Channel Coding for Gaussian Channels with Mean and Variance Constraints

Adeel Mahmood and Aaron B. Wagner

School of Electrical and Computer Engineering, Cornell University

Abstract

We consider channel coding for Gaussian channels with the recently introduced mean and variance cost constraints. Through matching converse and achievability bounds, we characterize the optimal first- and second-order performance. The main technical contribution of this paper is an achievability scheme which uses random codewords drawn from a mixture of three uniform distributions on $(n-1)$-spheres of radii R_1, R_2 and R_3, where $R_i = O(\sqrt{n})$ and $|R_i - R_j| = O(1)$. To analyze such a mixture distribution, we prove a lemma giving a uniform $O(\log n)$ bound, which holds with high probability, on the log ratio of the output distributions Q_i^{cc} and Q_j^{cc}, where Q_i^{cc} is induced by a random channel input uniformly distributed on an $(n-1)$-sphere of radius R_i. To facilitate the application of the usual central limit theorem, we also give a uniform $O(\log n)$ bound, which holds with high probability, on the log ratio of the output distributions Q_i^{cc} and Q_i^*, where Q_i^* is induced by a random channel input with i.i.d. components.

Index Terms

Channel coding, Gaussian channels, second-order coding rate, random coding, mixture distribution

I. INTRODUCTION

The two common forms of cost (or power) constraints in channel coding have been the maximal cost constraint specified by $c(\mathbf{X}) \leq \Gamma$ almost surely, or the expected cost constraint specified by $\mathbb{E}[c(\mathbf{X})] \leq \Gamma$, where $\mathbf{X} \in \mathbb{R}^n$ is a random channel input vector and $c(\cdot)$ is an additively separable cost function defined as

$$c(\mathbf{X}) := \frac{1}{n} \sum_{i=1}^{n} c(X_i).$$

Recent works [1] and [2] introduced the mean and variance (m.v.) cost constraint, specified by

$$\mathbb{E}[c(\mathbf{X})] \leq \Gamma$$
$$\text{Var}(c(\mathbf{X})) \leq \frac{V}{n},$$

for discrete memoryless channels (DMCs). With a variance parameter V, the mean and variance (m.v.) cost constraint generalizes the two existing frameworks in the sense that $V \to 0$ recovers the first- and second-order coding performance of the maximal cost constraint [1] and $V \to \infty$ recovers the expected cost constraint. Beyond generalization, the m.v. cost constraint for $0 < V < \infty$ is shown to have practical advantages over both prior cost models. Unlike the maximal cost constraint, it allows for an improved second-order coding performance with feedback [1, Theorem 3], [2, Theorem 2]; even without feedback, the coding performance under the m.v. cost constraint is superior [2, Theorem 1]. In particular, for DMCs with a unique capacity-cost-achieving distribution, second-order coding performance improvement via feedback is possible if and only if $V > 0$. Unlike the expected cost constraint, the m.v. cost constraint enforces a controlled, ergodic use of transmission power [2, (5)]. This is essential for several practical reasons such as operating circuitry in the linear regime, minimizing power consumption, and reducing interference with other terminals. The m.v. cost constraint allows the power to fluctuate above the threshold in a manner consistent with a noise process, thus making it a more realistic and natural cost model in practice than the restrictive maximal cost constraint.

The aforementioned performance improvements under the m.v. cost constraint were shown for DMCs only, although the ergodic behavior enforced by the m.v. cost constraint is generally true. In this paper, we investigate additive white Gaussian noise (AWGN) channels with the mean and variance cost constraint with the cost function taken to be $c(x) = x^2$. Specifically, we characterize the optimal first- and second-order coding rates under the m.v. cost constraint for AWGN channels subject to a non-vanishing average probability of error $\epsilon > 0$. Simultaneously, we also characterize the optimal average error probability as a function of the second-order rate with the baseline first-order rate fixed as the capacity-cost function [3, (9.17)]. The latter is motivated by the fact, which itself follows from our results, that the strong converse [3, p. 208] holds under the m.v. cost constraint. This is an interesting finding since the strong converse does not hold for AWGN channels subject to expectation-only constraint [4, Theorem 77]. However, if one considers maximal probability of error instead of average probability of error, then the strong converse holds for both m.v. cost constraint and expectation-only constraint [5, (18)]. More results on

AWGN channels can be found in [6] and [7]. In this paper, we only focus on the average error probability and non-feedback framework.

Our achievability proof relies on a random coding scheme where the channel input has a distribution supported on at most three concentric spheres of radii R_1, R_2 and R_3, where $R_i = O(\sqrt{n})$ and $|R_i - R_j| = O(1)$. Each uniform distribution on a sphere induces a distribution Q_i^{cc} on the channel output, and a mixture of such input uniform distributions induces a mixture distribution of Q_1^{cc}, Q_2^{cc} and Q_3^{cc} on the output. For $\mathbf{Y} \sim Q_i^{cc}$, we have that $\mathbb{E}[||\mathbf{Y}||^2] = O(n)$ and $||\mathbf{Y}||^2$ concentrates in probability over a set encompassing $\pm\sqrt{n}$ deviations around its mean. Furthermore, the probability density function Q_i^{cc} is given in terms of the modified Bessel function of the first kind. To assist in the analysis of the mixture distribution, we use a uniform asymptotic expansion of the Bessel function followed by traditional series expansions to bound the log ratio of Q_i^{cc} and Q_j^{cc}. Remarkably, the zeroth- and first-order terms, which are $O(n)$ and $O(\sqrt{n})$ respectively, cancel out, giving us an $O(\log n)$ bound. This $O(\log n)$ bounds holds uniformly over a set on which $||\mathbf{Y}||^2$ concentrates. Moreover, to facilitate the application of the central limit theorem, we give a similar $O(\log n)$ bound on the log ratio of Q_i^{cc} and Q^*, where Q^* is the output distribution induced by an i.i.d. channel input. The discussion of this paragraph is formalized in Lemmas 5, 6 and 7 in Section III. The achievability theorem and its proof are also given in Section III.

For the proof of the matching converse result, the main technical component involves obtaining convergence of the distribution of the normalized sum of information densities to a standard Gaussian CDF. Let $\mathbf{Y} = (Y_1, \ldots, Y_n)$ denote the random channel output when the input to the AWGN channel, denoted by W, is some fixed vector $\mathbf{x} = (x_1, \ldots, x_n)$. The sum of the information densities is given by

$$\sum_{i=1}^{n} \log \frac{W(Y_i|x_i)}{Q^*(Y_i)}. \tag{1}$$

A well-known observation is that the distribution of (1) depends on $\mathbf{x}$ only through its average power $c(\mathbf{x}) = ||\mathbf{x}||^2/n$ (see Lemma 2). Unlike the maximal cost constraint, $c(\mathbf{x})$ is not uniformly bounded in the mean and variance cost constraint framework. Hence, we prove that the normalized sum converges uniformly in distribution to the standard Gaussian, where the convergence is only uniform over typical vectors $\mathbf{x} \in \mathbb{R}^n$. For atypical vectors, we use standard concentration arguments. Section IV is devoted to the converse theorem and its proof.

II. PRELIMINARIES

We write $\mathbf{x} = (x_1, \ldots, x_n)$ to denote a vector and $\mathbf{X} = (X_1, \ldots, X_n)$ to denote a random vector in $\mathbb{R}^n$. For any $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, let $\mathcal{N}(\mu, \sigma^2)$ denote the Gaussian distribution with mean μ and variance σ^2. Let Φ denote the standard Gaussian CDF. Let $\chi_n^2(\lambda)$ denote the noncentral chi-squared distribution with n degrees of freedom and noncentrality parameter λ. If two random variables X and Y have the same distribution, we write $X \stackrel{d}{=} Y$. The modified Bessel function of the first kind of order ν is denoted by $I_\nu(x)$. We will write $\log$ to denote logarithm to the base e and $\exp(x)$ to denote e to the power of x. Define S_R^{n-1} to be the $(n-1)$-sphere of radius R centered at the origin, i.e., $S_R^{n-1} := \{\mathbf{x} \in \mathbb{R}^n : ||\mathbf{x}|| = R\}$.

Let $\mathcal{P}(\mathbb{R}^n)$ denote the set of all probability distributions over $\mathbb{R}^n$. If $P \in \mathcal{P}(\mathbb{R}^n)$ is an n-fold product distribution induced by some $P' \in \mathcal{P}(\mathbb{R})$, then we write

$$P(\mathbf{x}) = P'(\mathbf{x}) = \prod_{i=1}^{n} P'(x_i)$$

with some abuse of notation.

The additive white Gaussian noise (AWGN) channel models the relationship between the channel input $\mathbf{X}$ and output $\mathbf{Y}$ over n channel uses as $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$, where $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, N \cdot I_n)$ represents independent and identically distributed (i.i.d.) Gaussian noise with variance $N > 0$. The noise vector $\mathbf{Z}$ is independent of the input $\mathbf{X}$. For a single time step, we use W to denote the conditional probability distribution associated with the channel:

$$W(\cdot|x) := \mathcal{N}(x, N).$$

Let the cost function $c : \mathbb{R} \to [0, \infty)$ be given by $c(x) = x^2$. For a channel input sequence $\mathbf{x} \in \mathbb{R}^n$,

$$c(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} c(x_i) = \frac{||\mathbf{x}||^2}{n}.$$

For $\Gamma > 0$, the capacity-cost function is defined as

$$C(\Gamma) = \max_{P:\mathbb{E}_P[c(X)] \leq \Gamma} I(P, W), \tag{2}$$

where

$$\mathbb{E}_P[c(X)] = \int_{\mathbb{R}} x^2 dP(x).$$

Definition 1: We use P^* to denote the capacity-cost-achieving distribution in (2) and $Q^* = P^*W$ to denote the induced output distribution. We define

$$\nu_x := \mathrm{Var}\left(\log \frac{W(Y|x)}{Q^*(Y)}\right), \quad \text{where } Y \sim W(\cdot|x),$$

$$V(\Gamma) := \int_{\mathbb{R}} \nu_x P^*(x)dx.$$

Lemma 1: We have $P^* = \mathcal{N}(0, \Gamma)$ and $Q^* = \mathcal{N}(0, \Gamma + N)$. Thus, the capacity-cost function is given by

$$C(\Gamma) = \frac{1}{2}\log\left(1 + \frac{\Gamma}{N}\right).$$

Furthermore, for any $x \in \mathbb{R}$ and $Y \sim W(\cdot|x)$,

$$\mathbb{E}\left[\log \frac{W(Y|x)}{Q^*(Y)}\right] = C(\Gamma) - C'(\Gamma)(\Gamma - c(x)) \tag{3}$$

$$C'(\Gamma) = \frac{1}{2(\Gamma + N)}$$

$$\nu_x = \frac{\Gamma^2 + 2x^2 N}{2(N + \Gamma)^2}$$

$$V(\Gamma) = \frac{\Gamma^2 + 2\Gamma N}{2(N + \Gamma)^2}. \tag{4}$$

Proof: The fact that $P^* = \mathcal{N}(0, \Gamma)$ can be found in [3, (9.17)]. The remaining content of Lemma 1 follows from elementary calculus.

Lemma 2 (Spherical Symmetry): Let $\mathbf{x} \in \mathbb{R}^n$ be the channel input. Let $Y_i = x_i + Z_i$. Define

$$T_i := \log \frac{W(Y_i|x_i)}{Q^*(Y_i)} - \mathbb{E}\left[\log \frac{W(Y_i|x_i)}{Q^*(Y_i)}\right].$$

Then

$$\sum_{i=1}^n T_i \overset{d}{=} -\frac{\Gamma}{2(\Gamma + N)}\Lambda + \frac{Nnc(\mathbf{x})}{2\Gamma(\Gamma + N)} + \frac{n\Gamma}{2(\Gamma + N)},$$

where $\Lambda \sim \chi_n^2\left(\frac{Nnc(\mathbf{x})}{\Gamma^2}\right)$. Hence, the distribution of $\sum_{i=1}^n T_i$ depends on $\mathbf{x}$ only through its cost $c(\mathbf{x})$. Hence, we can write

$$\sum_{i=1}^n T_i \overset{d}{=} \sum_{i=1}^n \tilde{T}_i \tag{5}$$

where $\tilde{T}_i$'s are i.i.d. and

$$\tilde{T}_i \overset{d}{=} \log \frac{W(Y|\sqrt{c(\mathbf{x})})}{Q^*(Y)} - \mathbb{E}\left[\log \frac{W(Y|\sqrt{c(\mathbf{x})})}{Q^*(Y)}\right], \tag{6}$$

where $Y \sim \mathcal{N}(\sqrt{c(\mathbf{x})}, N)$ in (6).

Proof: The observation of spherical symmetry with respect to the channel input is standard in most works on AWGN channels (see, e.g., [8]). For convenience and completeness, we have given a proof of this lemma in Appendix A.

Corollary 1: With the same setup as in Lemma 2, we have

$$\mathrm{Var}\left(\sum_{i=1}^n T_i\right) = \frac{n\Gamma^2 + 2Nnc(\mathbf{x})}{2(\Gamma + N)^2}.$$

Proof: Use the equality in distribution given in (5) and Lemma 1.

With a blocklength n and a fixed rate $R > 0$, let $\mathcal{M}_R = \{1, \ldots, \lceil \exp(nR) \rceil\}$ denote the message set. Let $M \in \mathcal{M}_R$ denote the random message drawn uniformly from the message set.

Definition 2: An (n, R) code for an AWGN channel consists of an encoder f which, for each message $m \in \mathcal{M}_R$, chooses an input $\mathbf{X} = f(m) \in \mathbb{R}^n$, and a decoder g which maps the output $\mathbf{Y}$ to $\hat{m} \in \mathcal{M}_R$. The code (f, g) is random if f or g is random.

Definition 3: An (n, R, Γ, V) code for an AWGN channel is an (n, R) code such that $\mathbb{E}[c(\mathbf{X})] \leq \Gamma$ and $\mathrm{Var}(c(\mathbf{X})) \leq V/n$, where the message $M \sim \mathrm{Unif}(\mathcal{M}_R)$ has a uniform distribution over the message set $\mathcal{M}_R$.

Given $\epsilon \in (0, 1)$, define

$$M^*(n, \epsilon, \Gamma, V) := \max\{\lceil \exp(nR) \rceil : \bar{P}_e(n, R, \Gamma, V) \le \epsilon\},$$

where $\bar{P}_e(n, R, \Gamma, V)$ denotes the minimum average error probability attainable by any random (n, R, Γ, V) code.

