

The Third-Order Term in the Normal Approximation for the AWGN Channel

Vincent Y. F. Tan[†] and Marco Tomamichel^{*}

Abstract

This paper shows that, under the average error probability formalism, the third-order term in the normal approximation for the additive white Gaussian noise channel with a maximal or equal power constraint is at least $\frac{1}{2} \log n + O(1)$. This matches the upper bound derived by Polyanskiy-Poor-Verdú (2010).

I. INTRODUCTION

The most important continuous alphabet channel in communication systems is the discrete-time additive white Gaussian noise (AWGN) channel in which at each time i , the output of the channel Y_i is the sum of the input X_i and Gaussian noise Z_i . Shannon showed in his original paper [1] that launched the field of information theory that the capacity of the AWGN channel is

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

where P is the signal-to-noise ratio (SNR). More precisely, let $M^*(W^n, \varepsilon, P)$ be the maximum number of codewords that can be transmitted over n independent uses of an AWGN channel with SNR P and average error probability not exceeding $\varepsilon \in (0, 1)$. Then, combining the direct part in [1] and the strong converse by Shannon in [2] (also see Yoshihara [3] and Wolfowitz [4]), one sees that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log M^*(W^n, \varepsilon, P) = C(P) \quad \text{bits per channel use} \quad (2)$$

holds for every $\varepsilon \in (0, 1)$.

Recently, there has been significant renewed interest in studying the higher-order terms in the asymptotic expansion of non-asymptotic fundamental limits such as $\log M^*(W^n, \varepsilon, P)$. This line of analysis was pioneered by Strassen [5, Theorem 1.2] for discrete memoryless channels (DMCs) and is useful because it provides key insights into the amount of backoff from channel capacity for block codes of finite length n . For the AWGN channel, Hayashi [6, Theorem 5] showed that

$$\log M^*(W^n, \varepsilon, P) = nC(P) + \sqrt{nV(P)}\Phi^{-1}(\varepsilon) + o(\sqrt{n}) \quad (3)$$

where $\Phi^{-1}(\cdot)$ is the inverse of the Gaussian cumulative distribution function and

$$V(P) = \log^2 e \cdot \frac{P(P+2)}{2(P+1)^2} \quad \text{bits}^2 \text{ per channel use} \quad (4)$$

is termed the *Gaussian dispersion function* [7]. The first two terms in the expansion in (3) are collectively known the *normal approximation*. The functional form of $V(P)$ was already known to Shannon [2, Section X] who analyzed the behavior of the reliability function of the AWGN channel at rates close to capacity. Subsequently, the $o(\sqrt{n})$ remainder term in the expansion in (3) was refined by Polyanskiy-Poor-Verdú [7, Theorem 54, Eq. (294)] who showed that

$$O(1) \leq \log M^*(W^n, \varepsilon, P) - \left(nC(P) + \sqrt{nV(P)}\Phi^{-1}(\varepsilon) \right) \leq \frac{1}{2} \log n + O(1). \quad (5)$$

The same bounds hold under the maximum probability of error formalism.

[†] Department of Electrical and Computer Engineering (ECE), National University of Singapore (NUS) and Institute for Infocomm Research (I²R), Agency for Science, Technology and Research (A*STAR) (Email: vtan@nus.edu.sg)

^{*} Center for Quantum Technologies, National University of Singapore (Email: cqtmarco@nus.edu.sg)

Despite these impressive advances in the fundamental limits of coding over a Gaussian channel, the gap in the third-order term beyond the normal approximation in (5) calls for further investigations. The authors of the present paper showed for DMCs with positive ε -dispersion that the third-order term is no larger than $\frac{1}{2} \log n + O(1)$ [8, Theorem 1], matching a lower bound by Polyanskiy [9, Theorem 53] for non-singular channels (also called channels with positive reverse dispersion [9, Eq. (3.296)]). Altuğ and Wagner [10] showed for singular, symmetric DMCs that the third-order term is $O(1)$. Moulin [11] recently showed for a large class of channels (but *not* the AWGN channel) that the third-order term is $\frac{1}{2} \log n + O(1)$. In light of these existing results for DMCs, a reasonable conjecture would be that the third-order term for the Gaussian case is either $O(1)$ or $\frac{1}{2} \log n + O(1)$. In this paper, we show that in fact, the lower bound in (5) is loose. In particular, we establish that it can be improved to match the upper bound $\frac{1}{2} \log n + O(1)$. Our proof technique is similar to that developed by Polyanskiy [9, Theorem 53] to show that $\frac{1}{2} \log n + O(1)$ is achievable for non-singular DMCs. However, our proof is more involved due to the presence of power constraints on the codewords.

II. PROBLEM SETUP AND DEFINITIONS

Let W be an AWGN channel where the noise variance¹ is 1, i.e.

$$W(y|x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y-x)^2}{2}\right). \quad (6)$$

Let $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$ be two vectors in \mathbb{R}^n . Let $W^n(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n W(y_i|x_i)$ be the n -fold memoryless extension of W . An $(n, M, \varepsilon, P)_{\text{av}}$ -code for the AWGN channel W is a system $\{(\mathbf{x}(m), \mathcal{D}_m)\}_{m=1}^M$ where $\mathbf{x}(m) \in \mathbb{R}^n, m \in \{1, \dots, M\}$, are the codewords satisfying the maximal power constraint $\|\mathbf{x}(m)\|_2^2 \leq nP$, the sets $\mathcal{D}_m \subset \mathbb{R}^n$ are disjoint decoding regions and the *average probability of error* does not exceed ε , i.e.

$$\frac{1}{M} \sum_{m=1}^M W^n(\mathcal{D}_m^c | \mathbf{x}(m)) \leq \varepsilon. \quad (7)$$

Define $M^*(W^n, \varepsilon, P) := \max \{M \in \mathbb{N} : \exists \text{ an } (n, M, \varepsilon, P)_{\text{av}}\text{-code for } W\}$.

We also employ the Gaussian cumulative distribution function

$$\Phi(a) := \int_{-\infty}^a \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du \quad (8)$$

and define its inverse as $\Phi^{-1}(\varepsilon) := \sup\{a \in \mathbb{R} : \Phi(a) \leq \varepsilon\}$, which evaluates to the usual inverse for $0 < \varepsilon < 1$ and continuously extends to take values $\pm\infty$ outside that range.

III. MAIN RESULT AND REMARKS

Let us reiterate our main result.

Theorem 1. *For all $0 < \varepsilon < 1$ and $P \in (0, \infty)$,*

$$\log M^*(W^n, \varepsilon, P) \geq nC(P) + \sqrt{nV(P)}\Phi^{-1}(\varepsilon) + \frac{1}{2} \log n + O(1) \quad (9)$$

where $C(P)$ and $V(P)$ are the Gaussian capacity and dispersion functions respectively.

We make the following remarks before proving the theorem in the following section.

- 1) As mentioned in the Introduction, the upper bound on $\log M^*(W^n, \varepsilon, P)$ in (5) was first established by Polyanskiy-Poor-Verdú [7, Theorem 65]. They evaluated the meta-converse [7, Theorem 28] and appealed to the spherical symmetry in the Gaussian problem. The third-order term in the normal approximation was shown to be upper bounded by $\frac{1}{2} \log n + O(1)$ (under the average or maximum error probability formalism). Thus, one has

$$\log M^*(W^n, \varepsilon, P) = nC(P) + \sqrt{nV(P)}\Phi^{-1}(\varepsilon) + \frac{1}{2} \log n + O(1). \quad (10)$$

¹The assumption that the noise variance is 1 does not entail any loss of generality because we can simply scale the admissible power accordingly to ensure that the SNR is P .

The technique developed by the present authors in [8] can also be used to prove the $\frac{1}{2} \log n + O(1)$ upper bound on the third-order term.

