
- 2) The α -mixing coefficients (12) satisfy for some $r < 1$:

$$\sum_{k=1}^{\infty} k(\alpha_{\mathbf{H}}(k))^r < \infty. \quad (32)$$

- 3) For all $j > 1$ we have

$$\mathbb{P}[H_j H_0 \neq 0] > 0. \quad (33)$$

Then the dispersion $\mathbb{V}(P)$ of the coherent fading channel in Section II is given by (27). Furthermore, for any $0 < \epsilon < 1/2$ we have as $n \rightarrow \infty$

$$\log M^*(n, \epsilon) = nC(P) - \sqrt{n\mathbb{V}(P)}Q^{-1}(\epsilon) + o(\sqrt{n}), \quad (34)$$

regardless of whether ϵ is a maximal or average probability of error.

The proof is outlined in the Appendix.

The assumptions of Theorem 1 are not as restrictive and hard to verify as they may seem at first sight. Assumption 2, which implies ergodicity of \mathbf{H} , automatically holds for processes with finite memory (such as finite-order moving averages) and can usually be verified easily for other finite-order Markov processes. If \mathbf{H} is Gaussian, then the α -mixing coefficients can be tightly estimated from the spectral density of \mathbf{H} . In particular, if $S_{\mathbf{H}}(\omega)$ is a rational function of $e^{i\omega}$ then $\alpha_{\mathbf{H}}(k)$ decay exponentially; see Section V for more.

Assumption 2 also ensures that the first term in (27) makes sense. Indeed, although $\{H_i, i \in \mathbb{Z}\}$ possessing a spectral density implies that $\{\log(1 + PH_i^2), i \in \mathbb{Z}\}$ also has one [12], the continuity of the latter is not guaranteed. However, assumptions 1, 2 and [5, Lemma 1.3] imply continuity. Assumption 3 is necessary and sufficient for the uniqueness of the maximizer in

$$\max_{X^n} I(X^n; Y^n | H^n) \quad (35)$$

as can be seen from the argument in [7, Section 3.2]. Note that Assumption 3 is automatically satisfied if the distribution of H_0 has no atom at zero. Assumption 3 is independent of the other ones (e.g., let \mathbf{H} be an ergodic Markov chain with two states, $H = 0$ and $H = 1$, which transitions from 1 to 0 with probability 1). Although a mild requirement, we believe Theorem 1 still holds without Assumption 3.

Similarly, for the complex AWGN we have:

Theorem 2: For the complex AWGN with complex-valued fading process $\{H_i, i = 1, \dots\}$, in the assumptions of Theorem 1 the dispersion is given by

$$\begin{aligned} \mathbb{V}_c(P) &= \\ &\mathbb{L} [\log(1 + P|H|^2)] + \log^2 e \cdot \left(1 - \mathbb{E}^2 \left[\frac{1}{1 + P|H|^2} \right] \right) \end{aligned} \quad (36)$$

V. GAUSS-MARKOV FADING

We now proceed to investigate the behavior of the dispersion (27) with respect to the spectrum of the fading process. Before doing so, however, we need to check condition (32). To simplify the computation of the α -mixing coefficients, we first observe that

$$\alpha_{\mathbf{X}}(k) \leq \sup_{f,g} \mathbb{E} [f(X_0, X_{-1}, \dots)g(X_k, X_{k+1}, \dots)], \quad (37)$$

where the functions f and g are zero-mean and unit variance. The quantity on the right side of (37) is known as a ρ -mixing coefficient, which for Gaussian processes is easy to compute thanks to a beautiful observation of Sarmanov [13]. In particular, [14] gives an explicit formula and shows that for any Gaussian process whose spectral function $S_{\mathbf{X}}$ is rational in $e^{i\omega}$, the ρ -mixing coefficients decay exponentially. In view of (37), this automatically guarantees that any process obtained via finite-order auto-regressive moving average (ARMA) of a white Gaussian noise satisfies (32).³

³Moreover, for Gaussian processes, (37) is tight up to a universal constant factor [14, Theorem 2]. Hence, Gaussian processes with non-absolutely continuous spectral functions, must have $\alpha_{\mathbf{X}}(k) \geq \epsilon > 0$ for all k . Consequently, Theorem 1 is not applicable to such fading scenarios.

