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

Vincent Y. F. Tan
Department of Electrical and Computer Engineering,
National University of Singapore
Email: vtan@nus.edu.sg

Marco Tomamichel
Centre for Quantum Technologies,
National University of Singapore
Email: cqtmarco@nus.edu.sg

***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 improves on the lower bound by Polyanskiy-Poor-Verdú (2010) and matches the upper bound proved by the same authors.**

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 that the capacity of the AWGN channel [1] is

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

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, Thm. 2]), one sees that

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

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, Thm. 1.2] for discrete memoryless channels (DMCs) and is useful because it provides key insights into the amount of backoff from channel capacity for codes of finite blocklength n. For the AWGN channel, Hayashi [6, Thm. 5] showed that

$$\log M^*(W^n,\varepsilon,P) = n\mathsf{C}(P)+\sqrt{n\mathsf{V}(P)}\,\Phi^{-1}(\varepsilon)+o(\sqrt{n}) \tag{3}$$

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

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

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 $\mathsf{V}(P)$ 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, Thm. 54, Eq. (294)] who showed that

$$\begin{aligned} O(1) &\le \log M^*(W^n,\varepsilon,P) - \Big(n\mathsf{C}(P)+\sqrt{n\mathsf{V}(P)}\,\Phi^{-1}(\varepsilon)\Big) \\ &\le \frac{1}{2}\log n + O(1). \end{aligned} \tag{5}$$

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

Despite these 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, Thm. 1], matching a lower bound by Polyanskiy [9, Thm. 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, Thm. 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 variance[1] is 1, i.e.

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

[1]The 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.

Let $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ 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 $(n, M, \varepsilon, P)_{\rm av}$-*code* for the AWGN channel W is a system $\{(\mathbf{x}(m), \mathcal{D}_m)\}_{m=1}^M$ where $\mathbf{x}(m) \in \mathbb{R}^n, m \in \{1, \ldots, M\}$, 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.

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

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

The Gaussian cumulative distribution function is

$$\Phi(a) := \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{u^2}{2}\Big)\, \mathrm{d}u \tag{8}$$

and we 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 $P \in (0, \infty)$,*

$$\log M^*(W^n, \varepsilon, P) \geq n\mathsf{C}(P) + \sqrt{n\mathsf{V}(P)}\Phi^{-1}(\varepsilon) + \frac{1}{2}\log n + O(1) \tag{9}$$

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

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, Sec. 9.4] in which the capacity is given by the well-known *water-filling* solution. This improves on the result in [9, Thm. 81] by $\frac{1}{2}\log n$. However, this third-order achievability result does not match the converse bound given in [9, Thm. 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.

We also make an observation concerning the relation between prefactors of error exponents and the third-order terms in the normal approximation. In [2], Shannon derived exponential bounds on the error probability of optimal codes over a Gaussian channel using geometric arguments. It can be deduced using Shannon's results and the error exponent of an AWGN channel at high rates [13, Eq. (7.4.33)] $E(R)$ that the average probability of error for a code of size M satisfies

$$\mathrm{P}_{\mathrm{e}}^*(M, n) = \Theta\Big(\frac{\exp(-nE(R))}{n^{(1+|E'(R)|)/2}}\Big), \tag{10}$$

where the rate $R = \frac{1}{n}\log M$. 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, Thm. 1] and [9, Thm. 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, Prop. 1] and [7, Thm. 45]). Also see [18, Thm. 23]. These results suggest a connection between prefactors and third-order terms. The conjecture is that if the prefactor is $\Theta(n^{-1/2})$ (resp. $\Theta(n^{-(1+|E'(R)|)/2})$), then the third-order term is $O(1)$ (resp. $\frac{1}{2}\log n + O(1)$).

IV. Proof of Theorem 1

Because of space constraints, we only provide a sketch of the proof here. A detailed proof is available on the arXiv [19].