- 2) Our strategy for proving (9) parallels that for non-singular DMCs without cost constraints by Polyanskiy [9, Theorem 53]. It leverages on the random-coding union (RCU) bound [7, Theorem 16] and uses the log-likelihood ratio as the decoding metric, i.e. we do maximum likelihood decoding. However, the Gaussian problem involves cost (power) constraints and our random codebook generation strategy (which is similar to Shannon's [2]) involves drawing codewords independently and uniformly at random from the power sphere. Thus, a more delicate analysis (vis-à-vis [9, Theorem 53]) is required. In particular, one cannot directly employ the refined large-deviations result stated in [7, Lemma 47] which is crucial in showing the achievability of $\frac{1}{2} \log n + O(1)$. This is because [7, Lemma 47] requires independence of a collection random variables whereas the independence structure is lacking in the AWGN problem.
- 3) In Theorem 1, we considered a maximal power constraint on the codewords, i.e. $\|\mathbf{x}(m)\|_2^2 \leq nP$ for all m . It is easy to show that the third-order term is the same for the case of equal power constraints, i.e. $\|\mathbf{x}(m)\|_2^2 = nP$ for all m . However, the strong converse does not even hold [9, Theorem 77] under the *average probability of error* formalism and the *average power constraint across the codebook*, i.e. $\frac{1}{M} \sum_{m=1}^M \|\mathbf{x}(m)\|_2^2 \leq nP$. The ε -capacity depends on ε . We do not consider this case in this paper. Nonetheless, the strong converse and normal approximation do hold [7, Theorem 54] under the *maximum probability of error* formalism and average power constraint across the codebook but we do not consider this setup here. It is known [7, Eq. (295)] that the third-order term is sandwiched between $O(1)$ and $\frac{3}{2} \log n + O(1)$.
- 4) A straightforward extension of our proof technique (in particular, the application of Lemma 2 in Section IV-E) shows that the achievability of $\frac{1}{2} \log n + O(1)$ also holds for the problem of information transmission over *parallel Gaussian channels* [12, Section 9.4] in which the capacity is given by the well-known *water-filling* solution. See Appendix A for a description of the modifications to the proof of Theorem 1 to this setting. This improves on the result in [9, Theorem 81] by $\frac{1}{2} \log n$. However, this third-order achievability result does not match the converse bound given in [9, Theorem 80] in which it is shown that the third-order term is upper bounded by $\frac{k+1}{2} \log n + O(1)$ where $k \geq 1$ is the number of parallel Gaussian channels. We leave the closing of this gap for future research.
- 5) Finally, we make an observation concerning the relation between prefactors in the error exponents regime and the third-order terms in the normal approximation. In [2], Shannon derived exponential bounds on the average error probability of optimal codes over a Gaussian channel using geometric arguments. For *high rates* (i.e. rates above the critical rate and below capacity), he showed that [2, Eqs. (4)–(5)]

$$P_e^*(M, n) = \Theta\left(\frac{\exp(-nF(\varphi))}{\sqrt{n}}\right) \quad (11)$$

where $P_e^*(M, n)$ is the optimal average probability of error of a length- n block code of size $M \in \mathbb{N}$, $\varphi = \varphi(R)$ is a cone angle related to the signaling rate $R := \frac{1}{n} \log M$ as follows [2, Eq. (28)]

$$\exp(-nR) = \frac{(1 + O(\frac{1}{n})) \sin^n \varphi}{\sqrt{2\pi n} \sin \varphi \cos \varphi}, \quad (12)$$

and the exponent in (11) is defined as

$$F(\varphi) := \frac{P}{2} - \frac{\sqrt{P} G \cos \varphi}{2} - \log(G \sin \varphi), \quad \text{where} \quad (13)$$

$$G = G(\varphi) := \frac{1}{2}(\sqrt{P} \cos \varphi + \sqrt{P \cos^2 \varphi + 4}). \quad (14)$$

Furthermore for high rates, the error exponent (reliability function) of an AWGN channel is known and equals the sphere-packing exponent [13, Eq. (7.4.33)]

$$E(R) = \frac{P}{4\beta} \left((\beta + 1) - (\beta - 1) \sqrt{1 + \frac{4\beta}{P(\beta - 1)}} \right) + \frac{1}{2} \log \left(\beta - \frac{P(\beta - 1)}{2} \left[\sqrt{1 + \frac{4\beta}{P(\beta - 1)}} - 1 \right] \right) \quad (15)$$

where $\beta := \exp(2R)$. Simple algebra shows that $F(\theta) = E(\tilde{R}(\theta))$ when $\tilde{R}(\theta) := -\log \sin \theta$. Thus,

$$F(\varphi(R)) = E(\tilde{R}(\varphi(R))) \quad (16)$$

$$= E(-\log \sin(\varphi(R))) \quad (17)$$

$$= E\left(R - \frac{\log n}{2n} + \Theta\left(\frac{1}{n}\right)\right) \quad (18)$$

$$= E(R) - E'(R) \frac{\log n}{2n} + \Theta\left(\frac{1}{n}\right), \quad (19)$$

where (18) follows from (12) and (19) follows by Taylor expanding the continuously differentiable function $E(R)$. Note that $E'(R) \leq 0$. This leads to the conclusion that for high rates,

$$P_e^*(M, n) = \Theta\left(\frac{\exp(-nE(R))}{n^{(1+|E'(R)|)/2}}\right). \quad (20)$$

Thus, the prefactor of the AWGN channel is $\Theta(n^{-(1+|E'(R)|)/2})$. We showed in Theorem 1 that the third-order term is $\frac{1}{2} \log n + O(1)$. Somewhat surprisingly, this is analogous to the symmetric, discrete memoryless case. Indeed for non-singular, symmetric DMCs (such as the binary symmetric channel) the prefactor in the error exponents regime for high rates is $\Theta(n^{-(1+|E'(R)|)/2})$ [14]–[17] and for DMCs with positive ε -dispersion, the third-order term is $\frac{1}{2} \log n + O(1)$ (combining [8, Theorem 1] and [9, Theorem 53]). (Actually symmetry is not required for the third-order term to be $\frac{1}{2} \log n + O(1)$.) On the other hand, for singular, symmetric DMCs (such as the binary erasure channel), the prefactor is $\Theta(n^{-1/2})$ [14]–[17] and the third-order term is $O(1)$ (combining [10, Proposition 1] and [7, Theorem 45]). Also see [18, Theorem 23]. These results suggest a connection between prefactors and third-order terms. Indeed, a precise understanding of this connection is a promising avenue for further research.

IV. PROOF OF THEOREM 1

The proof, which is based on random coding, is split into several steps.

A. Random Codebook Generation And Encoding

We first start by defining the random coding distribution

$$f_{\mathbf{X}}(\mathbf{x}) := \frac{\delta(\|\mathbf{x}\|_2^2 - nP)}{S_n(\sqrt{nP})} \quad (21)$$

where $\delta(\cdot)$ is the Dirac delta and $S_n(r) = \frac{2\pi^{n/2}}{\Gamma(n/2)} r^{n-1}$ is the surface area of a radius- r sphere in \mathbb{R}^n . We sample M length- n codewords independently from $f_{\mathbf{X}}$. In other words, we draw codewords uniformly at random from the surface of the sphere in \mathbb{R}^n with radius \sqrt{nP} . The number of codewords M will be specified at the end of the proof in (81). These codewords are denoted as $\mathbf{x}(m) = (x_1(m), \dots, x_n(m))$, $m \in \{1, \dots, M\}$. To send message m , transmit codeword $\mathbf{x}(m)$.

B. Maximum-Likelihood Decoding

Let the induced output density be $f_{\mathbf{X}}W^n$, i.e.

$$f_{\mathbf{X}}W^n(\mathbf{y}) := \int_{\mathbf{x}'} f_{\mathbf{X}}(\mathbf{x}') W^n(\mathbf{y}|\mathbf{x}') d\mathbf{x}'. \quad (22)$$

Given $\mathbf{y} = (y_1, \dots, y_n)$, the decoder selects the message m satisfying

$$q(\mathbf{x}(m), \mathbf{y}) > \max_{\tilde{m} \in \{1, \dots, M\} \setminus \{m\}} q(\mathbf{x}(\tilde{m}), \mathbf{y}), \quad (23)$$

where the decoding metric is the log-likelihood ratio defined as

$$q(\mathbf{x}, \mathbf{y}) := \log \frac{W^n(\mathbf{y}|\mathbf{x})}{f_{\mathbf{X}}W^n(\mathbf{y})}. \quad (24)$$

If there is no unique $m \in \{1, \dots, M\}$ satisfying (23), declare an error. (This happens with probability zero.)