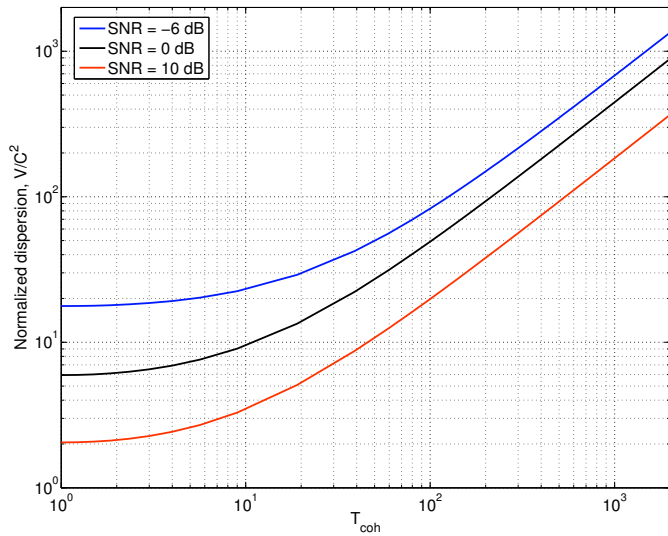


Fig. 1. Normalized dispersion of a coherent fading channel with Gauss-Markov fading process (38) as a function of coherence time.

For the purpose of illustration we consider the Gauss-Markov, or $AR(1)$, fading process:

$$H_i = aH_{i-1} + W_i, \quad W_i \sim \mathcal{N}(0, 1 - a^2), \quad (38)$$

where $0 \leq a < 1$. The spectral density is

$$S_{\mathbf{H}}(\omega) = \frac{1 - a^2}{1 + a^2 - 2a \cos \omega}, \quad (39)$$

whereas the coherence time is defined as

$$T_{coh} \triangleq \frac{\max_{\omega} S_{\mathbf{H}}(\omega)}{\frac{1}{2\pi} \int_{-\pi}^{\pi} S_{\mathbf{H}}(\omega) d\omega} = \frac{1 + a}{1 - a}. \quad (40)$$

Therefore, for memoryless fading $T_{coh} = 1$.

Note that $\alpha_{\mathbf{H}}(k)$ are easy to estimate since by the Markov property:

$$\alpha_{\mathbf{H}}(k) = \alpha(\sigma\{H_0\}, \sigma\{H_k\}) \quad (41)$$

and by (37) and [13] we get

$$\alpha_{\mathbf{H}}(k) \leq a^k. \quad (42)$$

This helps in the computation of $\mathbb{L}[C(PH^2)]$ since a firm exponentially decaying bound on the tail of the series in (10) can be given via [5, Lemma 1.3]. which allows for termination of the series (10) with a sharp estimate of precision. The second term in (27) is easily computed numerically.

The dependence of the dispersion on coherence time under the Gauss-Markov model is illustrated in Fig. 1. In view of (3) we plot the normalized dispersion $\frac{V}{C^2}$, where $C = 0.1403, 0.3848$ and 1.2527 bits/ch.use for $SNR = -6$ dB, 0 dB and 10 dB, respectively. Thus, for example, when $T_{coh} = 10^2$ achieving 90% of the capacity with block error rate 10^{-3} requires codes of length 80000, 50000 and 20000 for SNR of -6 dB, 0 dB and 10 dB, respectively. We notice that the required blocklength is approximately proportional to T_{coh} with the coefficient of proportionality dependent on the SNR . However, unlike the ad-hoc definition of coherence time

in (40), the notion of dispersion and the estimate in (3) are fully theoretically justified by Theorem 1.

REFERENCES

- [1] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [2] Y. Polyanskiy, "Channel coding: non-asymptotic fundamental limits," Ph.D. dissertation, Princeton Univ., Princeton, NJ, USA, 2010, available: <http://www.princeton.edu/~ypolyans>.
- [3] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communication aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.
- [4] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Dispersion of the Gilbert-Elliott channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1829–1848, Apr. 2011.
- [5] I. A. Ibragimov, "Some limit theorems for stationary processes," *Theor. Probability Appl.*, vol. 7, no. 4, 1962.
- [6] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, 1994.
- [7] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecom.*, vol. 10, no. 6, pp. 585–595, 1999.
- [8] X. Wu and R. Srikant, "Asymptotic behavior of error exponents in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1310–1325, 2007.
- [9] H. Shin and M. Win, "Gallager's exponent for MIMO channels: a reliability-rate tradeoff," *IEEE Trans. Comm.*, vol. 57, no. 4, pp. 972–985, 2009.
- [10] A. Lapidoth, "On the high SNR capacity of stationary Gaussian fading channels," in *Proc. 2003 41st Allerton Conference*, vol. 41, no. 1, Allerton Retreat Center, Monticello, IL, USA, 2003, pp. 410–419.
- [11] A. Ingber and M. Feder, "Finite blocklength coding for channels with side information at the receiver," in *Proc. 2010 IEEE 26-th Convention Of Electr. and Electron. Eng.*, Israel, Dec. 2010.
- [12] J. Blum and B. Eisenberg, "Spectral properties of processes derived from stationary Gaussian sequences," *Stoch. Proc. Appl.*, vol. 2, no. 2, pp. 177–181, 1974.
- [13] O. V. Sarmanov, "A maximal correlation coefficient," *Dokl. Akad. Nauk SSSR*, vol. 121, no. 1, 1958.
- [14] A. N. Kolmogorov and Y. A. Rozanov, "On strong mixing conditions for stationary Gaussian processes," *Theor. Probability Appl.*, vol. 5, no. 2, pp. 204–208, 1960, (English transl.).
- [15] V. Borkar, S. Mitter, and S. Venkatesh, "Variations on a theme by Neyman and Pearson," *Sankhya*, vol. 66, no. 2, pp. 292–305, 2004.

