On the Dispersions of the Gel'fand-Pinsker Channel and Dirty Paper Coding

Jonathan Scarlett

Abstract

This paper studies second-order coding rates for memoryless channels with a state sequence known non-causally at the encoder. In the case of finite alphabets, an achievability result is obtained using constant-composition random coding, and by using a small fraction of the block to transmit the type of the state sequence. For error probabilities less than $\frac{1}{2}$, it is shown that the second-order rate improves on an existing one based on i.i.d. random coding. In the Gaussian case (dirty paper coding) with an almost-sure power constraint, an achievability result is obtained used using random coding over the surface of a sphere, and using a small fraction of the block to transmit a quantized description of the state power. It is shown that the second-order asymptotics are identical to the single-user Gaussian channel of the same input power without a state.

Index Terms

Channels with state, Gel'fand-Pinsker channel, dirty paper coding, channel dispersion, second-order coding rate

I. Introduction

The problem of characterizing the second-order asymptotics of the highest achievable channel coding rate at a given error probability and increasing block length was studied by Strassen [1], and has recently regained significant attention following the works of Polyanskiy *et al.* [2] and Hayashi [3]. For discrete memoryless channels, the maximum number of codewords $M^*(n,\epsilon)$ of length n yielding an average error probability not exceeding $\epsilon \in (0,1)$ satisfies

$$\log M^*(n,\epsilon) = nC - \sqrt{nV}\mathsf{Q}^{-1}(\epsilon) + o(\sqrt{n}), \tag{1}$$

where C is the channel capacity, $\mathsf{Q}^{-1}(\cdot)$ is the inverse of the Q-function, and V is known as the *channel dispersion*. We can interpret C and V as being the mean and variance of the information density $i(x,y) \triangleq \log \frac{W(y|x)}{\sum_{\overline{x}} Q(\overline{x})W(y|\overline{x})}$ for some capacity-achieving input distribution Q. For the additive white Gaussian noise (AWGN) channel with maximal power P, an expansion of the form (1) holds with $V = \frac{P(2+P)}{2(1+P)^2}$ [2], [3].

J. Scarlett is with the Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, U.K. (e-mail: jmscarlett@gmail.com).

This work has been funded in part by the European Research Council under ERC grant agreement 259663, by the European Union's 7th Framework Programme (PEOPLE-2011-CIG) under grant agreement 303633 and by the Spanish Ministry of Economy and Competitiveness under grant TEC2012-38800-C03-03.

This work was presented in part at the International Zurich Seminar on Communiations, 2014.

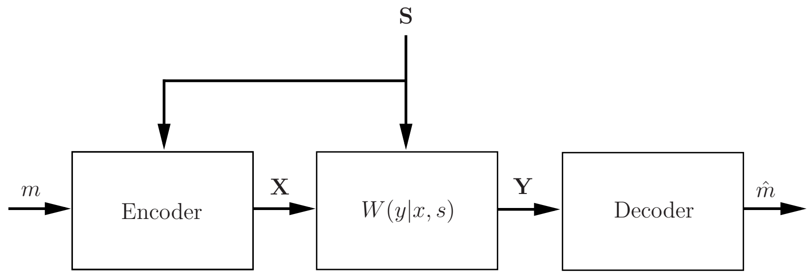

Figure 1. Channel model.

In this paper, we study the second-order asymptotics of channel coding with a random state known non-causally at the encoder, as studied by Gel'fand-Pinsker [4] and Costa [5] (see Figure 1). In the case of finite alphabets and unconstrained inputs, we give an achievability result of the form (1), with $\geq$ in place of the equality. In the case that the channel is Gaussian and the input is subject to an almost-sure power constraint (dirty paper coding [5]), we show that the second-order asymptotics are identical to those obtained when the state is absent, thus strengthening the well-known analogous result for the capacity.

A. Channel Model and Capacity

The alphabets of the input, output and state are denoted by $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{S}$ respectively. The channel is assumed to be memoryless with a transition law $W(y|x,s)$, and the state sequence $\boldsymbol{S} = (S_1, \cdots, S_n)$ is distributed according to $P_{\boldsymbol{S}}(\boldsymbol{s})$. The n-letter channel transition law is given by

$$W^n(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{s}) \triangleq \prod_{i=1}^{n} W(y_i|x_i,s_i). \tag{2}$$

The encoder takes as input the state sequence $\boldsymbol{s}$ and a message m equiprobable on the set $\{1, \cdots, M\}$, and transmits a codeword $\boldsymbol{x}^{(m)}(\boldsymbol{s})$. The decoder forms an estimate $\hat{m}$ based on $\boldsymbol{y}$, and an error is said to have occurred if $\hat{m} \neq m$. We study two variations of this setup, which we refer to as the discrete case and the Gaussian case.

In the discrete case, the alphabets $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{S}$ are assumed to be finite, the channel input is assumed to be unconstrained, and the state distribution is assumed to be i.i.d. on some distribution π, i.e. $P_{\boldsymbol{S}}(\boldsymbol{s}) = \prod_{i=1}^{n} \pi(s_i)$. The capacity is given by [4]

$$C = \max_{\mathcal{U}, Q_{U|S}, \phi(\cdot,\cdot)} I(U;Y) - I(U;S), \tag{3}$$

where the mutual informations are computed using the distribution

$$P_{SUY}(s,u,y) = \pi(s) Q_{U|S}(u|s) W(y|\phi(u,s),s) \tag{4}$$

and the maximum is over all finite alphabets $\mathcal{U}$, conditional distributions $Q_{U|S}$ and functions $\phi : \mathcal{U} \times \mathcal{S} \to \mathcal{X}$.