Since the denominator in (24), namely $f_{\mathbf{X}}W^n(\mathbf{y})$, is constant across all codewords, this is simply maximum-likelihood or, in this Gaussian case, minimum-Euclidean distance decoding. We will take advantage of the latter observation in our proof, more precisely the fact that

$$q(\mathbf{x}, \mathbf{y}) = \frac{n}{2} \log \frac{1}{2\pi} + \langle \mathbf{x}, \mathbf{y} \rangle - nP - \|\mathbf{y}\|_2^2 - \log f_{\mathbf{X}}W^n(\mathbf{y}) \quad (25)$$

only depends on the codeword through the inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i$. In fact, $q(\mathbf{x}, \mathbf{y})$ is equal to $\langle \mathbf{x}, \mathbf{y} \rangle$ up to a shift that only depends on $\|\mathbf{y}\|_2^2$.

Note that because $f_{\mathbf{X}}W^n$ is not a product density, $q(\mathbf{x}, \mathbf{y})$ is *not separable* (into a sum of n terms) unlike in the i.i.d. random coding case [9, Theorem 53].

C. The Random Coding Union (RCU) Bound

All the randomly drawn codewords satisfy the cost constraints with probability one. By using the same proof technique as that for the RCU bound [7, Theorem 16], we may assert that there exists an $(n, M, \varepsilon', P)_{\text{av}}$ -code satisfying

$$\varepsilon' \leq \mathbb{E} [\min \{1, M \Pr (q(\bar{\mathbf{X}}, \mathbf{Y}) \geq q(\mathbf{X}, \mathbf{Y}) | \mathbf{X}, \mathbf{Y}) \}] \quad (26)$$

where the random variables $(\bar{\mathbf{X}}, \mathbf{X}, \mathbf{Y})$ are distributed as $f_{\mathbf{X}}(\bar{\mathbf{x}}) \times f_{\mathbf{X}}(\mathbf{x}) \times W^n(\mathbf{y}|\mathbf{x})$. Now, introduce the function

$$g(t, \mathbf{y}) := \Pr (q(\bar{\mathbf{X}}, \mathbf{Y}) \geq t \mid \mathbf{Y} = \mathbf{y}). \quad (27)$$

Since $\bar{\mathbf{X}}$ is independent of \mathbf{X} , the probability in (26) can be written as

$$\Pr (q(\bar{\mathbf{X}}, \mathbf{Y}) \geq q(\mathbf{X}, \mathbf{Y}) | \mathbf{X}, \mathbf{Y}) = g(q(\mathbf{X}, \mathbf{Y}), \mathbf{Y}). \quad (28)$$

Furthermore, by Bayes rule, we have $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) \times f_{\mathbf{X}}W^n(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{x}) \times W^n(\mathbf{y}|\mathbf{x})$ and so

$$f_{\mathbf{X}}(\bar{\mathbf{x}}) = f_{\mathbf{X}}(\bar{\mathbf{x}}) \frac{f_{\mathbf{X}|\mathbf{Y}}(\bar{\mathbf{x}}|\mathbf{y})}{f_{\mathbf{X}|\mathbf{Y}}(\bar{\mathbf{x}}|\mathbf{y})} = f_{\mathbf{X}|\mathbf{Y}}(\bar{\mathbf{x}}|\mathbf{y}) \exp(-q(\bar{\mathbf{x}}, \mathbf{y})). \quad (29)$$

For a fixed sequence $\mathbf{y} \in \mathbb{R}^n$ and a constant $t \in \mathbb{R}$, multiplying both sides by $\mathbf{1}\{q(\bar{\mathbf{x}}, \mathbf{y}) \geq t\}$ and integrating over all $\bar{\mathbf{x}}$ yields the following alternative representation of $g(t, \mathbf{y})$:

$$g(t, \mathbf{y}) = \mathbb{E} [\exp(-q(\mathbf{X}, \mathbf{Y})) \mathbf{1}\{q(\mathbf{X}, \mathbf{Y}) \geq t\} \mid \mathbf{Y} = \mathbf{y}]. \quad (30)$$

D. A High-Probability Set

Consider the set of “typical” channel outputs whose norms are approximately $\sqrt{n(P+1)}$. More precisely, define

$$\mathcal{F} := \left\{ \mathbf{y} \in \mathbb{R}^n : \frac{1}{n} \|\mathbf{y}\|_2^2 \in [P+1-\delta, P+1+\delta] \right\}. \quad (31)$$

We claim that the probability of $\mathbf{Y} \in \mathcal{F}$ is large. First the union bound yields

$$\Pr(\mathbf{Y} \in \mathcal{F}^c) \leq \Pr\left(\frac{1}{n} \|\mathbf{X} + \mathbf{Z}\|_2^2 > P+1+\delta\right) + \Pr\left(\frac{1}{n} \|\mathbf{X} + \mathbf{Z}\|_2^2 < P+1-\delta\right). \quad (32)$$

Since the bounding of both probabilities can be done in a similar fashion, we focus on the first which may be written as

$$\Pr\left(\frac{1}{n} \|\mathbf{X} + \mathbf{Z}\|_2^2 > P+1+\delta\right) = \Pr\left(\frac{1}{n} (2\langle \mathbf{X}, \mathbf{Z} \rangle + \|\mathbf{Z}\|_2^2) > 1+\delta\right). \quad (33)$$

Define the following “typical” set of noises

$$\mathcal{G} := \left\{ \mathbf{z} \in \mathbb{R}^n : \frac{1}{n} \|\mathbf{z}\|_2^2 \leq 1 + \frac{\delta}{2} \right\}. \quad (34)$$

Since $\mathbf{Z} = (Z_1, \dots, Z_n) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n \times n})$, by the Chernoff bound (or, more precisely, by Cramer's theorem [19, Theorem 2.2.3] for χ_1^2 random variables), the probability that $\mathbf{Z} \in \mathcal{G}^c$ is upper bounded by $\exp(-\kappa_1 n \delta^2)$ for some constant $\kappa_1 > 0$. Now, we continue bounding the probability in (33) as follows:

$$\Pr\left(\frac{1}{n}(2\langle \mathbf{X}, \mathbf{Z} \rangle + \|\mathbf{Z}\|_2^2) > 1 + \delta\right) \leq \Pr\left(\frac{1}{n}(2\langle \mathbf{X}, \mathbf{Z} \rangle + \|\mathbf{Z}\|_2^2) > 1 + \delta \mid \mathbf{Z} \in \mathcal{G}\right) \Pr(\mathbf{Z} \in \mathcal{G}) + \Pr(\mathbf{Z} \in \mathcal{G}^c) \quad (35)$$

$$\leq \Pr\left(\frac{2}{n}\langle \mathbf{X}, \mathbf{Z} \rangle > \frac{\delta}{2} \mid \mathbf{Z} \in \mathcal{G}\right) \Pr(\mathbf{Z} \in \mathcal{G}) + \Pr(\mathbf{Z} \in \mathcal{G}^c) \quad (36)$$

$$\leq \Pr\left(\frac{1}{n} \sum_{i=1}^n X_i Z_i > \frac{\delta}{4}\right) + \Pr(\mathbf{Z} \in \mathcal{G}^c), \quad (37)$$

where in (36) we used the definition of \mathcal{G} . By spherical symmetry, we may take \mathbf{X} to be any point on the power sphere $\{\mathbf{x} : \|\mathbf{x}\|_2^2 = nP\}$. We take \mathbf{X} to be equal to $(\sqrt{nP}, 0, \dots, 0)$. Then the first term reduces to

$$\Pr\left(Z_1 > \frac{\delta}{4} \cdot \sqrt{\frac{n}{P}}\right) = 1 - \Phi\left(\frac{\delta}{4} \cdot \sqrt{\frac{n}{P}}\right) \leq \exp(-\kappa_2 n \delta^2), \quad (38)$$

where $\kappa_2 > 0$ is a constant. By putting all the bounds together and setting $\delta = n^{-1/3}$, we deduce that

$$\Pr(\mathbf{Y} \in \mathcal{F}) \geq 1 - \xi_n \quad (39)$$

where $\xi_n := \exp(-\kappa_3 n^{1/3})$ for some $\kappa_3 > 0$. Note that ξ_n decays faster than any polynomial.