APPENDIX PROOF OF THEOREM 1

Due to space limitations, we rely heavily on the notation and results of [1]. In particular, we assume familiarity with the definition of $\beta_{\alpha}(P, Q)$ in [1, (100)] and $\kappa_{\tau}(\mathbf{F}, Q_Y)$ in [1, (107)].

Achievability: Fix blocklength n and select the auxiliary output distribution

$$Q_{Y^n H^n}(y^n, h^n) = P_{H^n}(h^n) \prod_{j=1}^n \mathcal{N}(0, 1 + Ph_j^2). \quad (43)$$

We denote

$$\beta_{\alpha}^n(\mathbf{x}) \triangleq \beta_{\alpha}(P_{Y^n H^n | X^n = \mathbf{x}}, Q_{Y^n H^n}). \quad (44)$$

Henceforth \mathbf{x} is assumed to belong to the power sphere

$$\|\mathbf{x}\|^2 = nP. \quad (45)$$

To analyze the asymptotic behavior of $\beta_\alpha^n(\mathbf{x})$ we note that $\log \frac{P_{Y^n H^n | \mathbf{x}^n = \mathbf{x}}}{Q_{Y^n H^n}}$ under P is distributed as

$$\frac{\log e}{2} \sum_{i=1}^n \ln(1 + PH_i^2) + \frac{H_i^2 x_i^2 + 2H_i Z_i x_i - PH_i^2 Z_i^2}{1 + PH_i^2} \quad (46)$$

The expectation of (46) is equal to nC and variance to $nV(\mathbf{x})$

$$V(\mathbf{x}) = V_0 + V_1(\mathbf{x}) + V_2(\mathbf{x}) + v_n \quad (47)$$

$$V_0 = \mathbb{E}[V(PH_1^2)] + \mathbb{L}[C(PH^2)] \quad (48)$$

$$v_n = \frac{1}{n} \text{Var} \left[\sum_{i=1}^n C(PH_i^2) \right] - \mathbb{L}[C(PH^2)] \quad (49)$$

$$V_1(\mathbf{x}) = \frac{\log^2 e}{2P} \frac{1}{n} (\mathbf{d}, P\mathbf{1} - \mathbf{x}^2) \quad (50)$$

$$V_2(\mathbf{x}) = \frac{\log^2 e}{4P^2} \frac{1}{n} (\mathbf{A}_n(\mathbf{x}^2 - P\mathbf{1}), \mathbf{x}^2 - P\mathbf{1}) \quad (51)$$

$$(\mathbf{d})_j = \mathbb{E} \left[\frac{\sum_{i=1}^n \log(1 + PH_i^2)}{1 + PH_j^2} \right] \quad (52)$$

where $\mathbf{1}$ is an n -vector of all ones and \mathbf{A}_n is the $n \times n$ covariance matrix of $\left\{ \frac{1}{1+PH_i^2}, i = 1, \dots, n \right\}$. As in [1, Section III.J2], a central-limit theorem analysis of (46) implies

$$\log \beta_\alpha^n(\mathbf{x}) = -nC - \sqrt{nV(\mathbf{x})} Q^{-1}(\alpha) + o(\sqrt{n}) \quad (53)$$

Although as \mathbf{x} goes around the power sphere $V(\mathbf{x})$ experiences quite significant variations, for the most part it is very close to V in (27):

Lemma 3: For each n let \mathbf{x} be distributed uniformly on the power sphere (45). Then for each $\delta > 0$ we have as $n \rightarrow \infty$

$$\mathbb{P}[|V(\mathbf{x}) - V| > \delta] \rightarrow 0. \quad (54)$$

We now fix δ and denote

$$\mathcal{F} = \{\mathbf{x} : \|\mathbf{x}\|^2 = nP, V(\mathbf{x}) < V + \delta\}. \quad (55)$$

Given $Q_{Y^n H^n}$ and \mathcal{F} we define $\kappa_\tau(\mathcal{F}, Q_{Y^n H^n})$ as in [1, (107)]. The following is a simple lower bound for κ_τ :

Lemma 4: For any distribution P_X we have

$$\kappa_\tau(\mathcal{F}, Q_Y) \geq \beta_{\tau P_X[\mathcal{F}]}(P_Y, Q_Y), \quad (56)$$

where P_Y is the distribution induced by P_X over the channel $P_{Y|X}$.