Definition 4: Define the function $\mathcal{K} : \mathbb{R} \times [0, \infty) \to (0, 1)$ as

$$\mathcal{K}(r, V) = \inf_{\substack{\Pi: \\ \mathbb{E}[\Pi] = r \\ \mathrm{Var}(\Pi) \le V}} \mathbb{E}[\Phi(\Pi)]. \tag{7}$$

Lemma 3: The $\mathcal{K}$ function has the following properties:

1) The infimum in (7) is a minimum, and there exists a minimizer which is a discrete probability distribution with at most 3 point masses,
2) $\mathcal{K}(r, V)$ is (jointly) continuous in (r, V),
3) for any fixed V, $\mathcal{K}(r, V)$ is strictly increasing in r,
4) for any V and $0 < \epsilon < 1$, there always exists a unique r^* satisfying $\mathcal{K}(r^*, V) = \epsilon$,
5) for all $r \in \mathbb{R}$ and $V > 0$, the minimizing distribution in (7) has at least two point masses.

Proof: See [9, Lemmas 3 and 4] for the first four properties and [2, Theorem 1] for the fifth property above.

III. ACHIEVABILITY RESULT

The design of random channel codes (f, g) can be directly related to the design of the distribution $P \in \mathcal{P}(\mathbb{R}^n)$ of the codewords. Specifically, for each message $m \in \mathcal{M}_R$, the channel input $\mathbf{X}$ is chosen randomly according to P. Given an observed $\mathbf{y}$ at the decoder g, the decoder selects the message m with the lowest index that achieves the maximum over m of

$$\prod_{i=1}^{n} W(y_i | f(m)).$$

Any distribution $P \in \mathcal{P}(\mathbb{R}^n)$ can be used to construct an (n, R) channel code using the aforementioned construction. Moreover, any distribution P satisfying

$$\begin{aligned} \mathbb{E}[c(\mathbf{X})] &\le \Gamma \\ \mathrm{Var}(c(\mathbf{X})) &\le \frac{V}{n} \end{aligned} \tag{8}$$

can be used to construct a random (n, R, Γ, V) code.

Given a random (n, R, Γ, V) code based on an input distribution P, the following lemma gives an upper bound on the average error probability of the code in terms of the distribution of $(\mathbf{X}, \mathbf{Y})$ induced by P and the channel transition probability W.

Lemma 4: Consider an AWGN channel W with noise variance $N > 0$ and cost constraint $(\Gamma, V) \in (0, \infty) \times [0, \infty)$. For any distribution $P \in \mathcal{P}(\mathbb{R}^n)$ satisfying (8), and any n, θ and R,

$$\bar{P}_e(n, R, \Gamma, V) \le (P \circ W)\left(\frac{1}{n} \log \frac{W(\mathbf{Y}|\mathbf{X})}{PW(\mathbf{Y})} \le R + \theta\right) + e^{-n\theta}, \tag{9}$$

where $(\mathbf{X}, \mathbf{Y})$ have the joint distribution specified by

$$(P \circ W)(\mathbf{x}, \mathbf{y}) = P(\mathbf{x}) \prod_{k=1}^{n} W(y_k | x_k),$$

and PW denotes the marginal distribution of $\mathbf{Y}$. Furthermore, if for some α and ϵ,

$$\limsup_{n \to \infty} (P \circ W)\left(\frac{1}{n} \log \frac{W(\mathbf{Y}|\mathbf{X})}{PW(\mathbf{Y})} \le C(\Gamma) + \frac{\alpha}{\sqrt{n}}\right) < \epsilon, \tag{10}$$

then the distribution P gives rise to an achievable second-order coding rate of α, i.e.,

$$\liminf_{n \to \infty} \frac{\log M^*(n, \epsilon, \Gamma, V) - nC(\Gamma)}{\sqrt{n}} \ge \alpha. \tag{11}$$

Proof: The proof can be adapted from the proof of [10, Lemma 14] by (i) replacing controllers with distributions P satisfying (8) and (ii) replacing sums with integrals.

Lemma 4 is a starting point for proving our achievability result. Hence, a central quantity of interest in the proof is

$$\log \frac{W(\mathbf{Y}|\mathbf{X})}{PW(\mathbf{Y})}. \tag{12}$$

As discussed in the introduction, Lemmas 5, 6 and 7 are helpful in the analysis of (12), where P is a mixture distribution. Specifically, the lemmas are helpful in approximating the output distribution PW in terms of a simpler distribution.

Lemma 5: Consider a random vector $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$, where $\mathbf{X}$ and $\mathbf{Z}$ are independent, $\mathbf{X}$ is uniformly distributed on an $(n-1)$ sphere of radius R and $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, NI_n)$. Let Q^{cc} denote the PDF of $\mathbf{Y}$. Then

$$Q^{cc}(\mathbf{y}) = \frac{\Gamma\left(\frac{n}{2}\right)}{2(\pi N)^{n/2}} \cdot \exp\left(-\frac{R^2 + ||\mathbf{y}||^2}{2N}\right) \left(\frac{N}{R||\mathbf{y}||}\right)^{\frac{n}{2}-1} I_{\frac{n}{2}-1}\left(\frac{R||\mathbf{y}||}{N}\right).$$

Proof: The proof is given in Appendix B.

To state the next two lemmas in a succinct way, we need to introduce some notation.

Definition 5 (Multi-parameter and multi-variable big-O notation): Let $f, f_1, \ldots, f_m$ be functions of ϵ, Δ and n. We write

$$f(\epsilon, \Delta, n) = O\left(f_1(\epsilon, \Delta, n), \ldots, f_m(\epsilon, \Delta, n)\right)$$

if there exist positive constants $\epsilon_0, \Delta_0, n_0$ and $C_1, \ldots, C_m$ such that for all $n \geq n_0$, $|\epsilon| \leq \epsilon_0$ and $|\Delta| \leq \Delta_0$,

$$|f(\epsilon, \Delta, n)| \leq \sum_{i=1}^{m} C_i |f_i(\epsilon, \Delta, n)|.$$

Lemma 6: Consider a random vector $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$, where $\mathbf{X}$ and $\mathbf{Z}$ are independent, $\mathbf{X}$ is uniformly distributed on an $(n-1)$ sphere of radius $\sqrt{n\Gamma}$ and $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, NI_n)$. Let Q^{cc} denote the PDF of $\mathbf{Y}$. Consider another random vector $\mathbf{Y}' = \mathbf{X}' + \mathbf{Z}'$, where $\mathbf{X}'$ and $\mathbf{Z}'$ are independent, $\mathbf{X}'$ is uniformly distributed on an $(n-1)$ sphere of radius $\sqrt{n\Gamma'}$ and $\mathbf{Z}' \sim \mathcal{N}(\mathbf{0}, NI_n)$. Let Q_0^{cc} denote the PDF of $\mathbf{Y}'$. Let $\Gamma' = \Gamma + \epsilon$. Then

$$\sup_{\mathbf{y} \in \mathcal{P}_n^*} \left| \log \frac{Q_0^{cc}(\mathbf{y})}{Q^{cc}(\mathbf{y})} \right| = O\left(\log n, n\epsilon^2, n\Delta^2, n\epsilon\Delta\right),$$

where $\mathcal{P}_n^* = \left\{\mathbf{y} \in \mathbb{R}^n : \Gamma + N - \Delta \leq \frac{||\mathbf{y}||^2}{n} \leq \Gamma + N + \Delta\right\}$ and $\Delta \geq 0$.

Remark 1: The parameters ϵ and Δ in Lemma 6 may depend on n.

Proof: The proof of Lemma 6 is given in Appendix C.

Lemma 7: Consider a random vector $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$, where $\mathbf{X}$ and $\mathbf{Z}$ are independent, $\mathbf{X}$ is uniformly distributed on an $(n-1)$ sphere of radius $\sqrt{n\Gamma}$ and $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, NI_n)$. Let Q^{cc} denote the PDF of $\mathbf{Y}$ and let $Q^* = \mathcal{N}(\mathbf{0}, (\Gamma + N)I_n)$. Then

$$\sup_{\mathbf{y} \in \mathcal{P}_n^*} \left| \log \frac{Q^{cc}(\mathbf{y})}{Q^*(\mathbf{y})} \right| = O\left(\log n, n\Delta^2\right),$$

where $\mathcal{P}_n^* = \left\{\mathbf{y} \in \mathbb{R}^n : \Gamma + N - \Delta \leq \frac{||\mathbf{y}||^2}{n} \leq \Gamma + N + \Delta\right\}$ and $\Delta \geq 0$.

Remark 2: The parameter Δ in Lemma 7 may depend on n. In fact, in the proof of Theorem 1, we invoke Lemma 7 for $\Delta = \sqrt{\frac{\log n}{n}}$.

Proof: The proof of Lemma 7 is given in Appendix D.

Remark 3: The set $\mathcal{P}_n^*$ in both Lemmas 6 and 7 is a "high probability" set in the following sense. For $\Delta = \omega(n^{-1/2})$, we have $\mathbb{P}\left(\mathbf{Y} \in \mathcal{P}_n^*\right) \to 1$ as $n \to \infty$, if $\mathbb{E}\left[||\mathbf{Y}||^2/n\right] = \Gamma + N$ and $\mathrm{Var}(||\mathbf{Y}||^2) = O(n)$. The latter two conditions hold when $\mathbf{Y} \sim Q^*$ or $\mathbf{Y} \sim Q^{cc}$. This property is used multiple times in the proof of Theorem 1.

Remark 4: See also [8, Theorem 42] for a related result to Lemma 7.

Theorem 1: Fix an arbitrary $\epsilon \in (0, 1)$. Consider an AWGN channel with noise variance $N > 0$. Under the mean and variance cost constraint specified by the pair $(\Gamma, V) \in (0, \infty) \times [0, \infty)$, we have

$$\liminf_{n \to \infty} \frac{\log M^*(n, \epsilon, \Gamma, V) - nC(\Gamma)}{\sqrt{n}} \geq \max\left\{r \in \mathbb{R} : \mathcal{K}\left(\frac{r}{\sqrt{V(\Gamma)}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)}\right) \leq \epsilon\right\}.$$

Alternatively, for any second-order coding rate $r \in \mathbb{R}$,

$$\limsup_{n \to \infty} \bar{P}_e\left(n, C(\Gamma) + \frac{r}{\sqrt{n}}, \Gamma, V\right) \leq \mathcal{K}\left(\frac{r}{\sqrt{V(\Gamma)}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)}\right).$$

Remark 5: When $V = 0$, Theorem 1 recovers the optimal second-order coding rate corresponding to the maximal cost constraint $||\mathbf{X}|| \leq \sqrt{n\Gamma}$, as given in [11, Theorem 5]. In this special case, the achievability scheme in Theorem 1 simplifies to

taking the channel input to be uniformly distributed on a singular $(n-1)$-sphere of radius $\sqrt{n\Gamma}$. Thus, for $V = 0$, the proof of Theorem 1 is an alternative and more direct proof technique than that of [11, Theorem 5], which relied on a quantization argument and additional assumptions on convergence rates.

Remark 6: When $V > 0$, the achievability scheme involves the random channel input having a mixture distribution of two or three uniform distributions on $(n-1)$-spheres.

Let $r(\epsilon, \Gamma, V)$ denote the achievable SOCR for the mean and variance cost constraint in Theorem 1, which is also the optimal SOCR as shown later in Theorem 2. As remarked earlier, $r(\epsilon, \Gamma, 0) = \sqrt{V(\Gamma)}\Phi^{-1}(\epsilon)$ is the optimal SOCR for the maximal cost constraint. Fig. 1 plots the SOCR against the average error probability for a Gaussian channel with $\Gamma = 2$ and $N = 1$, showing improved SOCR for several values of V. In fact, [2, Theorem 1] shows that $r(\epsilon, \Gamma, V) > \sqrt{V(\Gamma)}\Phi^{-1}(\epsilon)$ for all $V > 0$.
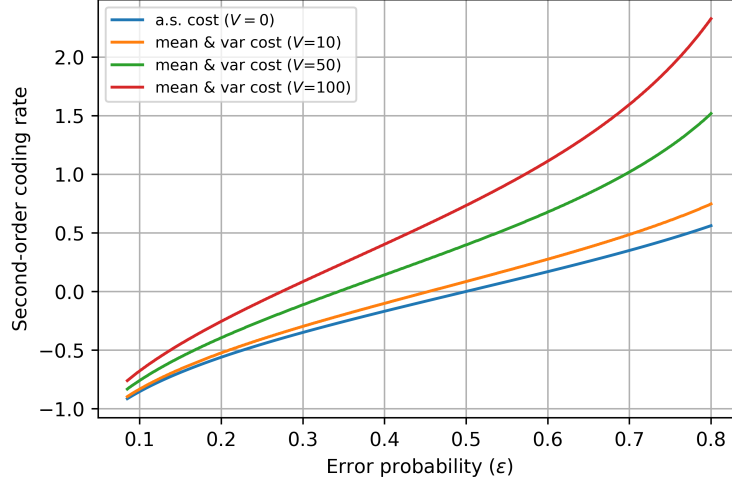

Fig. 1. The SOCR is compared between the almost-sure cost constraint and the mean and variance cost constraints for different values of V. The plots for the mean and variance cost constraints are technically lower bounds to the SOCR since they are obtained through a non-exhaustive search of the feasible region in the maximization and minimization in the formulation of $r(\epsilon, \Gamma, V)$.

Proof of Theorem 1: In view of Lemma 3, consider any distribution P_Π which achieves the minimum in

$$\mathcal{K}\left(\frac{r}{\sqrt{V(\Gamma)}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)}\right).$$

Let $\Pi \sim P_\Pi$, where we write

$$P_\Pi(\pi) = \begin{cases} p_1 & \pi = \pi_1 \\ p_2 & \pi = \pi_2 \\ p_3 & \pi = \pi_3. \end{cases}$$

Recall that $\mathbb{E}[\Pi] = \frac{r}{\sqrt{V(\Gamma)}}$ and $\mathrm{Var}(\Pi) \leq \frac{C'(\Gamma)^2 V}{V(\Gamma)}$. For each $j \in \{1, 2, 3\}$, let

$$\Gamma_j = \Gamma - \frac{\sqrt{V(\Gamma)}}{C'(\Gamma)\sqrt{n}}\pi_j + \frac{r}{C'(\Gamma)\sqrt{n}}. \tag{13}$$

We assume sufficiently large n so that $\Gamma_j > 0$ for all $j \in \{1, 2, 3\}$. Let P_j^* be the capacity-cost-achieving input distribution for $C(\Gamma_j)$ and Q_j^* be the corresponding optimal output distribution. Thus, $P_j^* = \mathcal{N}(0, \Gamma_j)$ and $Q_j^* = \mathcal{N}(0, \Gamma_j + N)$. Let Q_j^{cc} be the output distribution induced by the input distribution $\mathrm{Unif}(S_{R_j}^{n-1})$ for $R_j = \sqrt{n\Gamma_j}$.