E. Probability Of The Log-Likelihood Ratio Belonging To An Interval

We would like to upper bound $g(t, \mathbf{y})$ in (27) to evaluate the RCU bound. This we do in the next section. As an intermediate step, we consider the problem of upper bounding

$$h(\mathbf{y}; a, \mu) := \Pr(q(\mathbf{X}, \mathbf{Y}) \in [a, a + \mu] \mid \mathbf{Y} = \mathbf{y}), \quad (40)$$

where $a \in \mathbb{R}$ and $\mu > 0$ are some constants. Because \mathbf{Y} is fixed to some constant vector \mathbf{y} and $\|\mathbf{X}\|_2^2$ is also constant, $h(\mathbf{y}; a, \mu)$ can be rewritten using (25) as

$$h(\mathbf{y}; a, \mu) := \Pr(\langle \mathbf{X}, \mathbf{Y} \rangle \in [a', a' + \mu] \mid \mathbf{Y} = \mathbf{y}), \quad (41)$$

for some other constant $a' \in \mathbb{R}$. It is clear that $h(\mathbf{y}; a, \mu)$ depends on \mathbf{y} through its norm and so we may define (with an abuse of notation),

$$h(s; a, \mu) := h(\mathbf{y}; a, \mu), \quad \text{if } s = \frac{1}{n} \|\mathbf{y}\|_2^2. \quad (42)$$

In the rest of this section, we assume that $\mathbf{y} \in \mathcal{F}$ or, equivalently, $s \in [P + 1 - \delta, P + 1 + \delta]$.

By introducing the standard Gaussian random vector $\mathbf{Z} = (Z_1, \dots, Z_n) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n \times n})$, we have

$$h(s; a, \mu) = \Pr\left(\langle \mathbf{X}, \mathbf{X} + \mathbf{Z} \rangle \in [a', a' + \mu] \mid \|\mathbf{X} + \mathbf{Z}\|_2^2 = ns\right) \quad (43)$$

$$= \Pr\left(\sum_{i=1}^n X_i Z_i + nP \in [a', a' + \mu] \mid \|\mathbf{X} + \mathbf{Z}\|_2^2 = ns\right) \quad (44)$$

where (44) follows by the observation that $\langle \mathbf{X}, \mathbf{X} \rangle = nP$ with probability one. Now, define

$$\mathbf{x}_0 := (\sqrt{nP}, 0, \dots, 0) \quad (45)$$

to be a fixed vector on the power sphere. By spherical symmetry, we may pick \mathbf{X} in (44) to be equal to \mathbf{x}_0 . Thus, we have

$$h(s; a, \mu) = \Pr\left(Z_1 + \sqrt{nP} \in \left[\frac{a'}{\sqrt{nP}}, \frac{a' + \mu}{\sqrt{nP}}\right] \mid \|\mathbf{x}_0 + \mathbf{Z}\|_2^2 = ns\right). \quad (46)$$

In other words, we are conditioning on the event that the random vector $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n \times n})$ lands on the surface of a sphere of radius \sqrt{ns} centered at $-\mathbf{x}_0 = (-\sqrt{nP}, 0, \dots, 0)$. See Fig. 1. We are then asking what is the probability that the first component plus \sqrt{nP} belongs to the prescribed interval of length proportional to μ/\sqrt{n} .

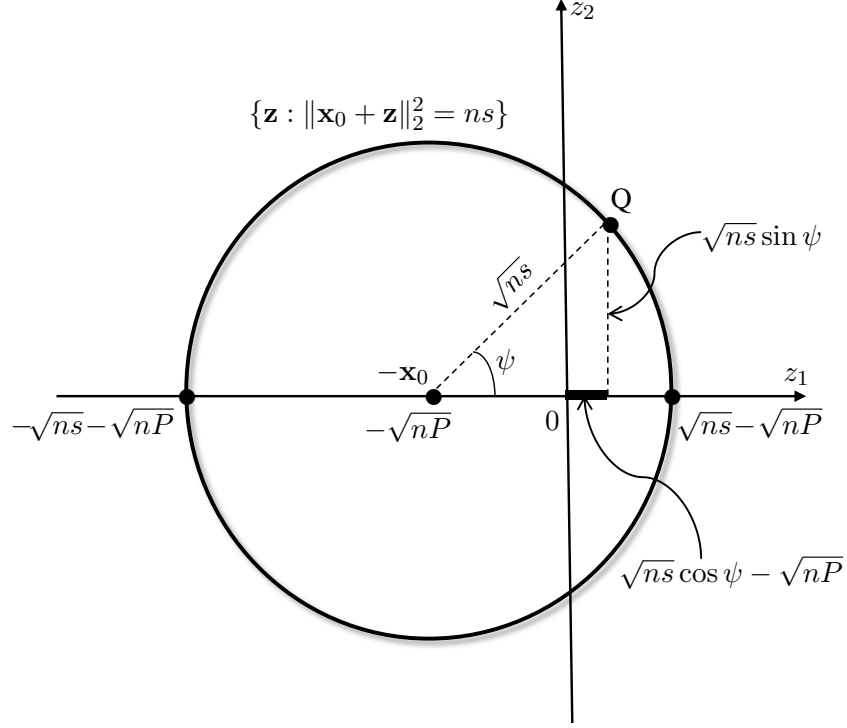


Fig. 1. Illustration of the relation between Z_1 and Ψ in (47) in two dimensions. The transformation of this figure to the U coordinate system via (49) translates the sphere to the origin and scales its radius to be 1.

Let us now derive the conditional density of Z_1 given the event $\mathcal{E} := \{\|\mathbf{x}_0 + \mathbf{Z}\|_2^2 = ns\}$. Denote this density as $f_{Z_1|\mathcal{E}}(z_1)$. Note that the support of $f_{Z_1|\mathcal{E}}(z_1)$ is $[-\sqrt{ns} - \sqrt{nP}, \sqrt{ns} - \sqrt{nP}]$. It is easier to find the conditional density of the angle $\Psi \in [0, 2\pi]$ given the event \mathcal{E} where Ψ and Z_1 are related as follows:

$$Z_1 = \sqrt{ns} \cos \Psi - \sqrt{nP}. \quad (47)$$

Again see Fig. 1. Now, we have

$$f_{\Psi|\mathcal{E}}(\psi) d\psi \propto (\sin^{n-2} \psi) \exp\left(-\frac{n}{2} [(\sqrt{s} \cos \psi - \sqrt{P})^2 + s \sin^2 \psi]\right) d\psi. \quad (48)$$

This follows because the area element (an $(n-1)$ -dimensional annulus of radius $\sqrt{ns} \sin \psi$ and width $d\psi$) is proportional to $\sin^{n-2} \psi$ (similar to Shannon's derivation in [2, Eq. (21)]) and the Gaussian weighting is proportional to $\exp\left(-\frac{n}{2} [(\sqrt{s} \cos \psi - \sqrt{P})^2 + s \sin^2 \psi]\right)$. This is just $\exp(-d^2/2)$ where d is the distance of the point described by ψ (point Q in Fig. 1) to the origin. We are obviously leveraging heavily on the radial symmetry of the problem around the first axis. Now, we consider the change of variables

$$U = \cos \Psi \quad (49)$$

resulting in

$$f_{U|\mathcal{E}}(u) du \propto (1 - u^2)^{(n-3)/2} \exp(n\sqrt{P}su) du. \quad (50)$$

Note that U takes values in $[-1, 1]$. More precisely, the conditional density of U given \mathcal{E} is

$$f_{U|\mathcal{E}}(u) = \frac{1}{F_n} (1 - u^2)^{(n-3)/2} \exp(n\sqrt{P}su) \mathbf{1}\{u \in [-1, 1]\}, \quad (51)$$

where the normalization constant is

$$F_n := \int_{-1}^1 (1 - u^2)^{(n-3)/2} \exp(n\sqrt{P}su) du. \quad (52)$$

The conditional density we have derived in (51)–(52) reduces to that by Stam [20, Eq. (3)] for the limiting case $P = 0$, i.e. the sphere is centered at the origin. It is of paramount importance to analyze how $\sup_{u \in [-1, 1]} f_{U|\mathcal{E}}(u)$

scales with n . The answer turns out to be $O(\sqrt{n})$. More formally, we state the following lemma whose proof is provided in Appendix B.