In the Gaussian case, the channel is described by

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{S} + \boldsymbol{Z}, \tag{5}$$

where $\boldsymbol{Z}$ is an i.i.d. noise sequence with $Z_i \sim N(0,1)$. That is, we have

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

The state distribution $P_{\boldsymbol{S}}$ is assumed to be arbitrary for now. The codewords are constrained to satisfy a power constraint of the form

$$\|\boldsymbol{x}^{(m)}(\boldsymbol{s})\|^2 \le nP, \qquad \forall m, \boldsymbol{s} \tag{7}$$

for some transmit power P. That is, we require $\|\boldsymbol{X}\|^2 \le nP$ almost surely. This is a stricter constraint than that considered in some previous works; see Section I-B for further discussion.

Here we provide an outline of the capacity results; see [5], [6, Sec. 7.7] for details. We first consider the case that $P_{\boldsymbol{S}}$ is i.i.d. on $\pi \sim N(0, P_\pi)$ for some $P_\pi > 0$. The capacity is given by (3) subject to the constraint $\mathbb{E}[\phi(U,S)^2] \le P$. We fix $\alpha > 0$ and choose

$$Q_{U|S}(\cdot|s) \sim N(-\alpha s, P) \tag{8}$$

$$\phi(u,s) = u - \alpha s, \tag{9}$$

which can be equivalently be written as

$$U = X + \alpha S \tag{10}$$

$$X \sim N(0,P), \tag{11}$$

where X is independent of S. Under these parameters, it can be shown that

$$I(U;Y) = \frac{1}{2}\log\left(\frac{(P+P_\pi+1)(P+\alpha^2 P_\pi)}{PP_\pi(1-\alpha)^2 + (P+\alpha^2 P_\pi)}\right) \tag{12}$$

$$I(U;S) = \frac{1}{2}\log\left(\frac{P+\alpha^2 P_\pi}{P}\right). \tag{13}$$

Furthermore, the optimal choice of α is given by

$$\alpha = \frac{P}{1+P}, \tag{14}$$

and yields

$$C = \frac{1}{2}\log(1+P). \tag{15}$$

Thus, the capacity is independent of the state power Π, and is the same as if the state sequence were absent (or equivalently, if it were known at the decoder).

In the case of a non-Gaussian i.i.d. state sequence with $\mathbb{E}[S_i^2] < \infty$, the capacity remains the same. For example, see [6, Sec. 7.7] for a proof based on connections with minimum mean square error (MMSE) estimation. Although we consider a general (possibly non-Gaussian and non-i.i.d.) state distribution in this paper, the parameter choices and mutual informations in (8)–(14) will play a major role in the analysis.

B. Discussion: Power Constraints and Common Randomness for Dirty Paper Coding

In general, the fundamental performance limits of channels with power constraints can vary depending on (i) the type of power constraint (e.g. almost-sure vs. averaged over a random variable), (ii) the availability of common randomness at the encoder and decoder, and (iii) whether the average or maximal error probability is being considered. Here we focus on the case of average error probability, and discuss some variations of dirty paper coding in which the former two properties differ. For each case we consider, the capacity will remain equal to (15), at least subject to mild technical conditions on $P_{\boldsymbol{S}}$.

Suppose that the power constraint is averaged over the randomness of the message and the state. In this case, we can show that the strong converse fails to hold, similarly to the AWGN channel without state [7, Sec. 4.3.3]. Since the capacity is given by (15), there exists a code of average power $\frac{P}{1-\epsilon}$, rate approaching $\frac{1}{2}\log\left(1+\frac{P}{1-\epsilon}\right)$, and vanishing error probability. By replacing the fraction ϵ of the codewords $\boldsymbol{x}^{(m)}(\boldsymbol{s})$ with the highest power (averaged over $\boldsymbol{S}$) by the all-zero codeword, we obtain a code of average power not exceeding P, rate approaching $\frac{1}{2}\log\left(1+\frac{P}{1-\epsilon}\right)$, and error probability approaching ϵ. Since the strong converse does hold under a maximal power constraint [7], we conclude that the existence of a code satisfying an average power constraint does not, in general, imply the existence of a code satisfying a maximal power constraint and having the same asymptotic rate and error probability.

The study of lattice coding for the dirty paper coding problem generally makes use of common randomness at the encoder and decoder in the form of a dither; see [8]–[10] and references therein. The power constraint considered in these works holds for all messages and state sequences, but it is averaged over the randomness of the dither. To the best of the author's judgment, the removal of this common randomness (if possible) would require relaxing the power constraint to be averaged over the message and state, and would thus recover the setting discussed in the previous paragraph in which the strong converse fails to hold.

As seen in (7), the setting we consider is stricter in the sense that the power constraint is an almost-sure constraint, and no common randomness is assumed. This is the same setup as that considered in [5], [11], among others.