Achievability Scheme: Let the random channel input $\mathbf{X}$ be such that with probability p_j, $\mathbf{X} \sim \mathrm{Unif}(S_{R_j}^{n-1})$. Denoting the distribution of $\mathbf{X}$ by P, we can write

$$P = \sum_{j=1}^{3} p_j \cdot \mathrm{Unif}(S_{R_j}^{n-1}).$$

The output distribution of $\mathbf{Y}$ induced by $P \circ W$ is

$$PW(\mathbf{y}) = \sum_{j=1}^{3} p_j Q_j^{cc}(\mathbf{y}).$$

We define

$$\mathcal{E}_n := (P \circ W)\left(\frac{1}{n}\log\frac{W(\mathbf{Y}|\mathbf{X})}{PW(\mathbf{Y})} \le C(\Gamma) + \frac{r}{\sqrt{n}}\right)$$

and show that $\limsup_{n\to\infty}\mathcal{E}_n < \epsilon$ which, by Lemma 4, would show that the random coding scheme achieves a second-order coding rate of r.

Analysis: We first write

$$\mathcal{E}_n = \sum_{j=1}^{3} p_j \mathbb{P}_{\mathbf{X}\sim\text{Unif}(S_{R_j}^{n-1})}\left(\frac{1}{n}\log\frac{W(\mathbf{Y}|\mathbf{X})}{PW(\mathbf{Y})} \le C(\Gamma) + \frac{r}{\sqrt{n}}\right). \tag{14}$$

To proceed further, we upper bound

$$\mathbb{P}_{\mathbf{X}\sim\text{Unif}(S_{R_j}^{n-1})}\left(\frac{1}{n}\log\frac{W(\mathbf{Y}|\mathbf{X})}{PW(\mathbf{Y})} \le C(\Gamma) + \frac{r}{\sqrt{n}}\right) \tag{15}$$

$$= \int_{\mathbf{y}\in\mathbb{R}^n} d\mathbf{y}\, Q_j^{cc}(\mathbf{y})\mathbb{P}\left(\frac{1}{n}\log\frac{W(\mathbf{y}|\mathbf{X})}{PW(\mathbf{y})} \le C(\Gamma) + \frac{r}{\sqrt{n}}\Big|\mathbf{Y}=\mathbf{y}\right)$$

$$\le \int_{\mathbf{y}\in\mathbb{R}^n} d\mathbf{y}\, Q_j^{cc}(\mathbf{y})\mathbb{P}\left(\frac{1}{n}\log\frac{W(\mathbf{y}|\mathbf{X})}{Q_i^{cc}(\mathbf{y})} \le C(\Gamma) + \frac{r}{\sqrt{n}}\Big|\mathbf{Y}=\mathbf{y}\right),$$

where $i \in \{1,2,3\}$ depends on $\mathbf{y}$ and is such that $Q_i^{cc}(\mathbf{y})$ assigns the highest probability to $\mathbf{y}$. We continue the derivation as

$$\mathbb{P}_{\mathbf{X}\sim\text{Unif}(S_{R_j}^{n-1})}\left(\frac{1}{n}\log\frac{W(\mathbf{Y}|\mathbf{X})}{PW(\mathbf{Y})} \le C(\Gamma) + \frac{r}{\sqrt{n}}\right)$$

$$\le \int_{\mathbf{y}\in\mathbb{R}^n} d\mathbf{y}\, Q_j^{cc}(\mathbf{y})\mathbb{P}\left(\log\frac{W(\mathbf{y}|\mathbf{X})}{Q_i^{cc}(\mathbf{y})} \le nC(\Gamma) + r\sqrt{n}\Big|\mathbf{Y}=\mathbf{y}\right)$$

$$= \int_{\mathbf{y}\in\mathbb{R}^n} d\mathbf{y}\, Q_j^{cc}(\mathbf{y})\mathbb{P}\left(\log\frac{W(\mathbf{y}|\mathbf{X})}{Q_j^{cc}(\mathbf{y})} \le nC(\Gamma) + r\sqrt{n} + \log\frac{Q_i^{cc}(\mathbf{y})}{Q_j^{cc}(\mathbf{y})}\Big|\mathbf{Y}=\mathbf{y}\right)$$

$$\le \int_{\mathbf{y}\in\mathbb{R}^n} d\mathbf{y}\, Q_j^{cc}(\mathbf{y})\mathbb{P}\left(\log\frac{W(\mathbf{y}|\mathbf{X})}{Q_j^{cc}(\mathbf{y})} \le nC(\Gamma) + r\sqrt{n} + \kappa_1\log n\Big|\mathbf{Y}=\mathbf{y}\right) + \delta_n^{(j)}. \tag{16}$$

In the last inequality above, we used Lemma 6. Specifically, in Lemma 6, let

- $\Gamma = \Gamma_j$,
- $\Gamma' = \Gamma_i$,
- $\epsilon = \Gamma_i - \Gamma_j$ so that $\epsilon = O\left(\frac{1}{\sqrt{n}}\right)$, and
- $\Delta = \sqrt{\frac{\log n}{n}}$.

Consequently, κ_1 is a constant from the result of Lemma 6 and

$$\delta_n^{(j)} = Q_j^{cc}\left(\left|\frac{\|\mathbf{Y}\|^2}{n} - \Gamma_j - N\right| > \Delta\right).$$

It is easy to check that for $\mathbf{Y}\sim Q_j^{cc}$, $\mathbb{E}\left[\|\mathbf{Y}\|^2\right] = n\Gamma_j + nN$ and $\text{Var}(\|\mathbf{Y}\|^2) = 4nN\Gamma_j + 2nN^2$. Thus, it is easy to see that $\delta_n^{(j)} \to 0$ as $n \to \infty$ using Chebyshev inequality.

Continuing the derivation from (16), we have

$$\mathbb{P}_{\mathbf{X}\sim\text{Unif}(S_{R_j}^{n-1})}\left(\frac{1}{n}\log\frac{W(\mathbf{Y}|\mathbf{X})}{PW(\mathbf{Y})} \le C(\Gamma) + \frac{r}{\sqrt{n}}\right)$$

$$\le \int_{\mathbf{y}\in\mathbb{R}^n} d\mathbf{y}\, Q_j^{cc}(\mathbf{y})\mathbb{P}\left(\log\frac{W(\mathbf{y}|\mathbf{X})}{Q_j^*(\mathbf{y})} \le nC(\Gamma) + r\sqrt{n} + \kappa_1\log n + \log\frac{Q_j^{cc}(\mathbf{y})}{Q_j^*(\mathbf{y})}\Big|\mathbf{Y}=\mathbf{y}\right) + \delta_n^{(j)}$$

$$\le \int_{\mathbf{y}\in\mathbb{R}^n} d\mathbf{y}\, Q_j^{cc}(\mathbf{y})\mathbb{P}\left(\log\frac{W(\mathbf{y}|\mathbf{X})}{Q_j^*(\mathbf{y})} \le nC(\Gamma) + r\sqrt{n} + \kappa\log n\Big|\mathbf{Y}=\mathbf{y}\right) + 2\delta_n^{(j)}. \tag{17}$$

In the last inequality above, we used Lemma 7. Specifically, in Lemma 7, let $\Gamma = \Gamma_j$ and $\Delta = \sqrt{\frac{\log n}{n}}$. Consequently, κ is a suitable bounding constant for both Lemma 6 and Lemma 7.

Continuing the derivation from (17), we have

$$\mathbb{P}_{\mathbf{X}\sim\mathrm{Unif}(S_{R_j}^{n-1})}\left(\frac{1}{n}\log\frac{W(\mathbf{Y}|\mathbf{X})}{PW(\mathbf{Y})}\leq C(\Gamma)+\frac{r}{\sqrt{n}}\right)$$

$$\leq \mathbb{P}_{\mathbf{X}\sim\mathrm{Unif}(S_{R_j}^{n-1})}\left(\log\frac{W(\mathbf{Y}|\mathbf{X})}{Q_j^*(\mathbf{Y})}\leq nC(\Gamma)+r\sqrt{n}+\kappa\log n\right)+2\delta_n^{(j)}$$

$$= \mathbb{P}_{\mathbf{X}\sim\mathrm{Unif}(S_{R_j}^{n-1})}\left(\sum_{m=1}^{n}\log\frac{W(Y_m|X_m)}{Q_j^*(Y_m)}-nC(\Gamma_j)\leq n\left(C(\Gamma)-C(\Gamma_j)\right)+r\sqrt{n}+\kappa\log n\right)+2\delta_n^{(j)}$$

$$\overset{(a)}{=} \mathbb{P}_{\mathbf{X}\sim\mathrm{Unif}(S_{R_j}^{n-1})}\left(\sum_{m=1}^{n}\left[\log\frac{W(Y_m|X_m)}{Q_j^*(Y_m)}-\mathbb{E}\left[\log\frac{W(Y_m|X_m)}{Q_j^*(Y_m)}\right]\right]\leq n\left(C(\Gamma)-C(\Gamma_j)\right)+r\sqrt{n}+\kappa\log n\right)+2\delta_n^{(j)}$$

$$\overset{(b)}{=} \mathbb{P}\left(\sum_{m=1}^{n}\tilde{T}_m\leq n\left(C(\Gamma)-C(\Gamma_j)\right)+r\sqrt{n}+\kappa\log n\right)+2\delta_n^{(j)}$$

$$\overset{(c)}{=} \mathbb{P}\left(\frac{1}{\sqrt{nV(\Gamma_j)}}\sum_{m=1}^{n}\tilde{T}_m\leq\sqrt{n}\frac{C(\Gamma)-C(\Gamma_j)}{\sqrt{V(\Gamma_j)}}+\frac{r}{\sqrt{V(\Gamma_j)}}+\frac{\kappa\log n}{\sqrt{nV(\Gamma_j)}}\right)+2\delta_n^{(j)}$$

$$\overset{(d)}{=} \Phi\left(\sqrt{n}\frac{C(\Gamma)-C(\Gamma_j)}{\sqrt{V(\Gamma_j)}}+\frac{r}{\sqrt{V(\Gamma_j)}}+\frac{\kappa\log n}{\sqrt{nV(\Gamma_j)}}\right)+\delta_n$$

$$\overset{(e)}{\leq} \Phi\left(\frac{\sqrt{n}}{2\sqrt{V(\Gamma_j)}}\left(\frac{\Gamma-\Gamma_j}{N+\Gamma}\right)+\frac{r}{\sqrt{V(\Gamma_j)}}+\frac{2\kappa\log n}{\sqrt{nV(\Gamma_j)}}\right)+\delta_n \qquad (18)$$

In equality (a) above, we used Lemma 1. Specifically, use Equation (3) in Lemma 1 with $\Gamma=\Gamma_j$.

In equality (b) above, we used Lemma 2, where the $\tilde{T}_m$'s are i.i.d. and

$$\tilde{T}_m=\log\frac{W(Y|\sqrt{\Gamma_j})}{Q_j^*(Y)}-\mathbb{E}\left[\log\frac{W(Y|\sqrt{\Gamma_j})}{Q_j^*(Y)}\right], \qquad (19)$$

where $Y\sim\mathcal{N}(\sqrt{\Gamma_j},N)$.

In equality (c) above, we normalize the sum to have unit variance, which follows from Corollary 1 and Equation (4).

In equality (d) above, we used the Berry-Esseen Theorem [12] to obtain convergence of the CDF of the normalized sum of i.i.d. random variables $\tilde{T}_m$'s to the standard normal CDF, with $\delta_n\to 0$ accounting for both the rate of convergence and $\delta_n^{(j)}\to 0$. In inequality (e) above, we used a Taylor series approximation, noting that $\Gamma-\Gamma_j=O(1/\sqrt{n})$.

Substituting (18) into (14), we obtain

$$\mathcal{E}_n\leq\sum_{j=1}^{3}p_j\Phi\left(\frac{\sqrt{n}}{2\sqrt{V(\Gamma_j)}}\left(\frac{\Gamma-\Gamma_j}{N+\Gamma}\right)+\frac{r}{\sqrt{V(\Gamma_j)}}+\frac{2\kappa\log n}{\sqrt{nV(\Gamma_j)}}\right)+\delta_n$$

for some redefined sequence $\delta_n\to 0$ as $n\to\infty$. Using Equation (13) and the formula for $C'(\Gamma)$ from Lemma 1, we can simplify the upper bound as

$$\mathcal{E}_n\leq\sum_{j=1}^{3}p_j\Phi\left(\frac{\sqrt{V(\Gamma)}}{\sqrt{V(\Gamma_j)}}\pi_j-\frac{r}{\sqrt{V(\Gamma_j)}}+\frac{r}{\sqrt{V(\Gamma_j)}}+\frac{2\kappa\log n}{\sqrt{nV(\Gamma_j)}}\right)+\delta_n.$$

Therefore, since $\Gamma_j\to\Gamma$ as $n\to\infty$, we have

$$\limsup_{n\to\infty}\mathcal{E}_n\leq\sum_{j=1}^{3}p_j\Phi\left(\pi_j\right)$$

$$=\mathcal{K}\left(\frac{r}{\sqrt{V(\Gamma)}},\frac{C'(\Gamma)^2V}{V(\Gamma)}\right). \qquad (20)$$

To complete the proof, we choose $r<r^*$ where

$$r^*=\max\left\{r:\mathcal{K}\left(\frac{r}{\sqrt{V(\Gamma)}},\frac{C'(\Gamma)^2V}{V(\Gamma)}\right)\leq\epsilon\right\}. \qquad (21)$$

Hence, $\limsup_{n\to\infty} \mathcal{E}_n < \epsilon$ because, from Lemma 3, $\mathcal{K}(\cdot,\cdot)$ is a strictly increasing function in the first argument. Hence, $\limsup_{n\to\infty} \mathcal{E}_n < \epsilon$ for $r < r^*$, thus establishing that any $r < r^*$ is an achievable second-order coding rate. Finally, letting $r \to r^*$ establishes an achievable second-order coding rate of r^*, matching the converse in Theorem 2.

The achievability result can also be stated in terms of an upper bound on the minimum average probability of error of (n, R, Γ, V) codes for a rate $R = C(\Gamma) + \frac{r}{\sqrt{n}}$. From Lemma 4, we have for $\theta = 1/n^\vartheta$ for $1 > \vartheta > 1/2$,

$$\bar{P}_e\left(n, C(\Gamma) + \frac{r}{\sqrt{n}}, \Gamma, V\right) \leq (P \circ W)\left(\frac{1}{n}\log\frac{W(\mathbf{Y}|\mathbf{X})}{PW(\mathbf{Y})} \leq C(\Gamma) + \frac{r}{\sqrt{n}} + \frac{1}{n^\vartheta}\right) + e^{-n^{1-\vartheta}}. \tag{22}$$

For any $r' > r$, we have $\frac{r}{\sqrt{n}} + \frac{1}{n^\vartheta} < \frac{r'}{\sqrt{n}}$ eventually, so

$$\limsup_{n\to\infty} \bar{P}_e\left(n, C(\Gamma) + \frac{r}{\sqrt{n}}, \Gamma, V\right)$$
$$\leq \limsup_{n\to\infty}(P \circ W)\left(\frac{1}{n}\log\frac{W(\mathbf{Y}|\mathbf{X})}{PW(\mathbf{Y})} \leq C(\Gamma) + \frac{r'}{\sqrt{n}}\right)$$
$$\leq \mathcal{K}\left(\frac{r'}{\sqrt{V(\Gamma)}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)}\right), \tag{23}$$

where the last inequality follows from (20). Letting $r' \to r$ in (23) and invoking continuity of the function $\mathcal{K}$ establishes the result.