A. Random Codebook Generation And Encoding

We first start by defining the random coding distribution

$$f_{\mathbf{X}}(\mathbf{x}) := \frac{\delta(\|\mathbf{x}\|_2^2 - nP)}{S_n(\sqrt{nP})} \tag{11}$$

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 (55). These codewords are denoted as $\mathbf{x}(m) = (x_1(m), \ldots, x_n(m)), m \in \{1, \ldots, M\}$. 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.

$$f_{\mathbf{X}}W^n(\mathbf{y}) := \int_{\mathbf{x}'} f_{\mathbf{X}}(\mathbf{x}')W^n(\mathbf{y}|\mathbf{x}')\, \mathrm{d}\mathbf{x}'. \tag{12}$$

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

$$q(\mathbf{x}(m), \mathbf{y}) > \max_{\tilde{m} \in \{1,\ldots,M\}\setminus\{m\}} q(\mathbf{x}(\tilde{m}), \mathbf{y}), \tag{13}$$

where the decoding metric is the log-likelihood ratio

$$q(\mathbf{x}, \mathbf{y}) := \log \frac{W^n(\mathbf{y}|\mathbf{x})}{f_{\mathbf{X}}W^n(\mathbf{y})}. \tag{14}$$

If there is no unique $m \in \{1, \ldots, M\}$ satisfying (13), declare an error. (This happens with probability zero.)

Since the denominator in (14), 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

$$q(\mathbf{x}, \mathbf{y}) = \frac{n}{2}\log\frac{1}{2\pi} + \langle \mathbf{x}, \mathbf{y}\rangle - nP - \|\mathbf{y}\|_2^2 - \log f_{\mathbf{X}}W^n(\mathbf{y}) \tag{15}$$

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 $\|\mathbf{y}\|_2^2$.

C. The Random Coding Union (RCU) Bound

All the randomly drawn codewords satisfy the cost constraints with probability one. The RCU bound [7, Thm. 16] states that there exists an $(n, M, \varepsilon', P)_{\text{av}}$-code satisfying

$$\varepsilon' \le \mathbb{E}\left[\min\left\{1, M \Pr\left(q(\bar{\mathbf{X}}, \mathbf{Y}) \ge q(\mathbf{X}, \mathbf{Y}) | \mathbf{X}, \mathbf{Y}\right)\right\}\right] \quad (16)$$

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

$$g(t, \mathbf{y}) := \Pr\left(q(\bar{\mathbf{X}}, \mathbf{Y}) \ge t \,\middle|\, \mathbf{Y} = \mathbf{y}\right). \quad (17)$$

Since $\bar{\mathbf{X}}$ is independent of $\mathbf{X}$, the probability in (16) can be written as

$$\Pr\left(q(\bar{\mathbf{X}}, \mathbf{Y}) \ge q(\mathbf{X}, \mathbf{Y}) | \mathbf{X}, \mathbf{Y}\right) = g(q(\mathbf{X}, \mathbf{Y}), \mathbf{Y}). \quad (18)$$

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

$$f_{\mathbf{X}}(\bar{\mathbf{x}}) = f_{\mathbf{X}}(\bar{\mathbf{x}}) \frac{f_{\mathbf{X}|\mathbf{Y}}(\bar{\mathbf{x}}|\mathbf{y})}{f_{\mathbf{X}|\mathbf{Y}}(\bar{\mathbf{x}}|\mathbf{y})} = f_{\mathbf{X}|\mathbf{Y}}(\bar{\mathbf{x}}|\mathbf{y}) \exp(-q(\bar{\mathbf{x}}, \mathbf{y})). \quad (19)$$

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

$$g(t, \mathbf{y}) = \mathbb{E}\left[\exp(-q(\mathbf{X}, \mathbf{Y})) \mathbf{1}\{q(\mathbf{X}, \mathbf{Y}) \ge t\} \,\middle|\, \mathbf{Y} = \mathbf{y}\right]. \quad (20)$$

D. A High-Probability Set

Consider the set of "typical" channel outputs whose ℓ_2-norms are approximately $\sqrt{n(P+1)}$. More precisely, define

$$\mathcal{F} := \left\{\mathbf{y} \in \mathbb{R}^n : \frac{1}{n}\|\mathbf{y}\|_2^2 \in [P+1-\delta, P+1+\delta]\right\}. \quad (21)$$

It is easy to show via Chernoff bounds that

$$\Pr(\mathbf{Y} \in \mathcal{F}) \ge 1 - \xi_n =: 1 - \exp(-\kappa n \delta^2) \quad (22)$$

where $\kappa = \kappa(P) > 0$. We take $\delta = n^{-1/3}$ throughout which means that $\xi_n = \exp(-\kappa n^{1/3})$ and this sequence decays faster than any polynomial.