C. Previous Work

For unconstrained channels with state known at the encoder, Watanabe *et al.* [12] and Yassaee *et al.* [13] provided alternative derivations of the same result using different techniques based on i.i.d. random coding. In order to state the result, we introduce some definitions. We say that a triplet (n, M, ϵ) is achievable if there exists a code with block length n containing at least M messages and yielding an average error probability not exceeding ϵ, and we define

$$M^*(n,\epsilon) \triangleq \max\Big\{M \,:\, (n,M,\epsilon) \text{ is achievable}\Big\}. \tag{16}$$

Letting $P_{Y|U}$, P_Y, etc. denote the marginals of (4), we define the information densities

$$i(u,s) \triangleq \log\frac{Q_{U|S}(u|s)}{P_U(u)} \tag{17}$$

$$i(u,y) \triangleq \log\frac{P_{Y|U}(y|u)}{P_Y(y)} \tag{18}$$

with a slight abuse of notation. Furthermore, for a 2×2 positive semi-definite matrix $\boldsymbol{V}$, we define the set

$$\mathsf{Q}_{\mathrm{inv}}(\boldsymbol{V}, \epsilon) \triangleq \Big\{ \boldsymbol{z} \in \mathbb{R}^2 \,:\, \mathbb{P}\big[\boldsymbol{Z} \preceq \boldsymbol{z}\big] \geq 1 - \epsilon \Big\}, \tag{19}$$

where $\boldsymbol{Z} \sim N(\mathbf{0}, \boldsymbol{V})$, and $\preceq$ denotes element-wise inequality. It was shown in [12], [13] that

$$\log M^*(n, \epsilon) \geq nC - \sqrt{n}\tilde{R} + O(\log n), \tag{20}$$

where

$$\tilde{R} \triangleq \min_{(\tilde{R}_1, \tilde{R}_2) \in \mathsf{Q}_{\mathrm{inv}}(\boldsymbol{V}, \epsilon)} \tilde{R}_1 + \tilde{R}_2 \tag{21}$$

$$\boldsymbol{V} \triangleq \mathrm{Cov} \begin{bmatrix} -i(U, S) \\ i(U, Y) \end{bmatrix}. \tag{22}$$

For the case that an input constraint is present (e.g. dirty paper coding), a similar expansion was provided in [12] using a 3×3 covariance matrix $\boldsymbol{V}$, with the third entry added to capture the probability that the random i.i.d. codeword violates the constraint.

A study of the second-order asymptotics of the modulo-lattice additive noise channel was provided by Jiang and Liu [10]. By a data-processing argument, their result provides an achievable second-order expansion of the rate for dirty paper coding with common randomness at the encoder and decoder, and with a power constraint which is averaged over the common randomness. In particular, [10, Thm. 1] bears a strong resemblance to Theorem 2 below. However, it should be noted that our setting assumes a stricter power constraint without common randomness, as discussed in Section I-B. Furthermore, the analysis in [10] is significantly different from ours.

For related work on random-coding error exponents, see [9], [14], [15] and references therein.

D. Contributions

As stated previously, the main contributions of this paper are a second-order achievability result for the discrete case, and a conclusive characterization of the second-order asymptotics for the Gaussian case. In the discrete case with a target error probability less than $\frac{1}{2}$, we show that our result can be weakened to (20). For the Gaussian case, we show that the dispersion is the same as that of the AWGN channel of the same input power without a state.

Our result for the discrete case is based on constant-composition random coding, which has recently been shown to yield gains in the second-order performance of other network information theory problems [16], [17]. In the Gaussian case, we use a variant of random coding according to a uniform distribution on a shell, which has been used for the single-user Gaussian channel [2] and the Gaussian multiple-access channel [18]. In both cases, we reduce the problem to that of a genie-aided setting by using a small fraction of the block length to inform the decoder of a property of the state sequence, namely, its empirical distribution or its quantized power (e.g. see [14]). A key part of our analysis for the discrete case makes use of techniques recently introduced by Tomamichel and Tan [19, Lemmas 17-18].

E. Notation

Bold symbols are used for vectors (e.g. $\boldsymbol{x}$), and the corresponding i-th entry is written using a subscript (e.g. x_i). Given two vectors, say $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, we define the inner product $\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle = \sum_i x_{1,i} x_{2,i}$, the ℓ_2-norm $\|\boldsymbol{x}_1\| = \sqrt{\langle \boldsymbol{x}_1, \boldsymbol{x}_1 \rangle}$, and the ℓ_∞-norm $\|\boldsymbol{x}_1\|_\infty = \max_i |x_{1,i}|$.

The marginals of a joint distribution P_{XY} are denoted by P_X and P_Y. Probability, expectation and variance are respectively denoted by $\mathbb{P}[\cdot]$, $\mathbb{E}[\cdot]$, and $\mathrm{Var}[\cdot]$. When the meaning is clear, we will use shorthands such as $\mathbb{P}[\cdot \,|\, x]$ to denote conditioning on events such as $X = x$. For two sequences f_n and g_n, we write $f_n = O(g_n)$ if $|f_n| \le c|g_n|$ for some c and sufficiently large n, $f_n = o(g_n)$ if $\lim_{n\to\infty} \frac{f_n}{g_n} = 0$, and $f_n = \Theta(g_n)$ if both $f_n = O(g_n)$ and $g_n = O(f_n)$ hold.