∎

IV. CONVERSE RESULT

Lemma 8: Consider an AWGN channel W with noise variance $N > 0$ and cost constraint $(\Gamma, V) \in (0, \infty) \times (0, \infty)$. Then for every $n, \rho > 0$ and $\epsilon \in (0, 1)$,

$$\log M^*(n, \epsilon, \Gamma, V) \leq \log \rho - \log\left[\left(1 - \epsilon - \sup_{\overline{P} \in \mathcal{P}_{\Gamma,V}(\mathbb{R}^n)} \inf_{q \in \mathcal{P}(\mathbb{R}^n)} (\overline{P} \circ W)\left(\frac{W(\mathbf{Y}|\mathbf{X})}{q(\mathbf{Y})} > \rho\right)\right)^+\right], \tag{24}$$

where $\mathcal{P}_{\Gamma,V}(\mathbb{R}^n) \subset \mathcal{P}(\mathbb{R}^n)$ is the set of distributions $\overline{P}$ such that $\mathbb{E}\left[\sum_{i=1}^n c(X_i)\right] \leq n\Gamma$ and $\mathrm{Var}\left(\sum_{i=1}^n c(X_i)\right) \leq nV$ for $\mathbf{X} \sim \overline{P}$.

Proof: The proof is similar to that of [9, Lemma 2] and is omitted.

Theorem 2: Fix an arbitrary $\epsilon \in (0, 1)$. Consider an AWGN channel with noise variance $N > 0$. Under the mean and variance cost constraints specified by the pair $(\Gamma, V) \in (0, \infty) \times (0, \infty)$, we have

$$\limsup_{n\to\infty} \frac{\log M^*(n, \epsilon, \Gamma, V) - nC(\Gamma)}{\sqrt{n}} \leq \max\left\{r \in \mathbb{R} : \mathcal{K}\left(\frac{r}{\sqrt{V(\Gamma)}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)}\right) \leq \epsilon\right\}.$$

Alternatively, for any second-order coding rate $r \in \mathbb{R}$,

$$\liminf_{n\to\infty} \bar{P}_e\left(n, C(\Gamma) + \frac{r}{\sqrt{n}}, \Gamma, V\right) \geq \mathcal{K}\left(\frac{r}{\sqrt{V(\Gamma)}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)}\right).$$

Proof of Theorem 2:

We start with Lemma 8 and first upper bound

$$\sup_{\overline{P} \in \mathcal{P}_{\Gamma,V}(\mathbb{R}^n)} \inf_{q \in \mathcal{P}(\mathbb{R}^n)} (\overline{P} \circ W)\left(\frac{W(\mathbf{Y}|\mathbf{X})}{q(\mathbf{Y})} > \rho\right). \tag{25}$$

Let $\rho = \exp\left(nC(\Gamma) + \sqrt{n}r\right)$ in (25), where r is a number to be specified later. Let $\mathbf{X} \sim \overline{P}$ for any arbitrary $\overline{P} \in \mathcal{P}_{\Gamma,V}(\mathbb{R}^n)$ so that $\mathbb{E}\left[\sum_{i=1}^n c(X_i)\right] \leq n\Gamma$ and $\mathrm{Var}\left(\sum_{i=1}^n c(X_i)\right) \leq nV$.

Choosing $q = Q^*$ in (25), we have

$$(\overline{P} \circ W)\left(\log\frac{W(\mathbf{Y}|\mathbf{X})}{Q^*(\mathbf{Y})} > nC(\Gamma) + \sqrt{n}r\right)$$
$$= \int_{\mathbb{R}^n} d\overline{P}(\mathbf{x})W\left(\sum_{i=1}^n \log\frac{W(Y_i|x_i)}{Q^*(Y_i)} > nC(\Gamma) + \sqrt{n}r\right), \tag{26}$$

where in (26), $Y_1, Y_2, \ldots, Y_n$ are independent random variables and each $Y_i \sim \mathcal{N}(x_i, N)$. We define the empirical distribution of a vector $\mathbf{x}$ as

$$P_{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i},$$

where δ_x is the dirac delta measure at x. For any given channel input $\mathbf{x}$ and independent Y_i's where $Y_i \sim \mathcal{N}(x_i, N)$, we have from Corollary 1 that

$$\mathrm{Var}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\log \frac{W(Y_i|x_i)}{Q^*(Y_i)} - \mathbb{E}\left[\log \frac{W(Y_i|x_i)}{Q^*(Y_i)}\right]\right]\right)$$
$$= \int \nu_t P_{\mathbf{x}}(t) dt$$
$$= \frac{\Gamma^2 + 2Nc(\mathbf{x})}{2(N+\Gamma)^2}$$

For any $\beta \geq 0$, define

$$\delta_\beta := \sup_{\Gamma - \beta < t \leq \Gamma + \beta} \left|\frac{\Gamma^2 + 2Nt}{2(N+\Gamma)^2} - V(\Gamma)\right|$$

and note that $\delta_\beta \to 0$ as $\beta \to 0$. Define the function

$$\psi_\beta(t) := \begin{cases} V(\Gamma) + \delta_\beta & \text{if } t \leq \Gamma \\ V(\Gamma) - \delta_\beta & \text{if } t > \Gamma. \end{cases}$$

Fix any $0 < \Delta < 1/2$ and $\eta > 0$. Then select $\beta > 0$ such that $\delta_\beta < V(\Gamma)$,

$$\sup_{t \in \mathbb{R}} \left|\frac{1}{\sqrt{V(\Gamma)}} - \frac{1}{\sqrt{\psi_\beta(t)}}\right| \leq \frac{\Delta}{C'(\Gamma)\sqrt{V}} \tag{27}$$

and

$$\sup_{t \in \mathbb{R}} \left|\frac{1}{V(\Gamma)} - \frac{1}{\psi_\beta(t)}\right| \leq \frac{\Delta/2}{C'(\Gamma)^2 V}. \tag{28}$$

Also define

$$\varphi(r) := \begin{cases} V(\Gamma) - \delta_\beta & \text{if } r \leq 0 \\ V(\Gamma) + \delta_\beta & \text{if } r > 0. \end{cases} \tag{29}$$

With Δ, η and β fixed as above[1], we divide the set of sequences $\mathbf{x} \in \mathbb{R}^n$ into three subsets:

$$\mathcal{P}_{n,1} := \{\mathbf{x} : c(\mathbf{x}) \leq \Gamma - \beta\},$$
$$\mathcal{P}_{n,2} := \{\mathbf{x} : \Gamma - \beta < c(\mathbf{x}) \leq \Gamma + \beta\},$$
$$\mathcal{P}_{n,3} := \{\mathbf{x} : c(\mathbf{x}) > \Gamma + \beta\}.$$

For $\mathbf{x} \in \mathcal{P}_{n,1}$, we have

$$W\left(\sum_{i=1}^{n} \log \frac{W(Y_i|x_i)}{Q^*(Y_i)} > nC(\Gamma) + \sqrt{n}r\right)$$
$$\overset{(a)}{=} W\left(\sum_{i=1}^{n} \left[\log \frac{W(Y_i|x_i)}{Q^*(Y_i)} - \mathbb{E}\left[\log \frac{W(Y_i|x_i)}{Q^*(Y_i)}\right]\right] > nC'(\Gamma)(\Gamma - c(\mathbf{x})) + \sqrt{n}r\right)$$
$$\leq W\left(\sum_{i=1}^{n} \left[\log \frac{W(Y_i|x_i)}{Q^*(Y_i)} - \mathbb{E}\left[\log \frac{W(Y_i|x_i)}{Q^*(Y_i)}\right]\right] > nC'(\Gamma)\beta + \sqrt{n}r\right)$$
$$\overset{(b)}{\leq} W\left(\sum_{i=1}^{n} \left[\log \frac{W(Y_i|x_i)}{Q^*(Y_i)} - \mathbb{E}\left[\log \frac{W(Y_i|x_i)}{Q^*(Y_i)}\right]\right] > \frac{nC'(\Gamma)\beta}{2}\right)$$
$$\overset{(c)}{\leq} \frac{4}{nC'(\Gamma)^2\beta^2}\left(\frac{\Gamma^2 + 2N\Gamma}{2(N+\Gamma)^2}\right) \tag{30}$$

[1]η is used later.

for sufficiently large n. In equality (a), we used Lemma 1. Inequality (b) holds because r is a constant. Inequality (c) follows by applying Chebyshev's inequality and Corollary 1, the latter of which gives us that

$$
\begin{aligned}
\mathrm{Var} & \left(\sum_{i=1}^n \left[\log \frac{W(Y_i|x_i)}{Q^*(Y_i)} - \mathbb{E}\left[\log \frac{W(Y_i|x_i)}{Q^*(Y_i)} \right] \right] \right) \\
& = \frac{n\Gamma^2 + 2nNc(\mathbf{x})}{2\left(N+\Gamma\right)^2} \\
& \le n\left(\frac{\Gamma^2 + 2N\Gamma}{2(N+\Gamma)^2} \right).
\end{aligned}
$$

For $\mathbf{x} \in \mathcal{P}_{n,2}$, we have

$$
\begin{aligned}
& W\left(\sum_{i=1}^n \log \frac{W(Y_i|x_i)}{Q^*(Y_i)} > nC(\Gamma) + \sqrt{n}r \right) \\
& \overset{(a)}{=} W\left(\sum_{i=1}^n \left[\log \frac{W(Y_i|x_i)}{Q^*(Y_i)} - \mathbb{E}\left[\log \frac{W(Y_i|x_i)}{Q^*(Y_i)} \right] \right] > nC'(\Gamma)\left(\Gamma - c(\mathbf{x}) \right) + \sqrt{n}r \right) \\
& \overset{(b)}{=} W\left(\frac{1}{\sqrt{\int n\,\nu_t P_{\mathbf{x}}(t)dt}} \sum_{i=1}^n \left[\log \frac{W(Y_i|x_i)}{Q^*(Y_i)} - \mathbb{E}\left[\log \frac{W(Y_i|x_i)}{Q^*(Y_i)} \right] \right] > \right. \\
& \qquad\qquad \left. \frac{\sqrt{n}C'(\Gamma)\left(\Gamma - c(\mathbf{x}) \right)}{\sqrt{\int \nu_t P_{\mathbf{x}}(t)dt}} + \frac{r}{\sqrt{\int \nu_t P_{\mathbf{x}}(t)dt}} \right) \\
& \overset{(c)}{=} W\left(\frac{1}{\sqrt{\int n\,\nu_t P_{\mathbf{x}}(t)dt}} \sum_{i=1}^n \tilde{T}_i > \frac{\sqrt{n}C'(\Gamma)\left(\Gamma - c(\mathbf{x}) \right)}{\sqrt{\int \nu_t P_{\mathbf{x}}(t)dt}} + \frac{r}{\sqrt{\int \nu_t P_{\mathbf{x}}(t)dt}} \right) \\
& \overset{(d)}{\le} 1 - \Phi\left(\frac{\sqrt{n}C'(\Gamma)\left(\Gamma - c(\mathbf{x}) \right)}{\sqrt{\int \nu_t P_{\mathbf{x}}(t)dt}} + \frac{r}{\sqrt{\int \nu_t P_{\mathbf{x}}(t)dt}} \right) + \frac{\kappa_{\mathbf{x}}}{\sqrt{n}} \\
& \overset{(e)}{\le} 1 - \Phi\left(\frac{\sqrt{n}C'(\Gamma)\left(\Gamma - c(\mathbf{x}) \right)}{\sqrt{\int \nu_t P_{\mathbf{x}}(t)dt}} + \frac{r}{\sqrt{\int \nu_t P_{\mathbf{x}}(t)dt}} \right) + \frac{\kappa}{\sqrt{n}} \\
& \le 1 - \Phi\left(\frac{\sqrt{n}C'(\Gamma)\left(\Gamma - c(\mathbf{x}) \right)}{\sqrt{\psi_\beta(c(\mathbf{x}))}} + \frac{r}{\sqrt{\varphi(r)}} \right) + \frac{\kappa}{\sqrt{n}} \qquad (31)
\end{aligned}
$$

for sufficiently large n. In equality (a), we used Lemma 1. In equality (b), we normalize the sum to have unit variance, where we write

$$
\begin{aligned}
& \int_{-\infty}^{\infty} n\,\nu_t P_{\mathbf{x}}(t)dt \\
& = \mathrm{Var}\left(\sum_{i=1}^n \left[\log \frac{W(Y_i|x_i)}{Q^*(Y_i)} - \mathbb{E}\left[\log \frac{W(Y_i|x_i)}{Q^*(Y_i)} \right] \right] \right) \\
& = \frac{n\Gamma^2 + 2nNc(\mathbf{x})}{2\left(N+\Gamma\right)^2}.
\end{aligned}
$$

In equality (c), we use Lemma 2. In inequality (d), we apply the Berry-Esseen Theorem [12], where $\kappa_{\mathbf{x}}$ is a constant depending on the second- and third-order moments of $\{\tilde{T}_i\}_{i=1}^n$. In inequality (e), we use the fact that the distribution of each $\tilde{T}_i$ depends on $\mathbf{x}$ only through its cost $c(\mathbf{x})$. Since $c(\mathbf{x})$ is uniformly bounded over the set $\mathcal{P}_{n,2}$, the constant $\kappa_{\mathbf{x}}$ can be uniformly upper bounded over all $\mathbf{x} \in \mathcal{P}_{n,2}$ by some constant κ that does not depend on $\mathbf{x}$ at all.

For $\mathcal{P}_{n,3}$, the following lemma shows that the probability of this set under $\overline{P}$ is small.