E. Probability That The Decoding Metric Lies In An Interval

We would like to upper bound $g(t, \mathbf{y})$ in (17) to evaluate the RCU bound. This we do in the next section. As an intermediate step, we consider the problem of upper bounding

$$h(\mathbf{y}; a, \mu) := \Pr\left(q(\mathbf{X}, \mathbf{Y}) \in [a, a+\mu] \,\middle|\, \mathbf{Y} = \mathbf{y}\right), \quad (23)$$

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 (15) as

$$h(\mathbf{y}; a, \mu) := \Pr\left(\langle \mathbf{X}, \mathbf{Y} \rangle \in [a', a'+\mu] \,\middle|\, \mathbf{Y} = \mathbf{y}\right), \quad (24)$$

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),

$$h(s; a, \mu) := h(\mathbf{y}; a, \mu), \quad \text{if} \quad s = \frac{1}{n}\|\mathbf{y}\|_2^2. \quad (25)$$

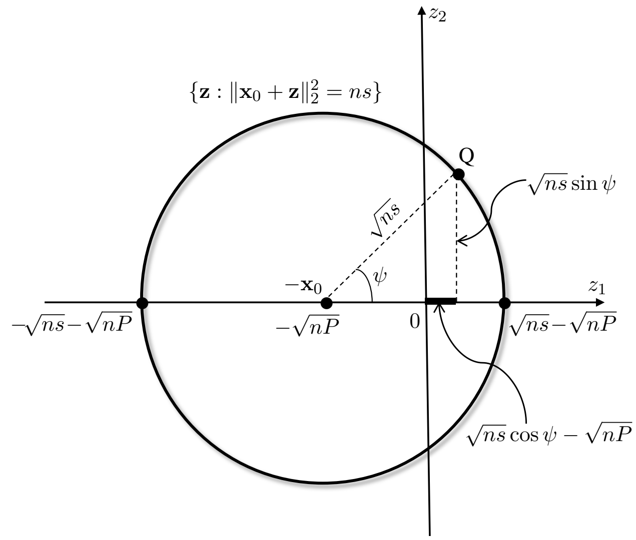

Fig. 1. Illustration of the relation between Z_1 and Ψ in (30) in two dimensions. The transformation of this figure to the U coordinate system via (32) translates the sphere to the origin and scales its radius to be 1.

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 $\mathbf{Z} = (Z_1, \ldots, Z_n) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n \times n})$, we have

$$\begin{aligned} &h(s; a, \mu) \\ &= \Pr\left(\langle \mathbf{X}, \mathbf{X} + \mathbf{Z} \rangle \in [a', a'+\mu] \,\middle|\, \|\mathbf{X} + \mathbf{Z}\|_2^2 = ns\right) \quad (26) \\ &= \Pr\left(\sum_{i=1}^{n} X_i Z_i + nP \in [a', a'+\mu] \,\middle|\, \|\mathbf{X} + \mathbf{Z}\|_2^2 = ns\right) \quad (27) \end{aligned}$$

where (27) follows by the observation that $\langle \mathbf{X}, \mathbf{X} \rangle = nP$ with probability one. Now, define

$$\mathbf{x}_0 := \left(\sqrt{nP}, 0, \ldots, 0\right) \quad (28)$$

to be a fixed vector on the power sphere. By spherical symmetry, we may pick $\mathbf{X}$ in (27) to be $\mathbf{x}_0$. Thus, we have

$$h(s; a, \mu) = \Pr\left(Z_1 + \sqrt{nP} \in \left[\frac{a'}{\sqrt{nP}}, \frac{a'+\mu}{\sqrt{nP}}\right] \,\middle|\, \|\mathbf{x}_0 + \mathbf{Z}\|_2^2 = ns\right). \quad (29)$$

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 $-\mathbf{x}_0 = (-\sqrt{nP}, 0, \ldots, 0)$. 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 $\mu/\sqrt{n}$.

Let us now derive the conditional density of Z_1 given $\mathcal{E} := \{\|\mathbf{x}_0 + \mathbf{Z}\|_2^2 = ns\}$. 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 $\mathcal{E}$ where Ψ and Z_1 are related as follows:

$$Z_1 = \sqrt{ns}\cos\Psi - \sqrt{nP}. \quad (30)$$

Again see Fig. 1. Now, we have

$$f_{\Psi|\mathcal{E}}(\psi) \propto \left(\sin^{n-2}\psi\right)\exp\left(-\frac{n}{2}\left[(\sqrt{s}\cos\psi-\sqrt{P})^2+s\sin^2\psi\right]\right). \quad (31)$$