II. Statement of Main Results

In this section, we present a formal statement of our main results, along with some discussions and comparisons to existing results.

Theorem 1. *Consider a discrete-memoryless Gel'fand-Pinsker channel described by $\pi(s)$ and $W(y|x,s)$. Let $\mathcal{U}$, $Q_{U|S}$ and $\phi(u,s)$ by any set of capacity-achieving parameters in* (3)*, and let P_{SUY}, $i(u,s)$ and $i(u,y)$ be as given in* (4)*,* (17) *and* (18) *under these parameters. If $\mathbb{E}\big[\mathrm{Var}[i(U,Y) \,|\, S, U]\big] > 0$, then*

$$\log M^*(n,\epsilon) \ge nC - \sqrt{nV}\mathsf{Q}^{-1}(\epsilon) + O(\log n), \tag{23}$$

for $\epsilon \in (0,1)$, where

$$V \triangleq \mathbb{E}\big[\mathrm{Var}[i(U,Y) \,|\, S, U]\big] + \mathrm{Var}\big[\mathbb{E}[i(U,Y) - i(U,S) \,|\, S]\big] \tag{24}$$

$$= \mathrm{Var}\big[i(U,Y) - i(U,S)\big]. \tag{25}$$

Proof: See Section III. ■

In the proof of Theorem 1, we first prove (23) with V of the form given in (24), and then show that (24) and (25) coincide under any input distribution $Q_{U|S}$ which maximizes $I(U;Y) - I(U;S)$ in (3). We will see in Section III-B8 that (25) exceeds (24) more generally; the inequality may be strict if $Q_{U|S}$ is suboptimal. This is analogous to the single-user setting, where constant-composition random coding can be used to prove the achievability of (1) with $V = \mathbb{E}[\mathrm{Var}[i(X,Y) \,|\, X]]$. This is upper bounded by $\mathrm{Var}[i(X,Y)]$, but equality holds under any capacity-achieving input distribution [2].

In Section V, we discuss the difficulties in removing the assumption $\mathbb{E}\big[\mathrm{Var}[i(U,Y) \,|\, S, U]\big] > 0$ in Theorem 1.

Theorem 2. *Consider the dirty paper coding setup with input power $P > 0$. For any sequence of state distributions $P_{\boldsymbol{S}}$ (indexed by n) such that*

$$\mathbb{P}\big[\|\boldsymbol{S}\|^2 > n\Pi\big] = O\Big(\frac{\log n}{\sqrt{n}}\Big) \tag{26}$$

for some $\Pi < \infty$, we have

$$\log M^*(n,\epsilon) = nC - \sqrt{nV}\mathsf{Q}^{-1}(\epsilon) + O(\log n), \tag{27}$$

for $\epsilon \in (0,1)$*, where*

$$V \triangleq \frac{P(2+P)}{2(1+P)^2}. \tag{28}$$

Proof: The converse part follows by revealing the state sequence to the decoder and using the converse result for the AWGN channel without state [2]. The achievability part is proved in Section IV. ∎

The assumption in (26) is mild, allowing for any state sequence distribution yielding a uniformly bounded (yet arbitrarily large) power with probability $1 - O\big(\frac{\log n}{\sqrt{n}}\big)$. In particular, the state sequence need not be i.i.d. nor even ergodic, and may be deterministic. Furthermore, the right-hand of (26) side can be weakened to $o(1)$ at the expense of weakening the $O(\log n)$ term in (27) to $o(\sqrt{n})$.

In the case of an i.i.d. state with $S_i \sim \pi$, Chebyshev's inequality reveals that a sufficient condition for (26) to hold is that $\mathbb{E}_\pi[S^4] < \infty$. In the special case that $\pi \sim N(0, P_\pi)$ for some $P_\pi > 0$, substituting the capacity-achieving parameters (see (8)–(9) and (14)) into (24) yields precisely the dispersion in (28), thus establishing a connection with the discrete case. More precisely, the first term in (24) equals $\frac{P(2+P)}{2(1+P)^2}$, and the second term is zero.

A. Comparisons to Existing Results

We begin by showing that, for any $\epsilon \in (0, \frac{1}{2})$, Theorem 1 yields a second-order term which is no worse than that of (20), i.e. $\sqrt{V}\mathsf{Q}^{-1}(\epsilon) \le \tilde{R}$. We claim that

$$\tilde{R} \ge \big(\sqrt{\mathrm{Var}[i(U,S)]} + \sqrt{\mathrm{Var}[i(U,Y)]}\big)\mathsf{Q}^{-1}(\epsilon). \tag{29}$$

To see this, we note from (19) that any $(\tilde{R}_1, \tilde{R}_2) \in \mathsf{Q}_{\mathrm{inv}}(\boldsymbol{V}, \epsilon)$ must satisfy $\tilde{R}_1 \ge \sqrt{V_1}\mathsf{Q}^{-1}(\epsilon)$ and $\tilde{R}_2 \ge \sqrt{V_2}\mathsf{Q}^{-1}(\epsilon)$, where V_1 and V_2 are the diagonal entries of $\boldsymbol{V}$. Furthermore, we can expand (25) as

$$V = \mathrm{Var}[i(U,S)] - 2\mathrm{Cov}[i(U,S), i(U,Y)] + \mathrm{Var}[i(U,Y)] \tag{30}$$

$$\le \mathrm{Var}[i(U,S)] + 2\sqrt{\mathrm{Var}[i(U,S)]\mathrm{Var}[i(U,Y)]} + \mathrm{Var}[i(U,Y)], \tag{31}$$

where (31) follows from the Cauchy-Schwarz inequality. The desired result follows from the identity $\sqrt{V_1} + \sqrt{V_2} = \sqrt{V_1 + 2\sqrt{V_1 V_2} + V_2}$, which is easily verified by squaring both sides.