Lemma 9: We have

$$
\overline{P}\left(\mathcal{P}_{n,3} \right) \le \frac{V}{n\beta^2}. \qquad (32)
$$

Proof: Let $\mu_i = \mathbb{E}[c(X_i)]$. For any $\beta > 0$,

$$\overline{P}\left(c(\mathbf{X}) > \Gamma + \beta\right)$$

$$= \overline{P}\left(\sum_{i=1}^{n}(c(X_i) - \mu_i) + \sum_{i=1}^{n}\mu_i \geq n\Gamma + n\beta\right)$$

$$\leq \overline{P}\left(\sum_{i=1}^{n}(c(X_i) - \mu_i) \geq n\beta\right)$$

$$\leq \overline{P}\left(\left(\sum_{i=1}^{n}(c(X_i) - \mu_i)\right)^2 \geq n^2\beta^2\right)$$

$$\leq \frac{1}{n^2\beta^2}\mathbb{E}\left[\left(\sum_{i=1}^{n}(c(X_i) - \mu_i)\right)^2\right]$$

$$\leq \frac{V}{n\beta^2}.$$

$\blacksquare$

Using the results in (30), (31) and (32), we can upper bound (26) as

$$(\overline{P} \circ W)\left(\log\frac{W(\mathbf{Y}|\mathbf{X})}{Q^*(\mathbf{Y})} > nC(\Gamma) + \sqrt{n}r\right)$$

$$\leq 1 + \frac{4}{nC'(\Gamma)^2\beta^2}\left(\frac{\Gamma^2 + 2N\Gamma}{2(N + \Gamma)^2}\right) + \frac{V}{n\beta^2} + \frac{\kappa}{\sqrt{n}} -$$

$$\mathbb{E}\left[\Phi\left(\frac{\sqrt{n}C'(\Gamma)(\Gamma - c(\mathbf{X}))}{\sqrt{\psi_\beta(c(\mathbf{X}))}} + \frac{r}{\sqrt{\varphi(r)}}\right)\right]. \tag{33}$$

To further upper bound the above expression, we need to obtain a lower bound to

$$\inf_{\mathbf{X}}\mathbb{E}\left[\Phi\left(\frac{\sqrt{n}C'(\Gamma)(\Gamma - c(\mathbf{X}))}{\sqrt{\psi_\beta(c(\mathbf{X}))}} + \frac{r}{\sqrt{\varphi(r)}}\right)\right],$$

where the infimum is over all random vectors $\mathbf{X}$ such that $\mathbb{E}[c(\mathbf{X})] \leq \Gamma$ and $\text{Var}(c(\mathbf{X})) \leq V/n$. Without loss of generality, we can assume $\mathbb{E}[c(\mathbf{X})] = \Gamma$ since the function Φ is monotonically increasing and the function

$$c \mapsto \frac{\Gamma - c}{\sqrt{\psi_\beta(c)}}$$

is monotonically nonincreasing.

Note that

$$\left|\mathbb{E}\left[\frac{\sqrt{n}C'(\Gamma)}{\sqrt{\psi_\beta(c(\mathbf{X}))}}(\Gamma - c(\mathbf{X}))\right] - \mathbb{E}\left[\frac{\sqrt{n}C'(\Gamma)}{\sqrt{V(\Gamma)}}(\Gamma - c(\mathbf{X}))\right]\right|$$

$$= \left|\mathbb{E}\left[\sqrt{n}C'(\Gamma)(\Gamma - c(\mathbf{X}))\left[\frac{1}{\sqrt{\psi_\beta(c(\mathbf{X}))}} - \frac{1}{\sqrt{V(\Gamma)}}\right]\right]\right|$$

$$\leq \sqrt{n}C'(\Gamma)\mathbb{E}\left[|\Gamma - c(\mathbf{X})| \cdot \left|\frac{1}{\sqrt{\psi_\beta(c(\mathbf{X}))}} - \frac{1}{\sqrt{V(\Gamma)}}\right|\right]$$

$$\overset{(a)}{\leq} \sqrt{n}C'(\Gamma)\frac{\Delta}{C'(\Gamma)\sqrt{V}}\mathbb{E}\left[|\Gamma - c(\mathbf{X})|\right]$$

$$\overset{(b)}{\leq} \Delta. \tag{34}$$

In inequality (a), we used (27). In inequality (b), we used the fact that $\mathbb{E}\left[\sqrt{(\Gamma - c(\mathbf{X}))^2}\right] \leq \sqrt{\mathrm{Var}(c(\mathbf{X}))}$. Hence, from (34), we have

$$\mathbb{E}\left[\frac{\sqrt{n}C'(\Gamma)}{\sqrt{\psi_\beta(c(\mathbf{X}))}}\left(\Gamma - c(\mathbf{X})\right)\right]$$
$$\geq \mathbb{E}\left[\frac{\sqrt{n}C'(\Gamma)}{\sqrt{V(\Gamma)}}\left(\Gamma - c(\mathbf{X})\right)\right] - \Delta$$
$$= -\Delta. \tag{35}$$

Now define $X' = \Gamma - c(\mathbf{X})$ so that $\mathbb{E}[X'] = 0$ and $\mathrm{Var}(X') \leq V/n$. Also define

$$\Psi_\beta(X') := \begin{cases} V(\Gamma) + \delta_\beta & \text{if } X' \geq 0 \\ V(\Gamma) - \delta_\beta & \text{if } X' < 0. \end{cases}$$

Then note that

$$\left|\mathrm{Var}\left(\frac{\Gamma - c(\mathbf{X})}{\sqrt{\psi_\beta(c(\mathbf{X}))}}\right) - \mathrm{Var}\left(\frac{\Gamma - c(\mathbf{X})}{\sqrt{V(\Gamma)}}\right)\right|$$
$$= \left|\mathrm{Var}\left(\frac{X'}{\sqrt{\Psi_\beta(X')}}\right) - \mathrm{Var}\left(\frac{X'}{\sqrt{V(\Gamma)}}\right)\right|$$
$$= \left|\mathbb{E}\left[\frac{X'^2}{\Psi_\beta(X')} - \frac{X'^2}{V(\Gamma)}\right] + \left(\mathbb{E}\left[\frac{X'}{\sqrt{V(\Gamma)}}\right]\right)^2 - \left(\mathbb{E}\left[\frac{X'}{\sqrt{\Psi_\beta(X')}}\right]\right)^2\right|$$
$$= \left|\mathbb{E}\left[\left(\frac{X'}{\sqrt{\Psi_\beta(X')}} + \frac{X'}{\sqrt{V(\Gamma)}}\right)\left(\frac{X'}{\sqrt{\Psi_\beta(X')}} - \frac{X'}{\sqrt{V(\Gamma)}}\right)\right] + \right.$$
$$\left.\left(\mathbb{E}\left[\frac{X'}{\sqrt{V(\Gamma)}}\right] + \mathbb{E}\left[\frac{X'}{\sqrt{\Psi_\beta(X')}}\right]\right)\left(\mathbb{E}\left[\frac{X'}{\sqrt{V(\Gamma)}}\right] - \mathbb{E}\left[\frac{X'}{\sqrt{\Psi_\beta(X')}}\right]\right)\right|$$
$$\leq \left|\mathbb{E}\left[\left(\frac{X'}{\sqrt{\Psi_\beta(X')}} + \frac{X'}{\sqrt{V(\Gamma)}}\right)\left(\frac{X'}{\sqrt{\Psi_\beta(X')}} - \frac{X'}{\sqrt{V(\Gamma)}}\right)\right]\right| + $$
$$\left|\left(\mathbb{E}\left[\frac{X'}{\sqrt{V(\Gamma)}}\right] + \mathbb{E}\left[\frac{X'}{\sqrt{\Psi_\beta(X')}}\right]\right)\left(\mathbb{E}\left[\frac{X'}{\sqrt{V(\Gamma)}}\right] - \mathbb{E}\left[\frac{X'}{\sqrt{\Psi_\beta(X')}}\right]\right)\right|$$
$$\overset{(a)}{\leq} \mathbb{E}[X'^2]\frac{\Delta/2}{C'(\Gamma)^2 V} + \left|\left(\mathbb{E}\left[\frac{X'}{\sqrt{V(\Gamma)}}\right] + \mathbb{E}\left[\frac{X'}{\sqrt{\Psi_\beta(X')}}\right]\right)\left(\mathbb{E}\left[\frac{X'}{\sqrt{V(\Gamma)}}\right] - \mathbb{E}\left[\frac{X'}{\sqrt{\Psi_\beta(X')}}\right]\right)\right|$$
$$\leq \mathbb{E}[X'^2]\frac{\Delta/2}{C'(\Gamma)^2 V} + \left|\mathbb{E}\left[\frac{X'}{\sqrt{V(\Gamma)}} + \frac{X'}{\sqrt{\Psi_\beta(X')}}\right]\right| \cdot \left|\mathbb{E}\left[\frac{X'}{\sqrt{V(\Gamma)}} - \frac{X'}{\sqrt{\Psi_\beta(X')}}\right]\right|$$
$$\leq \mathbb{E}[X'^2]\frac{\Delta/2}{C'(\Gamma)^2 V} + \left|\mathbb{E}\left[\frac{X'}{\sqrt{\Psi_\beta(X')}}\right]\right| \cdot \mathbb{E}\left[|X'| \cdot \left|\frac{1}{\sqrt{V(\Gamma)}} - \frac{1}{\sqrt{\Psi_\beta(X')}}\right|\right]$$
$$\overset{(b)}{\leq} \mathbb{E}[X'^2]\frac{\Delta/2}{C'(\Gamma)^2 V} + \frac{\Delta}{\sqrt{n}C'(\Gamma)} \cdot \frac{\Delta}{C'(\Gamma)\sqrt{V}}\mathbb{E}[|X'|]$$
$$\overset{(c)}{\leq} \frac{V}{n}\frac{\Delta/2}{C'(\Gamma)^2 V} + \frac{\Delta}{\sqrt{n}C'(\Gamma)} \cdot \frac{\Delta}{C'(\Gamma)\sqrt{V}}\sqrt{\frac{V}{n}}. \tag{36}$$

In inequality (a), we used (28). In inequality (b), we used (27) and (34), noting that $\mathbb{E}[X'] = 0$. In inequality (c), we used the inequality $\mathbb{E}[|X'|] \leq \sqrt{\mathbb{E}[X'^2]} = \sqrt{\mathrm{Var}(X')} \leq \sqrt{V/n}$. Therefore, from (36), we have

$$\mathrm{Var}\left(\frac{\sqrt{n}C'(\Gamma)}{\sqrt{\psi_\beta(c(\mathbf{X}))}}\left(\Gamma - c(\mathbf{X})\right)\right)$$
$$\leq \mathrm{Var}\left(\frac{\sqrt{n}C'(\Gamma)}{\sqrt{V(\Gamma)}}\left(\Gamma - c(\mathbf{X})\right)\right) + \Delta$$
$$\leq \frac{C'(\Gamma)^2 V}{V(\Gamma)} + \Delta. \tag{37}$$

Define the random variable Π' as

$$\Pi' := \frac{\sqrt{n}C'(\Gamma)}{\sqrt{\psi_\beta(c(\mathbf{X}))}} \left(\Gamma - c(\mathbf{X})\right)$$

so that, from (35) and (37), $\mathbb{E}[\Pi'] \geq -\Delta$ and $\mathrm{Var}(\Pi') \leq \frac{C'(\Gamma)^2 V}{V(\Gamma)} + \Delta$. Then, from (33), for sufficiently large n,

$$(\overline{P} \circ W)\left(\log \frac{W(\mathbf{Y}|\mathbf{X})}{Q^*(\mathbf{Y})} > nC(\Gamma) + \sqrt{n}r\right)$$

$$\leq 1 + 2\frac{\kappa}{\sqrt{n}} - \inf_{\Pi'} \mathbb{E}\left[\Phi\left(\Pi' + \frac{r}{\sqrt{\varphi(r)}}\right)\right]$$

$$\overset{(a)}{=} 1 + 2\frac{\kappa}{\sqrt{n}} - \inf_{\Pi} \mathbb{E}\left[\Phi(\Pi)\right]$$

$$\overset{(b)}{=} 1 + 2\frac{\kappa}{\sqrt{n}} - \mathcal{K}\left(-\Delta + \frac{r}{\sqrt{\varphi(r)}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)} + \Delta\right). \tag{38}$$

The infimum in equality (a) above is over all random variables Π such that $\mathbb{E}[\Pi] \geq -\Delta + \frac{r}{\sqrt{\varphi(r)}}$ and $\mathrm{Var}(\Pi) \leq \frac{C'(\Gamma)^2 V}{V(\Gamma)} + \Delta$. Equality (b) follows by the definition and properties of the $\mathcal{K}$ function given in Definition 4 and Lemma 3.

Using (38) to upper bound (25) and using that upper bound in the result of Lemma 8 (with $\rho = \exp\left(nC(\Gamma) + \sqrt{n}r\right)$), we obtain

$$\log M^*(n, \epsilon, \Gamma, V) \leq nC(\Gamma) + \sqrt{n}r - \log\left[\left(-\epsilon - 2\frac{\kappa}{\sqrt{n}} + \mathcal{K}\left(-\Delta + \frac{r}{\sqrt{\varphi(r)}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)} + \Delta\right)\right)^+\right]$$

$$\frac{\log M^*(n, \epsilon, \Gamma, V) - nC(\Gamma)}{\sqrt{n}} \leq r - \frac{1}{\sqrt{n}}\log\left[\left(-\epsilon - 2\frac{\kappa}{\sqrt{n}} + \mathcal{K}\left(-\Delta + \frac{r}{\sqrt{\varphi(r)}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)} + \Delta\right)\right)^+\right]. \tag{39}$$

For any given average error probability $\epsilon \in (0,1)$, we choose r in (39) as

$$r = \begin{cases} r_1^* + \eta & \text{if } \epsilon \in (0, \epsilon^*] \\ r_2^* + \eta & \text{if } \epsilon \in (\epsilon^*, 1), \end{cases} \tag{40}$$

where

$$r_1^* := \max\left\{r' : \mathcal{K}\left(-\Delta + \frac{r'}{\sqrt{V(\Gamma) - \delta_\beta}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)} + \Delta\right) \leq \epsilon\right\}, \tag{41}$$

$$r_2^* := \max\left\{r' : \mathcal{K}\left(-\Delta + \frac{r'}{\sqrt{V(\Gamma) + \delta_\beta}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)} + \Delta\right) \leq \epsilon\right\}, \tag{42}$$

$$\epsilon^* := \mathcal{K}\left(-\Delta - \frac{\eta}{\sqrt{V(\Gamma) - \delta_\beta}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)} + \Delta\right).$$

Note that in (40), $r \leq 0$ for $\epsilon \leq \epsilon^*$ and $r > 0$ for $\epsilon > \epsilon^*$.