This follows because the area element (an $(n-1)$-dimensional annulus of radius $\sqrt{ns}\sin\psi$ and width $\mathrm{d}\psi$) is proportional to $\sin^{n-2}\psi$ (similar to Shannon's derivation in [2, Eq. (21)]) and the Gaussian weighting is proportional to $\exp\big(-\frac{n}{2}\big[(\sqrt{s}\cos\psi-\sqrt{P})^2+s\sin^2\psi\big]\big)$. 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

$$U = \cos\Psi \quad (32)$$

resulting in the conditional density of U given $\mathcal{E}$ being

$$f_{U|\mathcal{E}}(u) = \frac{1}{F_n}(1-u^2)^{(n-3)/2}\exp\left(n\sqrt{Ps}u\right)\mathbf{1}\{u\in[-1,1]\}, \quad (33)$$

where the normalization constant is

$$F_n := \int_{-1}^{1}(1-u^2)^{(n-3)/2}\exp\left(n\sqrt{Ps}u\right)\mathrm{d}u. \quad (34)$$

The conditional density we have derived in (33)–(34) 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 exploits Laplace's method [21], [22] for approximating integrals of the form $\int_{[a,b]}\exp(n\alpha(u))\,\mathrm{d}u$ for some smooth function $\alpha(u)$.

Lemma 2. *Define the function*

$$L(P,s) := \frac{(2Ps)^2}{\sqrt{2\pi}}\cdot\sqrt{\frac{1+4Ps-\sqrt{1+4Ps}}{(\sqrt{1+4Ps}-1)^5}}. \quad (35)$$

The following bound holds:

$$\limsup_{n\to\infty}\frac{1}{\sqrt{n}}\sup_{u\in[-1,1]} f_{U|\mathcal{E}}(u) \le L(P,s). \quad (36)$$

Equipped with this lemma, let us consider the probability $h(s;a,\mu)$ in (29). We have

$$\begin{aligned} h(s;a,\mu) &= \Pr\left(\sqrt{ns}\,U\in\Big[\frac{a'}{\sqrt{nP}},\frac{a'+\mu}{\sqrt{nP}}\Big]\,\Big|\,\mathcal{E}\right) && (37)\\ &= \int_{a'/(n\sqrt{Ps})}^{(a'+\mu)/(n\sqrt{Ps})} f_{U|\mathcal{E}}(u)\,\mathrm{d}u && (38)\\ &\le \int_{a'/(n\sqrt{Ps})}^{(a'+\mu)/(n\sqrt{Ps})} 2\,L(P,s)\sqrt{n}\,\mathrm{d}u && (39)\\ &= \frac{2\,L(P,s)\,\mu}{\sqrt{nPs}}, && (40)\end{aligned}$$

where (37) follows from the fact that $Z_1=\sqrt{ns}\,U-\sqrt{nP}$ due to (30) and (32), and (39) 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 $\delta=n^{-1/3}\to 0$, we deduce that for all $\mathbf{y}\in\mathcal{F}$ and n sufficiently large (depending only on P),

$$h(\mathbf{y};a,\mu)\le K(P)\cdot\frac{\mu}{\sqrt{n}}, \quad (41)$$

for some function $K(P)$.

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

We now return to bounding $g(t,\mathbf{y})$ defined in (17). Again, we assume $\mathbf{y}\in\mathcal{F}$. The idea here is to consider the second form of $g(t,\mathbf{y})$ in (20) 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 (41) to each segment. This is modelled after the proof of [7, Lem. 47]. Indeed, we have

$$\begin{aligned} g(t,\mathbf{y}) &= \mathbb{E}\big[\exp(-q(\mathbf{X},\mathbf{Y}))\mathbf{1}\{q(\mathbf{X},\mathbf{Y})\ge t\}\,\big|\,\mathbf{Y}=\mathbf{y}\big]\\ &\le \sum_{l=0}^{\infty}\exp(-t-l\eta)\\ &\qquad\times\Pr\left(t+l\eta\le q(\mathbf{X},\mathbf{Y})<t+(l+1)\eta\,\big|\,\mathbf{Y}=\mathbf{y}\right) && (42)\\ &\le \sum_{l=0}^{\infty}\exp(-t-l\eta)\cdot\frac{K(P)\,\eta}{\sqrt{n}} && (43)\\ &= \frac{\exp(-t)}{1-\exp(-\eta)}\cdot\frac{K(P)\,\eta}{\sqrt{n}}. && (44)\end{aligned}$$