Next, we present a numerical example showing that the improvement over (20) can be strict. We revisit the example of memory with stuck-at faults given in [12]. The alphabets are given by $\mathcal{S} = \{0,1,2\}$ and $\mathcal{X} = \mathcal{Y} = \{0,1\}$, and we assume that $\pi = \big(\frac{p}{2}, \frac{p}{2}, 1-p\big)$ for some constant p. The channel is described as follows: $Y = 0$ (respectively, $Y = 1$) deterministically whenever $S = 0$ (respectively, $S = 1$), and the remaining transition probabilities $W(\cdot|\cdot, 2)$ are those of a binary symmetric channel (BSC) with crossover probability δ. The capacity is given by

$$C = (1-p)(1 - H_2(\delta)) \quad \text{bits/use},$$

and is achieved by the parameters $\mathcal{U} = \{0,1\}$, $Q_{U|S=0} = (1-\delta, \delta)$, $Q_{U|S=1} = (\delta, 1-\delta)$, $Q_{U|S=2} = \big(\frac{1}{2}, \frac{1}{2}\big)$ and $\phi(u,s) = u$. Choosing $\delta = 0.11$, $p = 0.1$ and $\epsilon = 0.001$, we computed the coefficient to $\sqrt{n}$ in (20) to be $\tilde{R} = 4.16$ bits/$\sqrt{\text{use}}$, whereas the coefficient in (23) is $\sqrt{V}\mathsf{Q}^{-1}(\epsilon) = 2.81$ bits/$\sqrt{\text{use}}$.

The following corollary shows that Theorem 1 recovers the achievability part of the dispersion for discrete memoryless channels with state known non-causally at both the encoder and decoder [19]. That is, in this special case there is a matching converse to Theorem 1.

Corollary 1. *Consider a discrete memoryless state-dependent channel $W(y|x,s)$ with state distribution $\pi(s)$, where the state sequence $\boldsymbol{S}$ is known non-causally at both the encoder and decoder. For each $s \in \mathcal{S}$, let C_s and V_s denote the capacity and dispersion of the channel $W(\cdot|\cdot,s)$. If $\mathbb{E}[V_S] > 0$, then*

$$\log M^*(n,\epsilon) \geq nC - \sqrt{nV}\mathsf{Q}^{-1}(\epsilon) + O(\log n), \tag{32}$$

where

$$C = \mathbb{E}[C_S] \tag{33}$$

$$V = \mathbb{E}[V_S] + \mathrm{Var}[C_S]. \tag{34}$$

Proof: See Appendix A. ∎

III. Proof for the Discrete Memoryless Gel'fand Pinsker Channel

We present a number of preliminary results in Section III-A, and we prove Theorem 1 in Section III-B. We make use of the method of types for finite alphabets [20, Ch. 2]. The empirical distribution (i.e. type) of a vector $\boldsymbol{x}$ is denoted by $\hat{P}_{\boldsymbol{x}}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{x_i = x\}$. The set of all types of length n on an alphabet $\mathcal{X}$ is denoted by $\mathcal{P}_n(\mathcal{X})$. The set of all sequences of length n with a given type P_X is denoted by $T^n(P_X)$, and similarly for joint types.

A. Preliminary Results

Throughout this subsection, we let $\mathcal{U}$, $Q_{U|S}$ and $\phi(\cdot,\cdot)$ be arbitrary.

1) A Genie-Aided Setting : We will prove Theorem 1 by first proving the following result for a genie-aided setting.

Theorem 3. *The statement of Theorem 1 holds true in the case that the empirical distribution $\hat{P}_{\boldsymbol{S}}$ of $\boldsymbol{S}$ is known at the decoder.*

We proceed by showing that Theorem 3 implies Theorem 1. The idea is to use the first $O(\log n)$ symbols to transmit the type $\hat{P}_{\boldsymbol{S}}$, and then use a second-order optimal code with state type known at the decoder for the remaining symbols. This technique was proposed in [14] in the context of random-coding error exponents.