As simple as it is, under regularity conditions this lower bound becomes tight upon taking the supremum over P_X [15].

For our purposes we select P_X to be the uniform distribution on the power-sphere (45). By Lemma 3 for all n sufficiently large we have $P_X[\mathcal{F}] > \frac{1}{2}$ and therefore from (56) we get for some constant K_1 :

$$\kappa_\tau(\mathcal{F}, Q_Y) \geq \beta_{\frac{\tau}{2}}(P_{Y^n H^n}, Q_{Y^n H^n}) \quad (57)$$

$$\geq \exp \left\{ -\frac{D(P_{Y^n H^n} \| Q_{Y^n H^n}) + \log 2}{\tau/2} \right\} \quad (58)$$

$$\geq \exp \left\{ -\frac{K_1}{\tau} \right\} \quad (59)$$

where in (57) $P_{Y^n H^n}$ is the distribution induced on the output of the fading channel by \mathbf{x} uniform on the power sphere, (58) follows from the data-processing inequality for divergence:

$$d(\beta_\alpha(P, Q) \| \alpha) \leq D(P \| Q), \quad (60)$$

and (59) from a computation $D(P_{Y^n H^n} \| Q_{Y^n H^n}) = O(1)$ as $n \rightarrow \infty$. Therefore, by the $\kappa\beta$ bound [1, Theorem 25] for each $\tau > 0$ there exists an (n, M, ϵ) code with

$$M \geq \frac{\kappa_\tau(\mathcal{F}, Q_{Y^n H^n})}{\sup_{\mathbf{x} \in \mathcal{F}} \beta_{1-\epsilon+\tau}^n(\mathbf{x})}. \quad (61)$$

From (61) via (53) and (59) we get

$$\log M^*(n, \epsilon) \geq -\frac{K_1}{\tau} + nC - \sqrt{n(V + \delta)} Q^{-1}(\epsilon - \tau) + o(\sqrt{n}). \quad (62)$$

Since τ and δ are arbitrary we conclude that the lower-bound in (34) is established.

Converse: Given a sequence of (n, M_n, ϵ) codes (average probability of error) we first notice that without loss of generality the encoder can be assumed deterministic. Next, as in the proof of [1, (286)] we reduce to the case of *maximal* probability of error. Furthermore, as in [1, Lemma 39] we reduce to the case when all of the codewords belong to the power sphere (45). Thus by the meta-converse [1, Theorem 30] with an auxiliary channel chosen as in (43) we have

$$\log M_n \leq -\inf_{\mathbf{x}} \log \beta_{1-\epsilon}^n(\mathbf{x}), \quad (63)$$

where the infimum is over all the codewords. If we extend the infimum to the entire power-sphere then in view of (53) we obtain for $\epsilon < 1/2$:

$$\log M_n \leq nC - \sqrt{n \inf_{\mathbf{x}} V(\mathbf{x})} Q^{-1}(\epsilon) + o(\sqrt{n}). \quad (64)$$

Note that due to (32) the vector \mathbf{d} in (50) is almost parallel to $\mathbf{1}$ and from (45) we have $(\mathbf{1}, \mathbf{x}^2 - P) = 0$. This shows that $V_1(\mathbf{x}) = o(1)$. Since $V_2(\mathbf{x}) \geq 0$ we obtain the upper bound

$$\log M_n \leq nC - \sqrt{nV_0} Q^{-1}(\epsilon) + o(\sqrt{n}). \quad (65)$$

Note that V_0 accounts for the first two terms in (30). Since the third term can be at most $\frac{\log^2 e}{8}$, (65) already gives a very good bound on the dispersion term. To get a tighter bound and conclude the proof of (34) we need to show that any capacity-achieving sequence of codes contains large subcodes with codewords fully on the set where $V(\mathbf{x}) \approx V$. Intuitively, this is true since by Lemma 3 only a tiny portion of the power sphere yields atypical values for $V(\mathbf{x})$. A rigorous and technical proof of this fact (omitted for space limitations) uses Assumption 3.