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

$$g(t,\mathbf{y})\le\frac{G\,\exp(-t)}{\sqrt{n}} \quad (45)$$

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 (16). Consider,

$$\begin{aligned}\varepsilon' &\le \mathbb{E}\left[\min\left\{1,Mg(q(\mathbf{X},\mathbf{Y}),\mathbf{Y})\right\}\right]\\ &\le \Pr(\mathbf{Y}\in\mathcal{F}^c)\\ &\quad+\mathbb{E}\Big[\min\left\{1,Mg(q(\mathbf{X},\mathbf{Y}),\mathbf{Y})\right\}\,\Big|\,\mathbf{Y}\in\mathcal{F}\Big]\cdot\Pr(\mathbf{Y}\in\mathcal{F}). && (46)\end{aligned}$$

The first term is bounded above by ξ_n and the second can be bounded above by

$$\mathbb{E}\left[\min\left\{1,\frac{MG\exp(-q(\mathbf{X},\mathbf{Y}))}{\sqrt{n}}\right\}\,\Big|\,\mathbf{Y}\in\mathcal{F}\right]\cdot\Pr(\mathbf{Y}\in\mathcal{F}) \quad (47)$$

due to (45) with $t=q(\mathbf{X},\mathbf{Y})$. Now we split the expectation into two parts depending on whether $q(\mathbf{x},\mathbf{y})>\log(MG/\sqrt{n})$ or otherwise, i.e.

$$\begin{aligned}&\mathbb{E}\left[\min\left\{1,\frac{MG\exp(-q(\mathbf{X},\mathbf{Y}))}{\sqrt{n}}\right\}\,\Big|\,\mathbf{Y}\in\mathcal{F}\right]\\ &\le\Pr\left(q(\mathbf{X},\mathbf{Y})\le\log\frac{MG}{\sqrt{n}}\,\Big|\,\mathbf{Y}\in\mathcal{F}\right)\\ &+\frac{MG}{\sqrt{n}}\mathbb{E}\left[\mathbf{1}\left\{q(\mathbf{X},\mathbf{Y})>\log\frac{MG}{\sqrt{n}}\right\}\exp(-q(\mathbf{X},\mathbf{Y}))\Big|\mathbf{Y}\in\mathcal{F}\right]. && (48)\end{aligned}$$

By applying (45) with $t = \log(MG/\sqrt{n})$, we know that the second term can be bounded above by $G/\sqrt{n}$.

Now let $f_Y^*(y) = \mathcal{N}(y; 0, P+1)$ be the capacity-achieving output distribution and $f_{\mathbf{Y}}^*(\mathbf{y}) = \prod_{i=1}^n f_Y^*(y_i)$ its n-fold memoryless extension. In Step 1 of the proof of Lem. 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 $f_{\mathbf{Y}}^*(\mathbf{y})$ can be bounded by a finite constant J, i.e.

$$\sup_{\mathbf{y}\in\mathcal{F}} \frac{f_{\mathbf{X}} W^n(\mathbf{y})}{f_{\mathbf{Y}}^*(\mathbf{y})} \le J. \tag{49}$$

Also see [23, Prop. 2]. We return to bounding the first term in (48). Using the definition of $q(\mathbf{x},\mathbf{y})$ in (14) and applying the bound in (49) yields

$$\begin{aligned}
&\Pr\left(q(\mathbf{X},\mathbf{Y}) \le \log \frac{MG}{\sqrt{n}} \,\middle|\, \mathbf{Y}\in\mathcal{F} \right) \\
&= \Pr\left(\log \frac{W^n(\mathbf{Y}|\mathbf{X})}{f_{\mathbf{X}} W^n(\mathbf{Y})} \le \log \frac{MG}{\sqrt{n}} \,\middle|\, \mathbf{Y}\in\mathcal{F} \right) \qquad (50) \\
&\le \Pr\left(\log \frac{W^n(\mathbf{Y}|\mathbf{X})}{f_{\mathbf{Y}}^*(\mathbf{Y})} \le \log \frac{MGJ}{\sqrt{n}} \,\middle|\, \mathbf{Y}\in\mathcal{F} \right). \qquad (51)
\end{aligned}$$

Thus, when we multiply the first term in (48) by $\Pr(\mathbf{Y}\in\mathcal{F})$, use Bayes rule and drop the event $\{\mathbf{Y}\in\mathcal{F}\}$, we see that the resultant product can be bounded above by

$$\Pr\left(\log \frac{W^n(\mathbf{Y}|\mathbf{X})}{f_{\mathbf{Y}}^*(\mathbf{Y})} \le \log \frac{MGJ}{\sqrt{n}} \right). \tag{52}$$