Therefore, for $\epsilon \leq \epsilon^*$, we have

$$\frac{\log M^*(n, \epsilon, \Gamma, V) - nC(\Gamma)}{\sqrt{n}} \leq r_1^* + \eta - \frac{1}{\sqrt{n}}\log\left[\left(-\epsilon - \frac{2\kappa}{\sqrt{n}} + \mathcal{K}\left(-\Delta + \frac{r_1^* + \eta}{\sqrt{V(\Gamma) - \delta_\beta}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)} + \Delta\right)\right)^+\right]. \tag{43}$$

Since $\mathcal{K}(\cdot, \cdot)$ is strictly increasing in the first argument, we have

$$\mathcal{K}\left(-\Delta + \frac{r_1^* + \eta}{\sqrt{V(\Gamma) - \delta_\beta}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)} + \Delta\right) > \epsilon$$

from the definition of r_1^* in (41) and the fact that $\eta > 0$. Hence, taking the limit supremum as $n \to \infty$ in (43), we obtain

$$\limsup_{n \to \infty} \frac{\log M^*(n, \epsilon, \Gamma, V) - nC(\Gamma)}{\sqrt{n}} \leq r_1^* + \eta \tag{44}$$

for $\epsilon \le \epsilon^*$. For $\epsilon > \epsilon^*$, a similar derivation gives us

$$\limsup_{n\to\infty} \frac{\log M^*(n, \epsilon, \Gamma, V) - nC(\Gamma)}{\sqrt{n}} \le r_2^* + \eta. \tag{45}$$

We now let Δ, η, β and δ_β go to zero in both (44) and (45). Then using the fact from Lemma 3 that $\mathcal{K}(\cdot, \cdot)$ is continuous, we obtain

$$\limsup_{n\to\infty} \frac{\log M^*(n, \epsilon, \Gamma, V) - nC(\Gamma)}{\sqrt{n}} \le \max\left\{ r : \mathcal{K}\left(\frac{r}{\sqrt{V(\Gamma)}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)} \right) \le \epsilon \right\}$$

for all $\epsilon \in (0, 1)$.

The converse result can also be stated in terms of a lower bound on the minimum average probability of error of (n, R, Γ, V) codes for a rate $R = C(\Gamma) + \frac{r}{\sqrt{n}}$. Starting again from Lemma 8, we have that for (n, R, Γ, V) codes with minimum average error probability ϵ,

$$\log\lceil \exp(nR) \rceil \le \log \rho - \log\left[\left(1 - \epsilon - \sup_{\overline{P} \in \mathcal{P}_{\Gamma, V}(\mathbb{R}^n)} \inf_{q \in \mathcal{P}(\mathbb{R}^n)} (\overline{P} \circ W)\left(\frac{W(\mathbf{Y}|\mathbf{X})}{q(\mathbf{Y})} > \rho \right) \right)^+ \right]. \tag{46}$$

Assume first that $r \le 0$ and let $\rho = \exp\left(nC(\Gamma) + r'\sqrt{n} \right)$ for some arbitrary $r' < r$. It directly follows from (38) and the definition of $\varphi(\cdot)$ in (29) that

$$\sup_{\overline{P} \in \mathcal{P}_{\Gamma, V}(\mathbb{R}^n)} \inf_{q \in \mathcal{P}(\mathbb{R}^n)} (\overline{P} \circ W)\left(\frac{W(\mathbf{Y}|\mathbf{X})}{q(\mathbf{Y})} > \rho \right)$$
$$\le 1 + 2\frac{\kappa}{\sqrt{n}} - \mathcal{K}\left(-\Delta + \frac{r}{\sqrt{V(\Gamma) - \delta_\beta}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)} + \Delta \right) \tag{47}$$

From (46) and (47), we have

$$\log\lceil \exp(nR) \rceil \le \log \rho - \log\left[\left(-\epsilon - 2\frac{\kappa}{\sqrt{n}} + \mathcal{K}\left(-\Delta + \frac{r'}{\sqrt{V(\Gamma) - \delta_\beta}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)} + \Delta \right) \right)^+ \right] \tag{48}$$

which evaluates to

$$-\epsilon - 2\frac{\kappa}{\sqrt{n}} + \mathcal{K}\left(-\Delta + \frac{r'}{\sqrt{V(\Gamma) - \delta_\beta}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)} + \Delta \right) \le e^{-(r - r')\sqrt{n}}.$$

Taking the limit as $n \to \infty$ and letting $r' \to r$, $\Delta \to 0$, $\beta \to 0$ and $\delta_\beta \to 0$, we obtain

$$\epsilon \ge \mathcal{K}\left(\frac{r}{\sqrt{V(\Gamma)}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)} \right) \tag{49}$$

for any $r \le 0$. For $r > 0$, let $\rho = \exp\left(nC(\Gamma) + r'\sqrt{n} \right)$ for some arbitrary $0 < r' < r$. Then from (38) and the definition of $\varphi(\cdot)$ in (29), we have that

$$\sup_{\overline{P} \in \mathcal{P}_{\Gamma, V}(\mathbb{R}^n)} \inf_{q \in \mathcal{P}(\mathbb{R}^n)} (\overline{P} \circ W)\left(\frac{W(\mathbf{Y}|\mathbf{X})}{q(\mathbf{Y})} > \rho \right)$$
$$\le 1 + 2\frac{\kappa}{\sqrt{n}} - \mathcal{K}\left(-\Delta + \frac{r'}{\sqrt{V(\Gamma) + \delta_\beta}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)} + \Delta \right). \tag{50}$$

Then a similar derivation to that used from (48) to (49) gives us

$$\epsilon \ge \mathcal{K}\left(\frac{r}{\sqrt{V(\Gamma)}}, \frac{C'(\Gamma)^2 V}{V(\Gamma)} \right)$$

for all $r > 0$.

∎

APPENDIX A
PROOF OF LEMMA 2

Let $Y \sim W(\cdot|x) = \mathcal{N}(x, N)$. We have

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

$$Q^*(y) = \frac{1}{\sqrt{2\pi(\Gamma + N)}} \exp\left(-\frac{y^2}{2(\Gamma + N)}\right)$$

$$\frac{W(Y|x)}{Q^*(Y)} = \frac{\sqrt{\Gamma + N}}{\sqrt{N}} \exp\left(-\frac{(Y-x)^2}{2N} + \frac{Y^2}{2(\Gamma + N)}\right)$$

$$\log\frac{W(Y|x)}{Q^*(Y)} = \frac{1}{2}\log\left(1 + \frac{\Gamma}{N}\right) - \frac{(Y-x)^2}{2N} + \frac{Y^2}{2(\Gamma + N)}$$

$$\mathbb{E}\left[\log\frac{W(Y|x)}{Q^*(Y)}\right] = \frac{1}{2}\log\left(1 + \frac{\Gamma}{N}\right) - \frac{\Gamma - x^2}{2(\Gamma + N)}.$$

Hence,

$$T_i = \frac{Y_i^2}{2(\Gamma + N)} - \frac{(Y_i - x_i)^2}{2N} + \frac{\Gamma - x_i^2}{2(\Gamma + N)}. \tag{51}$$

Using the relation $Y_i = x_i + Z_i$, we can write

$$\sum_{i=1}^n T_i = \sum_{i=1}^n \frac{(x_i + Z_i)^2}{2(\Gamma + N)} - \sum_{i=1}^n \frac{Z_i^2}{2N} + \frac{n\Gamma - nc(\mathbf{X})}{2(\Gamma + N)}$$

$$= \frac{nc(\mathbf{X})}{2(\Gamma + N)} + \sum_{i=1}^n \frac{Z_i^2}{2(\Gamma + N)} + \sum_{i=1}^n \frac{x_i Z_i}{\Gamma + N} - \sum_{i=1}^n \frac{Z_i^2}{2N} + \frac{n\Gamma - nc(\mathbf{X})}{2(\Gamma + N)}$$

$$= \sum_{i=1}^n \left[\frac{Z_i^2}{2(\Gamma + N)} + \frac{x_i Z_i}{\Gamma + N} - \frac{Z_i^2}{2N}\right] + \frac{n\Gamma}{2(\Gamma + N)}.$$

By completing the square and writing $Z_i = \sqrt{N}S_i$, where $\{S_i\}$ is i.i.d. $\mathcal{N}(0, 1)$, we can write

$$\sum_{i=1}^n T_i = -\frac{\Gamma}{2(\Gamma + N)}\sum_{i=1}^n \left(S_i - x_i\frac{\sqrt{N}}{\Gamma}\right)^2 + \frac{Nnc(\mathbf{X})}{2\Gamma(\Gamma + N)} + \frac{n\Gamma}{2(\Gamma + N)}.$$

Since

$$\sum_{i=1}^n \left(S_i - x_i\frac{\sqrt{N}}{\Gamma}\right)^2$$

has a noncentral chi-squared distribution with n degrees of freedom and noncentrality parameter λ given by

$$\lambda = \frac{Nnc(\mathbf{X})}{\Gamma^2},$$

the assertion of the lemma follows.

APPENDIX B
PROOF OF LEMMA 5

Define $\tilde{\mathbf{Y}} = \tilde{\mathbf{X}} + \tilde{\mathbf{Z}}$, where $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Z}}$ are independent, $\tilde{\mathbf{X}}$ is uniformly distributed on an $(n-1)$ sphere of radius $\frac{R}{\sqrt{N}}$ and $\tilde{\mathbf{Z}} \sim \mathcal{N}(\mathbf{0}, I_n)$. Let Q_1^{cc} denote the PDF of $\tilde{\mathbf{Y}}$. From [13, Proposition 1], we have

$$Q_1^{cc}(\tilde{\mathbf{y}}) = \frac{\Gamma\left(\frac{n}{2}\right)}{2\pi^{n/2}} \cdot \exp\left(-\frac{R^2 + N||\tilde{\mathbf{y}}||^2}{2N}\right)\left(\frac{\sqrt{N}}{R||\tilde{\mathbf{y}}||}\right)^{\frac{n}{2}-1} I_{\frac{n}{2}-1}\left(\frac{R||\tilde{\mathbf{y}}||}{\sqrt{N}}\right).$$

Since $\mathbf{Y} \stackrel{d}{=} \sqrt{N}\tilde{\mathbf{Y}}$, we can apply the change-of-variables formula

$$Q^{cc}(\mathbf{y}) = Q_1^{cc}\left(\frac{1}{\sqrt{N}}\mathbf{y}\right) \cdot \det\left(\frac{1}{\sqrt{N}}I_n\right)$$

to obtain the result.

APPENDIX C
PROOF OF LEMMA 6

From Lemma 5, we have

$$Q^{cc}(\mathbf{y}) = \frac{\Gamma\left(\frac{n}{2}\right)}{2(\pi N)^{n/2}} \cdot \exp\left(-\frac{n\Gamma + ||\mathbf{y}||^2}{2N}\right) \left(\frac{N}{\sqrt{n\Gamma}||\mathbf{y}||}\right)^{\frac{n}{2}-1} I_{\frac{n}{2}-1}\left(\frac{\sqrt{n\Gamma}||\mathbf{y}||}{N}\right)$$

and

$$Q_0^{cc}(\mathbf{y}) = \frac{\Gamma\left(\frac{n}{2}\right)}{2(\pi N)^{n/2}} \cdot \exp\left(-\frac{n\Gamma' + ||\mathbf{y}||^2}{2N}\right) \left(\frac{N}{\sqrt{n\Gamma'}||\mathbf{y}||}\right)^{\frac{n}{2}-1} I_{\frac{n}{2}-1}\left(\frac{\sqrt{n\Gamma'}||\mathbf{y}||}{N}\right).$$

Hence,

$$\frac{Q_0^{cc}(\mathbf{y})}{Q^{cc}(\mathbf{y})} = \frac{\exp\left(-\frac{n\Gamma'}{2N}\right)\left(\frac{N}{\sqrt{n\Gamma'}||\mathbf{y}||}\right)^{\frac{n}{2}-1} I_{\frac{n}{2}-1}\left(\frac{\sqrt{n\Gamma'}||\mathbf{y}||}{N}\right)}{\exp\left(-\frac{n\Gamma}{2N}\right)\left(\frac{N}{\sqrt{n\Gamma}||\mathbf{y}||}\right)^{\frac{n}{2}-1} I_{\frac{n}{2}-1}\left(\frac{\sqrt{n\Gamma}||\mathbf{y}||}{N}\right)}$$

$$= \exp\left(\frac{n}{2N}(\Gamma - \Gamma')\right) \left(\frac{\sqrt{n\Gamma}||\mathbf{y}||}{\sqrt{n\Gamma'}||\mathbf{y}||}\right)^{\frac{n}{2}-1} \frac{I_{\frac{n}{2}-1}\left(\frac{\sqrt{n\Gamma'}||\mathbf{y}||}{N}\right)}{I_{\frac{n}{2}-1}\left(\frac{\sqrt{n\Gamma}||\mathbf{y}||}{N}\right)}$$

$$= \exp\left(\frac{n}{2N}(\Gamma - \Gamma')\right) \left(\frac{\Gamma}{\Gamma'}\right)^{\frac{n}{4}-\frac{1}{2}} \frac{I_{\frac{n}{2}-1}\left(\frac{\sqrt{n\Gamma'}||\mathbf{y}||}{N}\right)}{I_{\frac{n}{2}-1}\left(\frac{\sqrt{n\Gamma}||\mathbf{y}||}{N}\right)}.$$

Hence,

$$\log\frac{Q_0^{cc}(\mathbf{y})}{Q^{cc}(\mathbf{y})}$$

$$= \frac{n}{2N}(\Gamma - \Gamma') + \left(\frac{n}{4} - \frac{1}{2}\right)\log\left(\frac{\Gamma}{\Gamma'}\right) + \log I_{\frac{n}{2}-1}\left(\frac{\sqrt{n\Gamma'}||\mathbf{y}||}{N}\right) - \log I_{\frac{n}{2}-1}\left(\frac{\sqrt{n\Gamma}||\mathbf{y}||}{N}\right)$$