We fix a sequence $g(n)$, define $\tilde{n} \triangleq n - g(n)$, and let $P_S \in \mathcal{P}_{\tilde{n}}(\mathcal{S})$ be the type of the last $\tilde{n}$ symbols of $\boldsymbol{S}$. The number of such types is upper bounded by $\tilde{M} \triangleq (n+1)^{|\mathcal{S}|-1}$. Using Gallager's random-coding bound [21, Sec. 5.6], we can transmit the type in $g(n)$ symbols with an error probability $\overline{p}_{e,0}$ satisfying

$$\overline{p}_{e,0} \leq e^{-g(n)\tilde{E}_r(\tilde{R})}, \tag{35}$$

where $\tilde{E}_r(\cdot)$ is the random-coding error exponent of the channel $\tilde{W}(y|x) \triangleq \sum_s \pi(s)W(y|x,s)$, and $\tilde{R} \triangleq \frac{1}{g(n)} \log \tilde{M} = \frac{|\mathcal{S}|-1}{g(n)} \log(n+1)$.[1] We choose $g(n) = K_0 \log(n+1)$, where K_0 is chosen to be sufficiently large so that $K_0 \tilde{E}_r(\tilde{R}) \geq 1$ (note that $\tilde{R} = \frac{|\mathcal{S}|-1}{K_0}$, so this is always possible). It follows from (35) and the choices of $\tilde{M}$ and $g(n)$ that

$$\overline{p}_{e,0} \leq e^{-K_0 \tilde{E}_r(\tilde{R}) \log(n+1)} \tag{36}$$

$$\leq e^{-\log(n+1)} \tag{37}$$

$$= \frac{1}{n+1}. \tag{38}$$

Thus, if $\big(n - K_0 \log(n+1), M, \epsilon - \frac{1}{n+1}\big)$ is achievable in the genie-aided setting, then (n, M, ϵ) is achievable in the absence of the genie. By performing a Taylor expansion of the square root and $\mathsf{Q}^{-1}(\cdot)$ function in (23), we conclude that Theorem 3 implies Theorem 1.

2) A Typical Set of State Types: In Section III, we will study an encoder and decoder which use different codebooks depending on the type P_S of the state sequence $\boldsymbol{S}$. Here we introduce a typical set of state types, defined by

$$\tilde{\mathcal{P}}_n \triangleq \left\{ P_S \in \mathcal{P}_n(\mathcal{S}) \,:\, \|P_S - \pi\|_\infty \leq \sqrt{\frac{\log n}{n}} \right\}. \tag{39}$$

We will see the second-order performance is unaffected by types falling outside $\tilde{\mathcal{P}}_n$, due to the fact that [19, Lemma 22]

$$\mathbb{P}\big[\hat{P}_{\boldsymbol{S}} \notin \tilde{\mathcal{P}}_n\big] = O\Big(\frac{1}{n^2}\Big). \tag{40}$$

It is clear from (39) that $f(P_S) \to f(\pi)$ for $P_S \in \tilde{\mathcal{P}}_n$ and any function $f(\cdot)$ which is continuous at π, and that $|f(P_S) - f(\pi)| = O\big(\sqrt{\frac{\log n}{n}}\big)$ for $P_S \in \tilde{\mathcal{P}}_n$ and any function $f(\cdot)$ which is continuously differentiable at π.

3) Type-Dependent Distributions : Here we present results regarding the approximation of a distribution by a type. For each $P_S \in \mathcal{P}_n(\mathcal{S})$, we define an approximation $Q_{U|S,n}^{(P_S)}$ of $Q_{U|S}$ as follows: For each $s \in \mathcal{S}$ with $P_S(s) > 0$, let $Q_{U|S,n}^{(P_S)}(\cdot|s)$ be a type in $\mathcal{P}_{nP_S(s)}(\mathcal{U})$ whose probabilities are $\frac{1}{nP_S(s)}$-close to $Q_{U|S}(\cdot|s)$ in terms of L_∞ norm. If $P_S(s) = 0$ then $Q_{U|S,n}^{(P_S)}(\cdot|s)$ is arbitrary (e.g. uniform).

Assuming without loss of generality that $\pi(s) > 0$ for all $s \in \mathcal{S}$, we have from (39) that $\min_s nP_S(s)$ grows linearly in n for all $P_S \in \tilde{\mathcal{P}}_n$. It follows from the above construction of $Q_{U|S,n}^{(P_S)}$ that

$$\left| Q_{U|S}(u|s) - Q_{U|S,n}^{(P_S)}(u|s) \right| = O\Big(\frac{1}{n}\Big) \tag{41}$$

uniformly in $P_S \in \tilde{\mathcal{P}}_n$ and (s,u).

Throughout Section III-B, we will make use of the following joint distributions:

$$P_{SUY}^{(P_S)}(s,u,y) \triangleq P_S(s)Q_{U|S}(u|s)W(y|\phi(u,s),s) \tag{42}$$

$$P_{SUY,n}^{(P_S)}(s,u,y) \triangleq P_S(s)Q_{U|S,n}^{(P_S)}(u|s)W(y|\phi(u,s),s). \tag{43}$$

[1] We could instead use the potentially stronger error exponents of [14], [15], but any exponent which is positive for sufficiently small positive rates suffices for our purposes.

Using (41), we immediately obtain that

$$\left|P^{(P_S)}_{SUY}(s,u,y) - P^{(P_S)}_{SUY,n}(s,u,y)\right| = O\Big(\frac{1}{n}\Big) \tag{44}$$

uniformly in $P_S \in \tilde{\mathcal{P}}_n$ and (s,u,y). We will use this result to approximate various expectations $\mathbb{E}_{P^{(P_S)}_{SUY,n}}[\cdot]$ by $\mathbb{E}_{P^{(P_S)}_{SUY}}[\cdot]$.