By spherical symmetry, we may choose $\mathbf{X} = (\sqrt{P}, \ldots, \sqrt{P})$. Working out the statistics of the information density and using the Berry-Esseen theorem [24, Sec. XVI.5], we see that

$$\begin{aligned}
&\Pr\left(\log \frac{W^n(\mathbf{Y}|\mathbf{X})}{f_{\mathbf{Y}}^*(\mathbf{Y})} \le \log \frac{MGJ}{\sqrt{n}} \,\middle|\, \mathbf{X} = (\sqrt{P}, \ldots, \sqrt{P}) \right) \\
&\quad \le \Phi\left(\frac{\log \frac{MGJ}{\sqrt{n}} - n\mathsf{C}(P)}{\sqrt{n\mathsf{V}(P)}} \right) + \frac{B}{\sqrt{n}}, \qquad (53)
\end{aligned}$$

where B is a constant that just depends on P.

Putting all the bounds together, we obtain

$$\varepsilon' \le \Phi\left(\frac{\log \frac{MGJ}{\sqrt{n}} - n\mathsf{C}(P)}{\sqrt{n\mathsf{V}(P)}} \right) + \frac{B}{\sqrt{n}} + \frac{G}{\sqrt{n}} + \xi_n. \tag{54}$$

Now choose

$$\begin{aligned}
\log M = n\mathsf{C}(P) + \sqrt{n\mathsf{V}(P)}\Phi^{-1}\left(\varepsilon - \frac{B+G}{\sqrt{n}} - \xi_n \right) \\
+ \frac{1}{2}\log n - \log(GJ) \qquad (55)
\end{aligned}$$

ensuring that $\varepsilon' \le \varepsilon$. Hence, there exists an $(n, M, \varepsilon, P)_{\mathrm{av}}$-code with M in (55). By Taylor expansions, this $\log M$ equals the lower bound in (9). This completes the proof. □

V. Future Work

In this work, it is shown that the third-order term in the asymptotic expansion for the AWGN channel is $\frac{1}{2}\log n$. This holds under the average error setting. It is not known if the same is true under the more stringent maximum error setting primarily because the RCU bound only holds under average error. This is left for future research.

Acknowledgements: VT sincerely thanks Shaowei Lin (I^{2}R, A*STAR) for many helpful explanations concerning approximation of integrals. The authors also thank Jonathan Scarlett (Cambridge) and Yücel Altuğ (Cornell) for discussions and constructive comments.

References

[1] C. E. Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423, 1948.

[2] C. E. Shannon. Probability of error for optimal codes in a Gaussian channel. *Bell Systems Technical Journal*, 38:611–656, 1959.

[3] K. Yoshihara. Simple proofs for the strong converse theorems in some channels. *Kodai Mathematical Journal*, 16(4):213–222, 1964.

[4] J. Wolfowitz. The coding of messages subject to chance errors. *Illinois Journal of Mathematics*, 1(4):591–606, 1957.

[5] V. Strassen. Asymptotische Abschätzungen in Shannons Informationstheorie. In *Trans. Third Prague Conf. Inf. Theory*, pages 689–723, Prague, 1962.

[6] M. Hayashi. Information spectrum approach to second-order coding rate in channel coding. *IEEE Trans. on Inf. Th.*, 55:4947–4966, Nov 2009.

[7] Y. Polyanskiy, H. V. Poor, and S. Verdú. Channel coding rate in the finite blocklength regime. *IEEE Trans. on Inf. Th.*, 56:2307–2359, 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.*, 59(11):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 edition, 2006.

[13] R. G. Gallager. *Information Theory and Reliable Communication*. Wiley, New York, 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.*, 59:2576–2595, May 2013.

[19] V. Y. F. Tan and M. Tomamichel. The third-order term in the normal approximation for the AWGN channel. `arXiv:1311.2337 [cs.IT]`, Nov 2013.

[20] A. J. Stam. Limit theorems for uniform distributions on spheres in high-dimensional Euclidean spaces. *Journal of Applied Probability*, 19(1):221–228, 1982.

[21] L. Tierney and J. B. Kadane. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86, Mar 1986.

[22] Z. Shun and P. McCullagh. Laplace approximation of high dimensional integrals. *Journal of the Royal Statistical Society, Series B (Methodology)*, 57(4):749–760, 1995.

[23] E. MolavianJazi and J. N. Laneman. A finite-blocklength perspective on Gaussian multi-access channels. `arXiv:1309.2343 [cs.IT]`, Sep 2013.

[24] W. Feller. *An Introduction to Probability Theory and Its Applications*. John Wiley and Sons, 2nd edition, 1971.