To approximate the Bessel function, we first rewrite it as

$$I_{\frac{n}{2}-1}\left(\frac{\sqrt{n\Gamma}||\mathbf{y}||}{N}\right) = I_\nu\left(\nu z\right),$$

where $\nu = \frac{n}{2} - 1$ and $z = \frac{2\sqrt{n\Gamma}||\mathbf{y}||}{N(n-2)}$. Since $\mathbf{y} \in \mathcal{P}_n^*$ by assumption, we have that z lies in a compact interval $[a, b] \subset (0, \infty)$ for sufficiently large n, where $0 < a < b < \infty$. Hence, we can use a uniform asymptotic expansion of the modified Bessel function (see [14, 10.41.3] whose interpretation is given in [14, 2.1(iv)]): as $\nu \to \infty$, we have

$$I_\nu(\nu z) = \frac{e^{\nu\eta}}{(2\pi\nu)^{1/2}(1+z^2)^{\frac{1}{4}}}\left(1 + O\left(\frac{1}{\nu}\right)\right), \tag{52}$$

where

$$\eta = \sqrt{1+z^2} + \log\left(\frac{z}{1 + (1+z^2)^{1/2}}\right).$$

Since $z \in (0, 1)$, the $O(1/\nu)$ term in (52) can be uniformly bounded over $\mathcal{P}_n^*$. Using the approximation in (52), we have

$$\log I_{\frac{n}{2}-1}\left(\frac{\sqrt{n\Gamma}||\mathbf{y}||}{N}\right) = \frac{n}{2}\sqrt{1 + \frac{4||\mathbf{y}||^2 n\Gamma}{N^2(n-2)^2}} + \frac{n}{2}\log\left(\frac{2||\mathbf{y}||\sqrt{n\Gamma}}{N(n-2)}\right) - \frac{n}{2}\log\left(1 + \sqrt{1 + \frac{4||\mathbf{y}||^2 n\Gamma}{N^2(n-2)^2}}\right) + O(\log n),$$

$$\tag{53}$$

where it is easy to see that the $O(\log n)$ term can be made to be uniformly bounded over $\mathcal{P}_n^*$. Using (53), we have

$$
\log I_{\frac{n}{2}-1}\left(\frac{\sqrt{n\Gamma'}\|\mathbf{y}\|}{N}\right) - I_{\frac{n}{2}-1}\left(\frac{\sqrt{n\Gamma}\|\mathbf{y}\|}{N}\right)
$$

$$
= \frac{n}{2}\sqrt{1+\frac{4\|\mathbf{y}\|^2 n\Gamma'}{N^2(n-2)^2}} + \frac{n}{2}\log\left(\frac{2\|\mathbf{y}\|\sqrt{n\Gamma'}}{N(n-2)}\right) - \frac{n}{2}\log\left(1+\sqrt{1+\frac{4\|\mathbf{y}\|^2 n\Gamma'}{N^2(n-2)^2}}\right) -
$$

$$
\frac{n}{2}\sqrt{1+\frac{4\|\mathbf{y}\|^2 n\Gamma}{N^2(n-2)^2}} - \frac{n}{2}\log\left(\frac{2\|\mathbf{y}\|\sqrt{n\Gamma}}{N(n-2)}\right) + \frac{n}{2}\log\left(1+\sqrt{1+\frac{4\|\mathbf{y}\|^2 n\Gamma}{N^2(n-2)^2}}\right) + O(\log n)
$$

$$
= \frac{n}{2}\sqrt{1+\frac{4\|\mathbf{y}\|^2 n\Gamma'}{N^2(n-2)^2}} + \frac{n}{4}\log\left(\frac{\Gamma'}{\Gamma}\right) - \frac{n}{2}\log\left(1+\sqrt{1+\frac{4\|\mathbf{y}\|^2 n\Gamma'}{N^2(n-2)^2}}\right) -
$$

$$
\frac{n}{2}\sqrt{1+\frac{4\|\mathbf{y}\|^2 n\Gamma}{N^2(n-2)^2}} + \frac{n}{2}\log\left(1+\sqrt{1+\frac{4\|\mathbf{y}\|^2 n\Gamma}{N^2(n-2)^2}}\right) + O(\log n).
$$

Hence,

$$
\log \frac{Q_0^{cc}(\mathbf{y})}{Q^{cc}(\mathbf{y})}
$$

$$
= \frac{n}{2N}(\Gamma - \Gamma') + \left(\frac{n}{4}-\frac{1}{2}\right)\log\left(\frac{\Gamma}{\Gamma'}\right) +
$$

$$
\frac{n}{2}\sqrt{1+\frac{4\|\mathbf{y}\|^2 n\Gamma'}{N^2(n-2)^2}} + \frac{n}{4}\log\left(\frac{\Gamma'}{\Gamma}\right) - \frac{n}{2}\log\left(1+\sqrt{1+\frac{4\|\mathbf{y}\|^2 n\Gamma'}{N^2(n-2)^2}}\right) -
$$

$$
\frac{n}{2}\sqrt{1+\frac{4\|\mathbf{y}\|^2 n\Gamma}{N^2(n-2)^2}} + \frac{n}{2}\log\left(1+\sqrt{1+\frac{4\|\mathbf{y}\|^2 n\Gamma}{N^2(n-2)^2}}\right) + O(\log n)
$$

$$
= \frac{n}{2N}(\Gamma - \Gamma') + \frac{1}{2}\log\left(\frac{\Gamma'}{\Gamma}\right) + \frac{n}{2}\sqrt{1+\frac{4\|\mathbf{y}\|^2 n\Gamma'}{N^2(n-2)^2}} - \frac{n}{2}\log\left(1+\sqrt{1+\frac{4\|\mathbf{y}\|^2 n\Gamma'}{N^2(n-2)^2}}\right) -
$$

$$
\frac{n}{2}\sqrt{1+\frac{4\|\mathbf{y}\|^2 n\Gamma}{N^2(n-2)^2}} + \frac{n}{2}\log\left(1+\sqrt{1+\frac{4\|\mathbf{y}\|^2 n\Gamma}{N^2(n-2)^2}}\right) + O(\log n).
$$

To simplify the notation, we write $\delta = \frac{\|\mathbf{y}\|^2}{n} - (\Gamma + N)$. Recall that $\Gamma' = \Gamma + \epsilon$ and $|\delta| \leq \Delta$. Then

$$
\log \frac{Q_0^{cc}(\mathbf{y})}{Q^{cc}(\mathbf{y})}
$$

$$
= -\frac{n}{2N}\epsilon + \frac{1}{2}\log\left(1+\frac{\epsilon}{\Gamma}\right) + \frac{1}{2}g_n(\delta,\epsilon) - \frac{1}{2}h_n(\delta,\epsilon) - \frac{1}{2}g_n(\delta,0) + \frac{1}{2}h_n(\delta,0) + O(\log n), \tag{54}
$$

where we define

$$
g_n(\delta,\epsilon) := \sqrt{n^2 + \frac{4\Gamma(\Gamma+N)}{N^2}\frac{n^4}{(n-2)^2} + \frac{4\epsilon(\Gamma+N)}{N^2}\frac{n^4}{(n-2)^2} + \frac{4\Gamma}{N^2}\frac{n^4}{(n-2)^2}\delta + \frac{4\epsilon}{N^2}\frac{n^4}{(n-2)^2}\delta}
$$

$$
h_n(\delta,\epsilon) := n\log\left(1+\sqrt{1+\frac{4n^2\Gamma(\Gamma+N+\delta)}{N^2(n-2)^2}+\frac{4n^2\epsilon(\Gamma+N+\delta)}{N^2(n-2)^2}}\right).
$$

Using the Taylor series approximation

$$
\frac{n^2}{(n-2)^2} = 1 + \frac{4}{n} + O\left(\frac{1}{n^2}\right),
$$

we have

$$
\begin{aligned}
(g_n(\delta,\epsilon))^2 \\
= n^2 + \frac{4\Gamma(\Gamma+N)}{N^2}\left(n^2+4n+O(1)\right) + \frac{4(\Gamma+N)}{N^2}\left(n^2+4n+O(1)\right)\epsilon + \\
\frac{4\Gamma}{N^2}\left(n^2+4n+O(1)\right)\delta + \frac{4}{N^2}\left(n^2+4n+O(1)\right)\epsilon\delta \\
= n^2\left(1 + \frac{4\Gamma(\Gamma+N)}{N^2} + \frac{4(\Gamma+N)}{N^2}\epsilon + \frac{4\Gamma}{N^2}\delta + \frac{4}{N^2}\epsilon\delta\right) + \\
n\left(\frac{16\Gamma(\Gamma+N)}{N^2} + \frac{16(\Gamma+N)}{N^2}\epsilon + \frac{16\Gamma}{N^2}\delta + \frac{16}{N^2}\epsilon\delta\right) + O(1).
\end{aligned}
$$

Let

$$
\begin{aligned}
A &= 1 + \frac{4\Gamma(\Gamma+N)}{N^2} \\
B &= \frac{4(\Gamma+N)}{N^2} \\
C &= \frac{4\Gamma}{N^2} \\
D &= \frac{4}{N^2} \\
E &= \frac{16\Gamma(\Gamma+N)}{N^2}
\end{aligned}
$$

so that

$$
\begin{aligned}
g_n(\delta,\epsilon) \\
= \sqrt{n^2\left(A+B\epsilon+C\delta+D\epsilon\delta\right) + n\left(E+4B\epsilon+4C\delta+4D\epsilon\delta\right) + O(1)} \\
= n\sqrt{A+B\epsilon+C\delta+D\epsilon\delta + \frac{E+4B\epsilon+4C\delta+4D\epsilon\delta}{n} + O(n^{-2})}
\end{aligned}
$$

Further set $K = A+B\epsilon+C\delta+D\epsilon\delta$ and $L = E+4B\epsilon+4C\delta+4D\epsilon\delta$. Then

$$
\begin{aligned}
g_n(\delta,\epsilon) &= n\sqrt{K + \frac{L}{n} + O(n^{-2})} \\
&= n\sqrt{K}\sqrt{1 + \frac{L}{Kn} + O(n^{-2})} \\
&= n\sqrt{K}\left(1 + \frac{L}{2Kn} + O(n^{-2})\right) \\
&= n\sqrt{K} + \frac{L}{2\sqrt{K}} + O(n^{-1}).
\end{aligned}
$$

Now

$$
\begin{aligned}
\sqrt{K} &= \sqrt{A+B\epsilon+C\delta+D\epsilon\delta} \\
&= \sqrt{A}\sqrt{1 + \frac{B}{A}\epsilon + \frac{C}{A}\delta + \frac{D}{A}\epsilon\delta} \\
&= \sqrt{A}\left(1 + \frac{B}{2A}\epsilon + \frac{C}{2A}\delta + O(\epsilon^2,\Delta^2,\epsilon\Delta)\right) \\
&= \sqrt{A} + \frac{B}{2\sqrt{A}}\epsilon + \frac{C}{2\sqrt{A}}\delta + O(\epsilon^2,\Delta^2,\epsilon\Delta).
\end{aligned}
$$

Thus

$$
n\sqrt{K} = n\sqrt{A} + \frac{B}{2\sqrt{A}}n\epsilon + \frac{C}{2\sqrt{A}}n\delta + O(n\epsilon^2,n\Delta^2,n\epsilon\Delta).
$$

Combining all the terms, we have

$$
g_n(\delta,\epsilon) = n\sqrt{1 + \frac{4\Gamma(\Gamma+N)}{N^2}} + \frac{4(\Gamma+N)}{2N^2\sqrt{1 + \frac{4\Gamma(\Gamma+N)}{N^2}}}n\epsilon + \frac{4\Gamma}{2N^2\sqrt{1 + \frac{4\Gamma(\Gamma+N)}{N^2}}}n\delta + O(1) + O(n\epsilon^2,n\Delta^2,n\epsilon\Delta).
$$

Note that

$$\sqrt{A} = \sqrt{1 + \frac{4\Gamma(\Gamma + N)}{N^2}} = \frac{2\Gamma + N}{N}.$$

Hence,

$$g_n(\delta, \epsilon) = n\frac{2\Gamma + N}{N} + \frac{2(\Gamma + N)}{N(2\Gamma + N)}n\epsilon + \frac{2\Gamma}{N(2\Gamma + N)}n\delta + O(1) + O(n\epsilon^2, n\Delta^2, n\epsilon\Delta). \tag{55}$$

Now we turn to

$$h_n(\delta, \epsilon) = n\log\left(1 + \sqrt{1 + \frac{4n^2\Gamma(\Gamma + N + \delta)}{N^2(n-2)^2} + \frac{4n^2\epsilon(\Gamma + N + \delta)}{N^2(n-2)^2}}\right).$$

We first simplify the expression inside the square root as follows:

$$\begin{aligned}
&1 + \frac{4n^2\Gamma(\Gamma + N + \delta)}{N^2(n-2)^2} + \frac{4n^2\epsilon(\Gamma + N + \delta)}{N^2(n-2)^2} \\
&= 1 + \frac{4\Gamma(\Gamma + N + \delta)}{N^2}\left(1 + \frac{4}{n} + O\left(\frac{1}{n^2}\right)\right) + \frac{4\epsilon(\Gamma + N + \delta)}{N^2}\left(1 + \frac{4}{n} + O\left(\frac{1}{n^2}\right)\right) \\
&= 1 + \frac{4\Gamma(\Gamma + N + \delta)}{N^2} + \frac{4\epsilon(\Gamma + N + \delta)}{N^2} + O\left(\frac{1}{n}\right) \\
&= A + C\delta + B\epsilon + D\epsilon\delta + O(n^{-1}) \\
&= A\left(1 + \frac{C}{A}\delta + \frac{B}{A}\epsilon + \frac{D}{A}\epsilon\delta + O(n^{-1})\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sqrt{A}\sqrt{1 + \frac{C}{A}\delta + \frac{B}{A}\epsilon + \frac{D}{A}\epsilon\delta + O(n^{-1})} \\
&= \sqrt{A}\left(1 + \frac{C}{2A}\delta + \frac{B}{2A}\epsilon + O(n^{-1}, \Delta^2, \epsilon^2, \epsilon\Delta)\right) \\
&= \sqrt{A} + \frac{C}{2\sqrt{A}}\delta + \frac{B}{2\sqrt{A}}\epsilon + O(n^{-1}, \Delta^2, \epsilon^2, \epsilon\Delta).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\log\left(1 + \sqrt{A} + \frac{C}{2\sqrt{A}}\delta + \frac{B}{2\sqrt{A}}\epsilon + O(n^{-1}, \Delta^2, \epsilon^2, \epsilon\Delta)\right) \\
&= \log\left(1 + \sqrt{A}\right) + \frac{1}{1 + \sqrt{A}}\left(\frac{C}{2\sqrt{A}}\delta + \frac{B}{2\sqrt{A}}\epsilon\right) + O(n^{-1}, \Delta^2, \epsilon^2, \epsilon\Delta) \\
&= \log\left(2\frac{\Gamma + N}{N}\right) + \frac{N}{2(\Gamma + N)}\frac{4\Gamma}{2N(2\Gamma + N)}\delta + \frac{N}{2(\Gamma + N)}\frac{4(\Gamma + N)}{2N(2\Gamma + N)}\epsilon + O(n^{-1}, \Delta^2, \epsilon^2, \epsilon\Delta) \\
&= \log\left(2\frac{\Gamma + N}{N}\right) + \frac{\Gamma}{(2\Gamma + N)(\Gamma + N)}\delta + \frac{1}{2\Gamma + N}\epsilon + O(n^{-1}, \Delta^2, \epsilon^2, \epsilon\Delta).
\end{aligned}$$

Finally,

$$h_n(\delta, \epsilon) = n\log\left(2\frac{\Gamma + N}{N}\right) + \frac{\Gamma}{(2\Gamma + N)(\Gamma + N)}n\delta + \frac{1}{2\Gamma + N}n\epsilon + O(n\Delta^2, n\epsilon^2, n\epsilon\Delta) + O(1). \tag{56}$$