Lemma 2. *Define the function*

$$L(P, s) := \frac{(2Ps)^2}{\sqrt{2\pi}} \cdot \sqrt{\frac{1 + 4Ps - \sqrt{1 + 4Ps}}{(\sqrt{1 + 4Ps} - 1)^5}}. \quad (53)$$

The following bound holds:

$$\limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \sup_{u \in [-1, 1]} f_{U|\mathcal{E}}(u) \leq L(P, s). \quad (54)$$

Equipped with this lemma, let us consider the probability $h(s; a, \mu)$ in (46). We have

$$h(s; a, \mu) = \Pr \left(\sqrt{ns} U \in \left[\frac{a'}{\sqrt{nP}}, \frac{a' + \mu}{\sqrt{nP}} \right] \middle| \mathcal{E} \right) \quad (55)$$

$$= \int_{a'/(n\sqrt{Ps})}^{(a' + \mu)/(n\sqrt{Ps})} f_{U|\mathcal{E}}(u) du \quad (56)$$

$$\leq \int_{a'/(n\sqrt{Ps})}^{(a' + \mu)/(n\sqrt{Ps})} 2 L(P, s) \sqrt{n} du \quad (57)$$

$$= \frac{2 L(P, s) \mu}{\sqrt{nPs}}, \quad (58)$$

where (55) follows from the fact that $Z_1 = \sqrt{ns} U - \sqrt{nP}$ due to (47) and (49), and (57) holds for all sufficiently large n (depending only on P and s) on account of Lemma 2.

Since $s \in [P + 1 - \delta, P + 1 + \delta]$ and $\delta = n^{-1/3} \rightarrow 0$, we deduce that for all $\mathbf{y} \in \mathcal{F}$ and n sufficiently large (depending only on P),

$$h(\mathbf{y}; a, \mu) \leq K(P) \cdot \frac{\mu}{\sqrt{n}}, \quad (59)$$

for some function $K(P)$. In fact, by the continuity of $s \mapsto L(P, s)$, the constant $K(P)$ can be taken to be

$$K(P) = \frac{3 L(P, P + 1)}{\sqrt{P(P + 1)}}. \quad (60)$$

F. Probability That The Decoding Metric Exceeds t For An Incorrect Codeword

We now return to bounding $g(t, \mathbf{y})$ defined in (27). Again, we assume $\mathbf{y} \in \mathcal{F}$. The idea here is to consider the second form of $g(t, \mathbf{y})$ in (30) and to slice the interval $[t, \infty)$ into non-overlapping segments $\{[t + l\eta, t + (l + 1)\eta) : l \in \mathbb{N} \cup \{0\}\}$ where $\eta > 0$ is a constant. Then we apply (59) to each segment. This is modelled after the proof of [7, Lemma 47]. Indeed, we have

$$\begin{aligned} g(t, \mathbf{y}) &= \mathbb{E}[\exp(-q(\mathbf{X}, \mathbf{Y})) \mathbf{1}\{q(\mathbf{X}, \mathbf{Y}) \geq t\} \mid \mathbf{Y} = \mathbf{y}] \\ &\leq \sum_{l=0}^{\infty} \exp(-t - l\eta) \Pr(t + l\eta \leq q(\mathbf{X}, \mathbf{Y}) < t + (l + 1)\eta \mid \mathbf{Y} = \mathbf{y}) \end{aligned} \quad (61)$$

$$\leq \sum_{l=0}^{\infty} \exp(-t - l\eta) \cdot \frac{K(P) \eta}{\sqrt{n}} \quad (62)$$

$$= \frac{\exp(-t)}{1 - \exp(-\eta)} \cdot \frac{K(P) \eta}{\sqrt{n}}. \quad (63)$$

Since η is a free parameter, we may choose it to be $\log 2$ yielding

$$g(t, \mathbf{y}) \leq \frac{G \exp(-t)}{\sqrt{n}} \quad (64)$$

where $G = G(P) = (2 \log 2) K(P)$.

G. Evaluating The RCU Bound

We now have all the necessary ingredients to evaluate the RCU bound in (26). Consider,

$$\begin{aligned}\varepsilon' &\leq \mathbb{E} \left[\min \{1, Mg(q(\mathbf{X}, \mathbf{Y}), \mathbf{Y})\} \right] \\ &\leq \Pr(\mathbf{Y} \in \mathcal{F}^c) + \mathbb{E} \left[\min \{1, Mg(q(\mathbf{X}, \mathbf{Y}), \mathbf{Y})\} \mid \mathbf{Y} \in \mathcal{F} \right] \cdot \Pr(\mathbf{Y} \in \mathcal{F})\end{aligned}\quad (65)$$

$$\leq \Pr(\mathbf{Y} \in \mathcal{F}^c) + \mathbb{E} \left[\min \left\{ 1, \frac{MG \exp(-q(\mathbf{X}, \mathbf{Y}))}{\sqrt{n}} \right\} \mid \mathbf{Y} \in \mathcal{F} \right] \cdot \Pr(\mathbf{Y} \in \mathcal{F}) \quad (66)$$

$$\leq \xi_n + \mathbb{E} \left[\min \left\{ 1, \frac{MG \exp(-q(\mathbf{X}, \mathbf{Y}))}{\sqrt{n}} \right\} \mid \mathbf{Y} \in \mathcal{F} \right] \cdot \Pr(\mathbf{Y} \in \mathcal{F}) \quad (67)$$

where (66) is due to (64) with $t = q(\mathbf{X}, \mathbf{Y})$ and (67) uses the bound in (39). Now we split the expectation into two parts depending on whether $q(\mathbf{x}, \mathbf{y}) > \log(MG/\sqrt{n})$ or otherwise, i.e.

$$\begin{aligned}\mathbb{E} \left[\min \left\{ 1, \frac{MG \exp(-q(\mathbf{X}, \mathbf{Y}))}{\sqrt{n}} \right\} \mid \mathbf{Y} \in \mathcal{F} \right] \\ \leq \Pr \left(q(\mathbf{X}, \mathbf{Y}) \leq \log \frac{MG}{\sqrt{n}} \mid \mathbf{Y} \in \mathcal{F} \right) + \frac{MG}{\sqrt{n}} \mathbb{E} \left[\mathbf{1} \left\{ q(\mathbf{X}, \mathbf{Y}) > \log \frac{MG}{\sqrt{n}} \right\} \exp(-q(\mathbf{X}, \mathbf{Y})) \mid \mathbf{Y} \in \mathcal{F} \right].\end{aligned}\quad (68)$$

By applying (64) with $t = \log(MG/\sqrt{n})$, we know that the second term can be bounded as

$$\frac{MG}{\sqrt{n}} \mathbb{E} \left[\mathbf{1} \left\{ q(\mathbf{X}, \mathbf{Y}) > \log \frac{MG}{\sqrt{n}} \right\} \exp(-q(\mathbf{X}, \mathbf{Y})) \mid \mathbf{Y} \in \mathcal{F} \right] \leq \frac{G}{\sqrt{n}}. \quad (69)$$