4) A Taylor Expansion of the Mutual Information: Let $I^{(P_S)}(U;S)$ and $I^{(P_S)}(U;Y)$ denote mutual informations with respect to the joint distribution $P^{(P_S)}_{USY}$ in (42), and define

$$I(P_S) \triangleq I^{(P_S)}(U;Y) - I^{(P_S)}(U;S). \tag{45}$$

We observe from (3) that $C = I(\pi)$ whenever the parameters $\mathcal{U}$, $Q_{U|S}$ and $\phi(\cdot,\cdot)$ achieve capacity.

In Section III-B7, we will make use of a linear approximation of $I(\cdot)$ given by

$$\tilde{I}(P_S) \triangleq \sum_s P_S(s)\mathbb{E}[i(U,Y) - i(U,S)\,|\,S=s] \tag{46}$$

$$= \sum_s P_S(s)\left(\sum_{u,y} Q_{U|S}(u|s)W(y|\phi(u,s),s)\log\frac{P^{(\pi)}_{Y|U}(y|u)}{P^{(\pi)}_Y(y)} - \sum_u Q_{U|S}(u|s)\log\frac{Q_{U|S}(u|s)}{P^{(\pi)}_U(u)}\right), \tag{47}$$

which equals the first-order Taylor approximation of $I(P_S)$ about $P_S = \pi$. More precisely, we show in Appendix B that

$$I(P_S) = \tilde{I}(P_S) + \Delta(P_S), \tag{48}$$

where

$$\max_{P_S \in \tilde{\mathcal{P}}_n} |\Delta(P_S)| \le \frac{K_1 \log n}{n} \tag{49}$$

for some constant K_1 and sufficiently large n.

B. *Proof of Theorem 1*

As stated previously, in order to prove Theorem 1, it suffices to prove Theorem 3. Thus, we henceforth assume that the state type P_S is known at the decoder.

1) Random-Coding Parameters: We consider a random-coding ensemble which is similar to that of [6, Sec. 7.6], the main difference being that we generate a different auxiliary codebook for each state type. The parameters are the auxiliary alphabet $\mathcal{U}$, input distribution $Q_{U|S}$, function $\phi : \mathcal{U} \times \mathcal{S} \to \mathcal{X}$, and number of auxiliary codewords $L^{(P_S)}$ per message for each state type $P_S \in \mathcal{P}_n(\mathcal{S})$. In accordance with the statement of Theorem 1, we assume that $\mathcal{U}$, $Q_{U|S}$ and ϕ are capacity-achieving.

2) Codebook Generation: For each state type $P_S \in \mathcal{P}_n(\mathcal{S})$ and each message $m = 1, \cdots, M$, we randomly generate an auxiliary codebook $\mathcal{C}^{(P_S)}_U$ containing $ML^{(P_S)}$ auxiliary codewords $\{\boldsymbol{U}^{(P_S)}(m,l)\}_{l=1}^{L^{(P_S)}}$, each of which is independently distributed according to the uniform distribution on the type class $T^n(P^{(P_S)}_{U,n})$ (see (43)):

$$P^{(P_S)}_{\boldsymbol{U}}(\boldsymbol{u}) = \frac{1}{|T^n(P^{(P_S)}_{U,n})|}\mathbb{1}\Big\{\boldsymbol{u} \in T^n(P^{(P_S)}_{U,n})\Big\}. \tag{50}$$

Each auxiliary codebook is revealed to the encoder and decoder.

3) Encoding and Decoding: Given the state sequence $\boldsymbol{S} \in T^n(P_S)$ and message m, the encoder sends

$$\phi^n(\boldsymbol{U}, \boldsymbol{S}) \triangleq \big(\phi(U_1, S_1), \cdots, \phi(U_n, S_n)\big), \tag{51}$$

where $\boldsymbol{U}$ is an auxiliary codeword $\boldsymbol{U}^{(P_S)}(m,l)$ in $\mathcal{C}_U^{(P_S)}$, with l chosen such that $(\boldsymbol{S}, \boldsymbol{U}) \in T^n(P_{SU,n}^{(P_S)})$. If multiple such auxiliary codewords exist, one of them is chosen arbitrarily. An error is declared if no such auxiliary codeword exists. Given the received vector $\boldsymbol{y}$ and the state type P_S, the decoder estimates m according to the pair $(\tilde{m}, \tilde{l})$ whose corresponding sequence $\boldsymbol{U}^{(P_S)}(\tilde{m}, \tilde{l})$ maximizes

$$i_n^{(P_S)}(\boldsymbol{u}, \boldsymbol{y}) \triangleq \sum_{i=1}^{n} i^{(P_S)}(u_i, y_i) \tag{52}$$

among the auxiliary codewords in $\mathcal{C}_U^{(P_S)}$, where

$$i^{(P_S)}(u, y) \triangleq \log \frac{P_{Y|U}^{(P_S)}(y|u)}{P_Y^{(P_S)}(y)} \tag{53}$$

with $P_{SUY}^{(P_S)}$ defined in (42). Ties are broken in an arbitrary fashion. Note that $P_{SUY}^{(\pi)}$ coincides with the joint distribution in (4), and hence $i^{(\pi)}(u, y)$ coincides with (18).