Substituting (55) and (56) in (54), we obtain

$$\log \frac{Q_0^{cc}(\mathbf{y})}{Q^{cc}(\mathbf{y})}$$

$$= -\frac{n}{2N}\epsilon + \frac{1}{2}\log\left(1 + \frac{\epsilon}{\Gamma}\right) + \frac{1}{2}g_n(\delta,\epsilon) - \frac{1}{2}h_n(\delta,\epsilon) - \frac{1}{2}g_n(\delta,0) + \frac{1}{2}h_n(\delta,0) + O(\log n)$$

$$= -\frac{n}{2N}\epsilon + \frac{1}{2}\log\left(1 + \frac{\epsilon}{\Gamma}\right) + \frac{1}{2}\left(n\frac{2\Gamma+N}{N} + \frac{2(\Gamma+N)}{N(2\Gamma+N)}n\epsilon + \frac{2\Gamma}{N(2\Gamma+N)}n\delta + O(1) + O(n\epsilon^2, n\Delta^2, n\epsilon\Delta)\right) -$$

$$\frac{1}{2}\left(n\log\left(2\frac{\Gamma+N}{N}\right) + \frac{\Gamma}{(2\Gamma+N)(\Gamma+N)}n\delta + \frac{1}{2\Gamma+N}n\epsilon + O(n\Delta^2, n\epsilon^2, n\epsilon\Delta) + O(1)\right) + O(\log n) -$$

$$\frac{1}{2}\left(n\frac{2\Gamma+N}{N} + \frac{2\Gamma}{N(2\Gamma+N)}n\delta + O(1) + O(n\Delta^2)\right) +$$

$$\frac{1}{2}\left(n\log\left(2\frac{\Gamma+N}{N}\right) + \frac{\Gamma}{(2\Gamma+N)(\Gamma+N)}n\delta + O(n\Delta^2) + O(1)\right)$$

$$= -\frac{n}{2N}\epsilon + \frac{1}{2}\frac{2(\Gamma+N)}{N(2\Gamma+N)}n\epsilon + O(1) + O(n\epsilon^2, n\Delta^2, n\epsilon\Delta) - \frac{1}{2}\frac{1}{2\Gamma+N}n\epsilon + O(\log n)$$

$$= O(\log n) + O(n\epsilon^2, n\Delta^2, n\epsilon\Delta).$$

APPENDIX D
PROOF OF LEMMA 7

From Lemma 5, we have

$$Q^{cc}(\mathbf{y}) = \frac{\Gamma\left(\frac{n}{2}\right)}{2(\pi N)^{n/2}} \cdot \exp\left(-\frac{n\Gamma + ||\mathbf{y}||^2}{2N}\right)\left(\frac{N}{\sqrt{n\Gamma}||\mathbf{y}||}\right)^{\frac{n}{2}-1} I_{\frac{n}{2}-1}\left(\frac{\sqrt{n\Gamma}||\mathbf{y}||}{N}\right).$$

From the standard formula for the multivariate Gaussian, we have

$$Q^*(\mathbf{y}) = \frac{1}{(2\pi(\Gamma+N))^{n/2}}\exp\left(-\frac{1}{2(\Gamma+N)}||\mathbf{y}||^2\right).$$

Then

$$\log\frac{Q^{cc}(\mathbf{y})}{Q^*(\mathbf{y})}$$

$$= \log\left(\Gamma\left(\frac{n}{2}\right)\right) - \frac{n\Gamma + ||\mathbf{y}||^2}{2N} + \frac{n}{2}\log\left(\frac{N}{\sqrt{n\Gamma}||\mathbf{y}||}\right) - \log\left(\frac{N}{\sqrt{n\Gamma}||\mathbf{y}||}\right) + \log\left(I_{\frac{n}{2}-1}\left(\frac{\sqrt{n\Gamma}||\mathbf{y}||}{N}\right)\right)$$

$$- \log(2) - \frac{n}{2}\log(\pi N) + \frac{n}{2}\log(2\pi(\Gamma+N)) + \frac{1}{2(\Gamma+N)}||\mathbf{y}||^2$$

$$= \log\left(\Gamma\left(\frac{n}{2}\right)\right) - \frac{||\mathbf{y}||^2\Gamma}{2N(\Gamma+N)} - \frac{n\Gamma}{2N} + \frac{n}{2}\log\left(\frac{N}{\sqrt{n\Gamma}||\mathbf{y}||}\right) - \log\left(\frac{N}{\sqrt{n\Gamma}||\mathbf{y}||}\right) + \log\left(I_{\frac{n}{2}-1}\left(\frac{\sqrt{n\Gamma}||\mathbf{y}||}{N}\right)\right)$$

$$- \log(2) + \frac{n}{2}\log\left(\frac{2\Gamma+2N}{N}\right)$$

$$\overset{(a)}{=} \frac{n}{2}\log\left(\frac{n}{2}\right) - \frac{n}{2} - \frac{1}{2}\log\left(\frac{n}{2}\right) + O(1) - \frac{||\mathbf{y}||^2\Gamma}{2N(\Gamma+N)} - \frac{n\Gamma}{2N} + \frac{n}{2}\log\left(\frac{N}{\sqrt{n\Gamma}||\mathbf{y}||}\right) - \log\left(\frac{N}{\sqrt{n\Gamma}||\mathbf{y}||}\right) +$$

$$\log\left(I_{\frac{n}{2}-1}\left(\frac{\sqrt{n\Gamma}||\mathbf{y}||}{N}\right)\right) - \log(2) + \frac{n}{2}\log\left(\frac{2\Gamma+2N}{N}\right)$$

$$= \frac{n}{2}\log(n) - \frac{n}{2} - \frac{1}{2}\log\left(\frac{n}{2}\right) - \frac{||\mathbf{y}||^2\Gamma}{2N(\Gamma+N)} - \frac{n\Gamma}{2N} + \frac{n}{2}\log\left(\frac{N}{\sqrt{n\Gamma}||\mathbf{y}||}\right) - \log\left(\frac{N}{\sqrt{n\Gamma}||\mathbf{y}||}\right) +$$

$$\log\left(I_{\frac{n}{2}-1}\left(\frac{\sqrt{n\Gamma}||\mathbf{y}||}{N}\right)\right) + \frac{n}{2}\log\left(\frac{\Gamma+N}{N}\right) + O(1) \tag{57}$$

In equality (a), we used an asymptotic expansion of the log gamma function (see, e.g., [14, 5.11.1]). To approximate the Bessel function, we first rewrite it as

$$I_{\frac{n}{2}-1}\left(\frac{\sqrt{n\Gamma}||\mathbf{y}||}{N}\right) = I_\nu\left(\nu z\right),$$

where $\nu = \frac{n}{2} - 1$ and $z = \frac{2\sqrt{n\Gamma}||\mathbf{y}||}{N(n-2)}$. Since $\mathbf{y} \in \mathcal{P}_n^*$ by assumption, we have that z lies in a compact interval $[a, b] \subset (0, \infty)$ for sufficiently large n, where $0 < a < b < \infty$. Hence, we can use a uniform asymptotic expansion of the modified Bessel function (see [14, 10.41.3] whose interpretation is given in [14, 2.1(iv)]): as $\nu \to \infty$, we have

$$I_\nu(\nu z) = \frac{e^{\nu\eta}}{(2\pi\nu)^{1/2}(1+z^2)^{\frac{1}{4}}}\left(1 + O\left(\frac{1}{\nu}\right)\right), \tag{58}$$

where

$$\eta = \sqrt{1+z^2} + \log\left(\frac{z}{1+(1+z^2)^{1/2}}\right).$$

Since $z \in (0, 1)$, the $O(1/\nu)$ term in (58) can be uniformly bounded over $\mathcal{P}_n^*$. Using the approximation in (58), we have

$$\log I_{\frac{n}{2}-1}\left(\frac{\sqrt{n\Gamma}||\mathbf{y}||}{N}\right) = \frac{n}{2}\sqrt{1 + \frac{4||\mathbf{y}||^2 n\Gamma}{N^2(n-2)^2}} + \frac{n}{2}\log\left(\frac{2||\mathbf{y}||\sqrt{n\Gamma}}{N(n-2)}\right) - \frac{n}{2}\log\left(1 + \sqrt{1 + \frac{4||\mathbf{y}||^2 n\Gamma}{N^2(n-2)^2}}\right) + O(\log n), \tag{59}$$

where it is easy to see that the $O(\log n)$ term can be made to be uniformly bounded over $\mathcal{P}_n^*$. Substituting (59) in (57), we obtain

$$\log\frac{Q^{cc}(\mathbf{y})}{Q^*(\mathbf{y})}$$

$$= \frac{n}{2}\log(n) - \frac{n}{2} - \frac{||\mathbf{y}||^2\Gamma}{2N(\Gamma+N)} - \frac{n\Gamma}{2N} + \frac{n}{2}\log\left(\frac{N}{\sqrt{n\Gamma}||\mathbf{y}||}\right) - \log\left(\frac{N}{\sqrt{n\Gamma}||\mathbf{y}||}\right) + \frac{n}{2}\sqrt{1 + \frac{4||\mathbf{y}||^2 n\Gamma}{N^2(n-2)^2}} +$$

$$\frac{n}{2}\log\left(\frac{2||\mathbf{y}||\sqrt{n\Gamma}}{N(n-2)}\right) - \frac{n}{2}\log\left(1 + \sqrt{1 + \frac{4||\mathbf{y}||^2 n\Gamma}{N^2(n-2)^2}}\right) + \frac{n}{2}\log\left(\frac{\Gamma+N}{N}\right) + O(\log n)$$

$$= -\frac{n}{2} - \frac{||\mathbf{y}||^2\Gamma}{2N(\Gamma+N)} - \frac{n\Gamma}{2N} + \frac{n}{2}\sqrt{1 + \frac{4||\mathbf{y}||^2 n\Gamma}{N^2(n-2)^2}} - \frac{n}{2}\log\left(1 + \sqrt{1 + \frac{4||\mathbf{y}||^2 n\Gamma}{N^2(n-2)^2}}\right)$$

$$+ \frac{n}{2}\log\left(\frac{2\Gamma+2N}{N}\right) + \log\left(\sqrt{n\Gamma}||\mathbf{y}||\right) + O(\log n)$$

$$= \left[\frac{1}{2}g_n(\delta, 0) - \frac{\Gamma}{N}n - \frac{n}{2} - \frac{\Gamma}{2N(\Gamma+N)}n\delta\right] + \left[\frac{n}{2}\log\left(\frac{2\Gamma+2N}{N}\right) - \frac{1}{2}h_n(\delta, 0)\right] + \log\left(\sqrt{n\Gamma}||\mathbf{y}||\right) + O(\log n), \tag{60}$$

where in the last equality above, we have

$$\delta = \frac{||\mathbf{y}||^2}{n} - (\Gamma + N) \tag{61}$$

$$g_n(\delta, 0) = \sqrt{n^2 + \frac{4\Gamma(\Gamma+N)}{N^2}\frac{n^4}{(n-2)^2} + \frac{4\Gamma}{N^2}\frac{n^4}{(n-2)^2}\delta}$$

$$h_n(\delta, 0) = n\log\left(1 + \sqrt{1 + \frac{4n^2\Gamma(\Gamma+N+\delta)}{N^2(n-2)^2}}\right)$$

to facilitate further analysis. Recall that $|\delta| \leq \Delta$.

From (55), we have

$$g_n(\delta, 0) = n\frac{2\Gamma+N}{N} + \frac{2\Gamma}{N(2\Gamma+N)}n\delta + O(n\Delta^2) + O(1).$$

Thus, we can simplify the first term in square brackets in (60) as

$$\frac{1}{2}g_n(\delta) - \frac{\Gamma}{N}n - \frac{n}{2} - \frac{\Gamma}{2N(\Gamma+N)}n\delta$$

$$= \frac{\Gamma}{N(2\Gamma+N)}n\delta + O(n\Delta^2) + O(1) - \frac{\Gamma}{2N(\Gamma+N)}n\delta$$

$$= \frac{\Gamma}{2(2\Gamma+N)(\Gamma+N)}n\delta + O(n\Delta^2) + O(1). \tag{62}$$

Similarly, from (56), we have

$$h_n(\delta, 0) = n\log\left(\frac{2\Gamma+2N}{N}\right) + \frac{\Gamma}{(2\Gamma+N)(\Gamma+N)}n\delta + O(1) + O\left(n\Delta^2\right).$$

Thus, we can simplify the second term in square brackets in (60) as

$$\frac{n}{2}\log\left(\frac{2\Gamma + 2N}{N}\right) - \frac{n}{2}\log\left(1 + \sqrt{1 + \frac{4n^2\Gamma(\Gamma + N + \delta)}{N^2(n-2)^2}}\right)$$
$$= -\frac{\Gamma}{2(2\Gamma + N)(\Gamma + N)}n\delta + O(1) + O\left(n\Delta^2\right). \tag{63}$$

Adding (62) and (63) gives us $O(n\Delta^2) + O(1)$. Hence, going back to (60), we have

$$\log\frac{Q^{cc}(\mathbf{y})}{Q^*(\mathbf{y})}$$
$$= \log\left(\sqrt{n\Gamma}||\mathbf{y}||\right) + O(\log n) + O(n\Delta^2)$$
$$= O(\log n) + O(n\Delta^2),$$

where the last equality follows from the assumption that $\mathbf{y} \in \mathcal{P}_n^*$.

REFERENCES

[1] A. Mahmood and A. B. Wagner, "Channel coding with mean and variance cost constraints," in *2024 IEEE International Symposium on Information Theory (ISIT)*, 2024, pp. 510–515.

[2] ——, "Improved channel coding performance through cost variability," 2024. [Online]. Available: https://arxiv.org/abs/2407.05260

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, N.J.: Wiley-Interscience, 2006.

[4] Y. Polyanskiy, "Channel coding: Non-asymptotic fundamental limits," Ph.D. dissertation, Dept. Elect. Eng., Princeton Univ., Princeton, NJ, USA, 2010.

[5] W. Yang, G. Caire, G. Durisi, and Y. Polyanskiy, "Optimum power control at finite blocklength," *IEEE Transactions on Information Theory*, vol. 61, no. 9, pp. 4598–4615, 2015.

[6] S. L. Fong and V. Y. F. Tan, "Asymptotic expansions for the awgn channel with feedback under a peak power constraint," in *2015 IEEE International Symposium on Information Theory (ISIT)*, 2015, pp. 311–315.

[7] ——, "A tight upper bound on the second-order coding rate of the parallel Gaussian channel with feedback," *IEEE Transactions on Information Theory*, vol. 63, no. 10, pp. 6474–6486, 2017.

[8] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[9] A. Mahmood and A. B. Wagner, "Channel coding with mean and variance cost constraints," 2024. [Online]. Available: https://arxiv.org/abs/2401.16417

[10] A. B. Wagner, N. V. Shende, and Y. Altuǧ, "A new method for employing feedback to improve coding performance," *IEEE Transactions on Information Theory*, vol. 66, no. 11, pp. 6660–6681, 2020.

[11] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 4947–4966, 2009.

[12] P. van Beek, "An application of Fourier methods to the problem of sharpening the Berry-Esseen inequality," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 23, no. 3, pp. 187–196, 1972.

[13] A. Dytso, M. Al, H. V. Poor, and S. Shamai Shitz, "On the capacity of the peak power constrained vector Gaussian channel: An estimation theoretic perspective," *IEEE Transactions on Information Theory*, vol. 65, no. 6, pp. 3907–3921, 2019.

[14] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, Eds., *NIST Handbook of Mathematical Functions*. New York: Cambridge University Press, 2010.