Now let $f_Y^*(y) = \mathcal{N}(y; 0, P+1)$ be the capacity-achieving output distribution and $f_Y^*(\mathbf{y}) = \prod_{i=1}^n f_Y^*(y_i)$ its n -fold memoryless extension. In Step 1 of the proof of Lemma 61 in [7], Polyanskiy-Poor-Verdú showed that on \mathcal{F} , the ratio of the induced output density $f_{\mathbf{X}} W^n(\mathbf{y})$ and $f_Y^*(\mathbf{y})$ can be bounded by a finite constant J , i.e.

$$\sup_{\mathbf{y} \in \mathcal{F}} \frac{f_{\mathbf{X}} W^n(\mathbf{y})}{f_Y^*(\mathbf{y})} \leq J. \quad (70)$$

Also see [21, Proposition 2]. We return to bounding the first term in (68). Using the definition of $q(\mathbf{x}, \mathbf{y})$ in (24) and applying the bound in (70) yields

$$\Pr \left(q(\mathbf{X}, \mathbf{Y}) \leq \log \frac{MG}{\sqrt{n}} \mid \mathbf{Y} \in \mathcal{F} \right) = \Pr \left(\log \frac{W^n(\mathbf{Y}|\mathbf{X})}{f_{\mathbf{X}} W^n(\mathbf{Y})} \leq \log \frac{MG}{\sqrt{n}} \mid \mathbf{Y} \in \mathcal{F} \right) \quad (71)$$

$$\leq \Pr \left(\log \frac{W^n(\mathbf{Y}|\mathbf{X})}{f_Y^*(\mathbf{Y})} \leq \log \frac{MGJ}{\sqrt{n}} \mid \mathbf{Y} \in \mathcal{F} \right). \quad (72)$$

Thus, when we multiply the first term in (68) by $\Pr(\mathbf{Y} \in \mathcal{F})$, use Bayes rule and drop the event $\{\mathbf{Y} \in \mathcal{F}\}$, we see that the product can be bounded as follows:

$$\Pr \left(q(\mathbf{X}, \mathbf{Y}) \leq \log \frac{MG}{\sqrt{n}} \mid \mathbf{Y} \in \mathcal{F} \right) \cdot \Pr(\mathbf{Y} \in \mathcal{F}) \leq \Pr \left(\log \frac{W^n(\mathbf{Y}|\mathbf{X})}{f_Y^*(\mathbf{Y})} \leq \log \frac{MGJ}{\sqrt{n}} \right). \quad (73)$$

The right-hand-side of (73) can be written as an average over $\mathbf{X} \sim f_{\mathbf{X}}$, i.e.

$$\Pr \left(\log \frac{W^n(\mathbf{Y}|\mathbf{X})}{f_Y^*(\mathbf{Y})} \leq \log \frac{MGJ}{\sqrt{n}} \right) = \int_{\mathbf{x}} f_{\mathbf{X}}(\mathbf{x}) \Pr \left(\log \frac{W^n(\mathbf{Y}|\mathbf{X})}{f_Y^*(\mathbf{Y})} \leq \log \frac{MGJ}{\sqrt{n}} \mid \mathbf{X} = \mathbf{x} \right) d\mathbf{x}. \quad (74)$$

By noting that $f_Y^*(\mathbf{y})$ is a product density,

$$\Pr \left(\log \frac{W^n(\mathbf{Y}|\mathbf{X})}{f_Y^*(\mathbf{Y})} \leq \log \frac{MGJ}{\sqrt{n}} \mid \mathbf{X} = \mathbf{x} \right) = \Pr \left(\sum_{i=1}^n \log \frac{W(Y_i|X_i)}{f_Y^*(Y_i)} \leq \log \frac{MGJ}{\sqrt{n}} \mid \mathbf{X} = \mathbf{x} \right). \quad (75)$$

The above probability does not depend on \mathbf{x} as long as it is on the power sphere $\{\mathbf{x} : \|\mathbf{x}\|_2^2 = nP\}$ because of spherical symmetry. Hence we may take $\mathbf{x} = (\sqrt{P}, \dots, \sqrt{P})$. It is then easy to check that the first two central moments of the information density are

$$\mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n \log \frac{W(Y_i|\sqrt{P})}{f_Y^*(Y_i)} \right] = C(P), \quad \text{and} \quad \text{Var} \left[\frac{1}{n} \sum_{i=1}^n \log \frac{W(Y_i|\sqrt{P})}{f_Y^*(Y_i)} \right] = \frac{V(P)}{n}. \quad (76)$$

Furthermore, the following third-absolute moment

$$\mathsf{T}(P) := \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\left| \log \frac{W(Y_i|\sqrt{P})}{f_Y^*(Y_i)} - \mathbb{E} \left[\log \frac{W(Y_i|\sqrt{P})}{f_Y^*(Y_i)} \right] \right|^3 \right] \quad (77)$$

is obviously bounded (note the scaling). See [22, Lemma 10 and Appendix A] for a precise analysis of third absolute moments of information densities involving Gaussians. This allows us to apply the Berry-Esseen theorem [23, Theorem 2 in Section XVI.5], which implies that

$$\Pr \left(\log \frac{W^n(\mathbf{Y}|\mathbf{X})}{f_{\mathbf{Y}}^*(\mathbf{Y})} \leq \log \frac{MGJ}{\sqrt{n}} \middle| \mathbf{X} = (\sqrt{P}, \dots, \sqrt{P}) \right) \leq \Phi \left(\frac{\log \frac{MGJ}{\sqrt{n}} - n\mathsf{C}(P)}{\sqrt{n\mathsf{V}(P)}} \right) + \frac{6\mathsf{T}(P)}{\sqrt{n\mathsf{V}(P)^3}}. \quad (78)$$

Let $B = B(P) := 6\mathsf{T}(P)/\mathsf{V}(P)^{3/2}$. We deduce that

$$\Pr \left(\log \frac{W^n(\mathbf{Y}|\mathbf{X})}{f_{\mathbf{Y}}^*(\mathbf{Y})} \leq \log \frac{MGJ}{\sqrt{n}} \right) \leq \Phi \left(\frac{\log \frac{MGJ}{\sqrt{n}} - n\mathsf{C}(P)}{\sqrt{n\mathsf{V}(P)}} \right) + \frac{B}{\sqrt{n}}. \quad (79)$$

Putting all the bounds together, we obtain

$$\varepsilon' \leq \Phi \left(\frac{\log \frac{MGJ}{\sqrt{n}} - n\mathsf{C}(P)}{\sqrt{n\mathsf{V}(P)}} \right) + \frac{B}{\sqrt{n}} + \frac{G}{\sqrt{n}} + \xi_n. \quad (80)$$

Now choose

$$\log M = n\mathsf{C}(P) + \sqrt{n\mathsf{V}(P)}\Phi^{-1} \left(\varepsilon - \frac{B+G}{\sqrt{n}} - \xi_n \right) + \frac{1}{2} \log n - \log(GJ) \quad (81)$$

ensuring that

$$\varepsilon' \leq \varepsilon. \quad (82)$$

Hence, there exists an $(n, M, \varepsilon, P)_{\text{av}}$ -code where M is given by (81). It is easily seen by Taylor expanding $\Phi^{-1}(\cdot)$ around ε that

$$\log M = n\mathsf{C}(P) + \sqrt{n\mathsf{V}(P)}\Phi^{-1}(\varepsilon) + \frac{1}{2} \log n + O(1). \quad (83)$$

This completes the proof of Theorem 1. \square

APPENDIX A

MODIFICATIONS OF THE PROOF TO THE PARALLEL GAUSSIAN CHANNELS SETTING

In this appendix, we give a sketch of how the proof of Theorem 1 can be used for the scenario where information is to be transmitted across k parallel Gaussian channels. See Section 9.4 of [12] for the precise problem setting. Let the input and output to the channel be $(\mathbf{X}_1, \dots, \mathbf{X}_k)$ and $(\mathbf{Y}_1, \dots, \mathbf{Y}_k)$ respectively. Let the independent noises of each of the channels have variances N_1, \dots, N_k and denote the total admissible power as P . Let $|\cdot|^+ := \max\{0, \cdot\}$ and set P_1, \dots, P_k be the power assignments that maximize the information capacity expression, i.e.

$$P_j = |\nu - N_j|^+ \quad (A.1)$$

where the Karush-Kuhn-Tucker multiplier ν is chosen to satisfy the total power constraint

$$\sum_{j=1}^k |\nu - N_j|^+ = P. \quad (A.2)$$

Let $\mathcal{P}^+ := \{j \in \{1, \dots, k\} : P_j > 0\}$. Clearly, (A.1) and (A.2) imply that \mathcal{P}^+ is non-empty if $P > 0$. We use the random coding distribution $f_{\mathbf{X}_1} \times \dots \times f_{\mathbf{X}_k}$ where each constituent distribution $f_{\mathbf{X}_j}$ is given by (21) with P_j in place of P there. Close inspection of the proof of Theorem 1 shows that the only estimate that needs to be verified is (58). For this, we consider the analogue of (44) which can be written as

$$h(s_1, \dots, s_k; a, \mu) = \Pr \left(\sum_{j=1}^k \sqrt{P_j} Z_{j1} \in \left[\frac{a_2}{\sqrt{n}}, \frac{a_2 + \mu}{\sqrt{n}} \right] \middle| \|\mathbf{X}_j + \mathbf{Z}_j\|_2^2 = ns_j, \forall j \in \{1, \dots, k\} \right), \quad (A.3)$$

where a_2 is related to a' in (44) by a constant shift. Note that the sum of the inner products $\sum_{j=1}^k \langle \mathbf{X}_j, \mathbf{Y}_j \rangle$ in the analogue of (41) reduces to $\sum_{j=1}^k \sqrt{P_j} Z_{j1} = \sum_{j \in \mathcal{P}^+} \sqrt{P_j} Z_{j1}$ once we have exploited spherical symmetry to choose $\mathbf{X}_j = \mathbf{x}_{j0} := (\sqrt{n P_j}, 0, \dots, 0)$ and moved all the constants to the right-hand-side. Let \mathcal{E} be the event $\{\|\mathbf{x}_{j0} + \mathbf{Z}_j\|_2^2 = n s_j, \forall j \in \{1, \dots, k\}\}$. By introducing the independent random variables $\{U_j : j \in \mathcal{P}^+\}$ that are related to $\{Z_{j1} : j \in \mathcal{P}^+\}$ analogously to (47), we see that (A.3) reduces to

$$h(s_1, \dots, s_k; a, \mu) = \Pr \left(\sum_{j \in \mathcal{P}^+} \sqrt{P_j s_j} U_j \in \left[\frac{a_3}{n}, \frac{a_3 + \mu}{n} \right] \middle| \mathcal{E} \right), \quad (\text{A.4})$$

where a_3 is related to a_2 by a constant shift. In principle, since the U_j 's are independent, we can use its distribution in (51) to find the distribution of $\sum_{j \in \mathcal{P}^+} \sqrt{P_j s_j} U_j$ by convolution and bound the probability using the steps that led to (58). However, the following method proves to be easier. Let l be any element in \mathcal{P}^+ then consider

$$\begin{aligned} & h(s_1, \dots, s_k; a, \mu) \\ &= \int \Pr \left(\sum_{j \in \mathcal{P}^+} \sqrt{P_j s_j} U_j \in \left[\frac{a_3}{n}, \frac{a_3 + \mu}{n} \right] \middle| \mathcal{E}, \{\forall j \in \mathcal{P}^+ \setminus \{l\}, U_j = u_j\} \right) \prod_{j \in \mathcal{P}^+ \setminus \{l\}} f_{U_j|\mathcal{E}}(u_j) du_j \end{aligned} \quad (\text{A.5})$$

$$= \int \Pr \left(\sqrt{P_l s_l} U_l \in \left[\frac{a_4}{n}, \frac{a_4 + \mu}{n} \right] \middle| \mathcal{E}, \{\forall j \in \mathcal{P}^+ \setminus \{l\}, U_j = u_j\} \right) \prod_{j \in \mathcal{P}^+ \setminus \{l\}} f_{U_j|\mathcal{E}}(u_j) du_j \quad (\text{A.6})$$

$$= \int \Pr \left(\sqrt{P_l s_l} U_l \in \left[\frac{a_4}{n}, \frac{a_4 + \mu}{n} \right] \middle| \mathcal{E} \right) \prod_{j \in \mathcal{P}^+ \setminus \{l\}} f_{U_j|\mathcal{E}}(u_j) du_j \quad (\text{A.7})$$

$$\leq \int \frac{2 L(P_l, s_l) \mu}{\sqrt{n P_l s_l}} \prod_{j \in \mathcal{P}^+ \setminus \{l\}} f_{U_j|\mathcal{E}}(u_j) du_j \quad (\text{A.8})$$

$$= \frac{2 L(P_l, s_l) \mu}{\sqrt{n P_l s_l}}, \quad (\text{A.9})$$

where (A.5) follows from the law of total probability; (A.6) follows by noting that $\{u_j : j \in \mathcal{P}^+ \setminus \{l\}\}$ are constants and defining a_4 to be related to a_3 by a constant shift; (A.7) is due to the joint independence of the random variables $\{U_j : j \in \mathcal{P}^+\}$; and finally (A.8), which holds for n sufficiently large, follows by the same reasoning in the steps that led to (58). Since $l \in \mathcal{P}^+$ is arbitrary,

$$h(s_1, \dots, s_k; a, \mu) \leq \min_{l \in \mathcal{P}^+} \frac{2 L(P_l, s_l) \mu}{\sqrt{n P_l s_l}}. \quad (\text{A.10})$$

We conclude that, just as in (59), the probability $h(\mathbf{y}_1, \dots, \mathbf{y}_k; a, \mu)$ is still bounded above by a constant multiple of μ/\sqrt{n} and the constant does not depend on a . The rest of the proof proceeds *mutatis mutandis*.

APPENDIX B PROOF OF LEMMA 2

We first find a lower bound for the normalization constant F_n defined in (52). Using the fact that $(1 - u^2)^{-3/2} \geq 1$, we have

$$F_n \geq \underline{F}_n := \int_{-1}^1 \exp(n\alpha(u)) du \quad (\text{B.1})$$

where the exponent is

$$\alpha(u) := \frac{1}{2} \log(1 - u^2) + \sqrt{P s} u. \quad (\text{B.2})$$

This exponent is maximized at

$$u^* = \frac{\sqrt{1 + 4 P s} - 1}{2 \sqrt{P s}}, \quad (\text{B.3})$$

which is in the interior of $[-1, 1]$ for finite P . Furthermore, the second derivative of α is

$$\alpha''(u) = -\frac{(1 + u^2)}{(1 - u^2)^2} \quad (\text{B.4})$$

which is always negative. Now we use Laplace's method to lower bound the definite integral in (B.1) with that of a Gaussian [24], [25]. We provide the details for the reader's convenience. Let $\epsilon \in (0, -\alpha''(u^*))$. By the continuity of $\alpha''(u)$ at u^* and Taylor's theorem, there exists a $\zeta \in (0, 1 - u^*)$ such that for any $u \in (u^* - \zeta, u^* + \zeta) \subset [-1, 1]$, we have $\alpha(u) \geq \alpha(u^*) + \frac{1}{2}(\alpha''(u^*) - \epsilon)(u - u^*)^2$. The following lower bounds hold:

$$\underline{F}_n \geq \int_{u^* - \zeta}^{u^* + \zeta} \exp(n\alpha(u)) du \quad (\text{B.5})$$

$$\geq \exp(n\alpha(u^*)) \int_{u^* - \zeta}^{u^* + \zeta} \exp\left(\frac{n}{2}(\alpha''(u^*) - \epsilon)(u - u^*)^2\right) du \quad (\text{B.6})$$

$$= \exp(n\alpha(u^*)) \sqrt{\frac{1}{n(-\alpha''(u^*) + \epsilon)}} \int_{-\zeta\sqrt{n(-\alpha''(u^*) + \epsilon)}}^{\zeta\sqrt{n(-\alpha''(u^*) + \epsilon)}} e^{-v^2/2} dv. \quad (\text{B.7})$$

We used the change of variables $v = \sqrt{n(-\alpha''(u^*) + \epsilon)}(u - u^*)$ in the final step. The integral in (B.7) tends to $\sqrt{2\pi}$ as n becomes large so

$$\liminf_{n \rightarrow \infty} \frac{\underline{F}_n}{\sqrt{\frac{2\pi}{n|\alpha''(u^*)|}} \exp(n\alpha(u^*))} \geq \sqrt{\frac{-\alpha''(u^*)}{-\alpha''(u^*) + \epsilon}}. \quad (\text{B.8})$$

Since $\epsilon > 0$ is arbitrary, we can rewrite (B.8) as

$$\underline{F}_n \geq \gamma_n \sqrt{\frac{2\pi}{n|\alpha''(u^*)|}} \exp(n\alpha(u^*)), \quad (\text{B.9})$$

for some sequence γ_n that converges to 1 as $n \rightarrow \infty$. Furthermore, the numerator of $f_{U|\mathcal{E}}(u)$ in (51) can be upper bounded as

$$(1 - u^2)^{(n-3)/2} \exp(n\sqrt{P}su) = \exp(n\beta_n(u)) \leq \exp(n\beta_n(u_n^*)) \quad (\text{B.10})$$

where the exponent is

$$\beta_n(u) := \left(\frac{1}{2} - \frac{3}{2n}\right) \log(1 - u^2) + \sqrt{P}su \quad (\text{B.11})$$

and the maximizer of $\beta_n(u)$ is

$$u_n^* := \frac{\sqrt{(1 - \frac{3}{n})^2 + 4Ps} - (1 - \frac{3}{n})}{2\sqrt{Ps}}. \quad (\text{B.12})$$

Clearly, $u_n^* \rightarrow u^*$ as $n \rightarrow \infty$. We have, by uniting (B.9) and (B.10), that

$$\sup_{u \in [-1, 1]} f_{U|\mathcal{E}}(u) \leq \frac{1}{\gamma_n} \sqrt{\frac{n|\alpha''(u^*)|}{2\pi}} \exp(n[\beta_n(u_n^*) - \alpha(u^*)]). \quad (\text{B.13})$$

Now, we examine the exponent $\beta_n(u_n^*) - \alpha(u^*)$ above. We have

$$\beta_n(u_n^*) - \alpha(u^*) \leq \beta_n(u_n^*) - \alpha(u_n^*) = \frac{3}{2n} \log \frac{1}{1 - (u_n^*)^2} \quad (\text{B.14})$$

where the inequality follows because u^* maximizes α and so $\alpha(u_n^*) \leq \alpha(u^*)$ and the equality is due to the definitions of $\alpha(u)$ and $\beta_n(u)$. Thus, (B.13) can be further upper bounded as

$$\sup_{u \in [-1, 1]} f_{U|\mathcal{E}}(u) \leq \frac{1}{\gamma_n} \cdot \sqrt{\frac{n|\alpha''(u^*)|}{2\pi}} \cdot \frac{1}{(1 - (u_n^*)^2)^{3/2}}. \quad (\text{B.15})$$

Dividing both sides by \sqrt{n} and taking the lim sup shows that the upper bound can be chosen to be

$$L(P, s) = \frac{1}{(1 - (u^*)^2)^{3/2}} \cdot \sqrt{\frac{|\alpha''(u^*)|}{2\pi}} = \sqrt{\frac{1 + (u^*)^2}{2\pi(1 - (u^*)^2)^5}}. \quad (\text{B.16})$$

This concurs with (53) after we substitute for the value of u^* in (B.3). \square

Acknowledgements

VT sincerely thanks Shaowei Lin (I²R, A*STAR) for many helpful explanations concerning approximation of integrals in high dimensions. The authors also thank Jonathan Scarlett (Cambridge) and Yücel Altuğ (Cornell) for discussions and constructive comments on the manuscript.

REFERENCES

- [1] C. E. Shannon, “A mathematical theory of communication,” *Bell Systems Technical Journal*, vol. 27, pp. 379–423, 1948.
- [2] C. E. Shannon, “Probability of error for optimal codes in a Gaussian channel,” *Bell Systems Technical Journal*, vol. 38, pp. 611–656, 1959.
- [3] K. Yoshihara, “Simple proofs for the strong converse theorems in some channels,” *Kodai Mathematical Journal*, vol. 16, no. 4, pp. 213–222, 1964.
- [4] J. Wolfowitz, *Coding Theorems of Information Theory*. Springer-Verlag, New York, 3rd ed., 1978.
- [5] V. Strassen, “Asymptotische Abschätzungen in Shannons Informationstheorie,” in *Trans. Third Prague Conf. Inf. Theory*, (Prague), pp. 689–723, 1962.
- [6] M. Hayashi, “Information spectrum approach to second-order coding rate in channel coding,” *IEEE Trans. on Inf. Th.*, vol. 55, pp. 4947–4966, Nov 2009.
- [7] Y. Polyanskiy, H. V. Poor, and S. Verdú, “Channel coding rate in the finite blocklength regime,” *IEEE Trans. on Inf. Th.*, vol. 56, pp. 2307–2359, May 2010.
- [8] M. Tomamichel and V. Y. F. Tan, “A tight upper bound for the third-order asymptotics of most discrete memoryless channels,” *IEEE Trans. on Inf. Th.*, vol. 59, pp. 7041–7051, Nov 2013.
- [9] Y. Polyanskiy, *Channel coding: Non-asymptotic fundamental limits*. PhD thesis, Princeton University, 2010.
- [10] Y. Altuğ and A. B. Wagner, “The third-order term in the normal approximation for singular channels,” arXiv:1309.5126 [cs.IT], Sep 2013.
- [11] P. Moulin, “The log-volume of optimal codes for memoryless channels, within a few nats,” arXiv:1311.0181 [cs.IT], Nov 2013.
- [12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 2nd ed., 2006.
- [13] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [14] Y. Altuğ and A. B. Wagner, “Refinement of the sphere packing bound for symmetric channels,” in *Proc. 49th Annual Allerton Conf. Communication, Control, and Computing*, 2011.
- [15] Y. Altuğ and A. B. Wagner, “A refinement of the random coding bound,” in *Proc. 50th Annual Allerton Conf. Communication, Control, and Computing*, 2012.
- [16] Y. Altuğ and A. B. Wagner, “Refinement of the sphere packing bound,” in *Int. Symp. Inf. Th.*, (Cambridge, MA), 2012.
- [17] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, “A derivation of the asymptotic random-coding prefactor,” in *Proc. 51st Annual Allerton Conf. Communication, Control, and Computing*, 2013. arXiv:1306.6203 [cs.IT].
- [18] Y. Polyanskiy, “Saddle point in the minimax converse for channel coding,” *IEEE Trans. on Inf. Th.*, vol. 59, pp. 2576–2595, May 2013.
- [19] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer, 2nd ed., 1998.
- [20] A. J. Stam, “Limit theorems for uniform distributions on spheres in high-dimensional Euclidean spaces,” *Journal of Applied Probability*, vol. 19, no. 1, pp. 221–228, 1982.
- [21] E. MolavianJazi and J. N. Laneman, “A finite-blocklength perspective on Gaussian multi-access channels,” arXiv:1309.2343 [cs.IT], Sep 2013.
- [22] J. Scarlett and V. Y. F. Tan, “Second-order asymptotics for the Gaussian MAC with degraded message sets,” arXiv:1310.1197 [cs.IT], Oct 2013.
- [23] W. Feller, *An Introduction to Probability Theory and Its Applications*. John Wiley and Sons, 2nd ed., 1971.
- [24] L. Tierney and J. B. Kadane, “Accurate approximations for posterior moments and marginal densities,” *Journal of the American Statistical Association*, vol. 81, pp. 82–86, Mar 1986.
- [25] Z. Shun and P. McCullagh, “Laplace approximation of high dimensional integrals,” *Journal of the Royal Statistical Society, Series B (Methodology)*, vol. 57, no. 4, pp. 749–760, 1995.