We consider the events

$$\mathcal{E}_1 \triangleq \left\{ \text{No } l \text{ exists with } (\boldsymbol{S}, \boldsymbol{U}^{(P_S)}(m,l)) \in T^n(P_{SU,n}^{(P_S)}) \right\} \tag{54}$$

$$\mathcal{E}_2 \triangleq \left\{ \text{Decoder chooses a message } \tilde{m} \neq m \right\}. \tag{55}$$

It follows from these definitions and (40) that the overall random-coding error probability $\overline{p}_e$ satisfies

$$\overline{p}_e \leq \sum_{P_S \in \tilde{\mathcal{P}}_n} \mathbb{P}\big[\hat{P}_{\boldsymbol{S}} = P_S\big] \Big(\mathbb{P}\big[\mathcal{E}_1 \,|\, P_S\big] + \mathbb{P}\big[\mathcal{E}_2 \,|\, P_S, \mathcal{E}_1^c\big] \Big) + O\Big(\frac{1}{n^2}\Big). \tag{56}$$

4) Analysis of $\mathcal{E}_1$: We study the probability of $\mathcal{E}_1$ conditioned on $\boldsymbol{S}$ having a given type $P_S \in \tilde{\mathcal{P}}_n$. Using the property of types in [22, Eq. (18)], we have for any $\boldsymbol{s} \in T^n(P_S)$ and $\boldsymbol{U}$ distributed according to (50) that

$$\mathbb{P}\big[(\boldsymbol{s}, \boldsymbol{U}) \in T^n(P_{SU,n}^{(P_S)})\big] \geq \frac{1}{p_0(n)} e^{-nI^{(P_S)}(U;S)}, \tag{57}$$

where $I^{(P_S)}(U;S)$ is defined in Section III-A4, and $p_0(n)$ is a polynomial. Since the codewords are generated independently, it follows that

$$\mathbb{P}\big[\mathcal{E}_1 \,|\, P_S\big] \leq \left(1 - \frac{1}{p_0(n)} e^{-nI^{(P_S)}(U;S)}\right)^{L^{(P_S)}} \tag{58}$$

$$\leq \left(\exp\left(- \frac{1}{p_0(n)} e^{-nI^{(P_S)}(U;S)} \right) \right)^{L^{(P_S)}} \tag{59}$$

$$= \exp\left(- \frac{1}{p_0(n)} e^{-n\left(I^{(P_S)}(U;S) - R_L^{(P_S)}\right)} \right), \tag{60}$$

where (59) follows since $1 - \alpha \leq e^{-\alpha}$, and in (60) we define

$$R_L^{(P_S)} \triangleq \frac{1}{n} \log L^{(P_S)}.$$

Choosing

$$R_L^{(P_S)} = I^{(P_S)}(U;S) + K_2 \frac{\log n}{n} \tag{61}$$

with K_2 equal to one plus the degree of the polynomial $p_0(n)$, we obtain from (60) that

$$\mathbb{P}\big[\mathcal{E}_1 \,|\, P_S\big] \le e^{-\psi n} \tag{62}$$

for some $\psi > 0$ and sufficiently large n.

5) Analysis of $\mathcal{E}_2$: We study the probability of $\mathcal{E}_2$ conditioned on $\boldsymbol{S}$ having a given type $P_S \in \tilde{\mathcal{P}}_n$, and also conditioned on $\mathcal{E}_1^c$. These conditions imply that $(\boldsymbol{s},\boldsymbol{u}) \in T^n(P_{SU,n}^{(P_S)})$ for all $(\boldsymbol{s},\boldsymbol{u})$ occurring with non-zero probability, and by the symmetry of the state sequence distribution and the codebook construction, all such $(\boldsymbol{s},\boldsymbol{u})$ are equally likely. It follows that the conditional distribution given P_S and $\mathcal{E}_1^c$ of the state sequence $\boldsymbol{S}$, auxiliary codeword $\boldsymbol{U}$, and received sequence $\boldsymbol{Y}$ is given by

$$(\boldsymbol{S},\boldsymbol{U},\boldsymbol{Y}) \sim P_{\boldsymbol{SU}}^{(P_S)}(\boldsymbol{s},\boldsymbol{u}) W^n(\boldsymbol{y}|\phi^n(\boldsymbol{u},\boldsymbol{s}),\boldsymbol{s}), \tag{63}$$

where

$$P_{\boldsymbol{SU}}^{(P_S)}(\boldsymbol{s},\boldsymbol{u}) \triangleq \frac{1}{\big|T^n(P_{SU,n}^{(P_S)})\big|} \mathbb{1}\Big\{(\boldsymbol{s},\boldsymbol{u}) \in T^n(P_{SU,n}^{(P_S)})\Big\}, \tag{64}$$

i.e. the uniform distribution on the type class $T^n(P_{SU,n}^{(P_S)})$.

Recall that the decoder maximizes $i_n^{(P_S)}(\boldsymbol{u},\boldsymbol{y})$ (see (52)) among the $ML^{(P_S)}$ auxiliary codewords in $\mathcal{C}_U^{(P_S)}$. Using the threshold bound for mismatched decoding [23], we have for any $\gamma^{(P_S)}$ that

$$\mathbb{P}\big[\mathcal{E}_2 \,|\, P_S, \mathcal{E}_1^c\big] \le \mathbb{P}\Big[i_n^{(P_S)}(\boldsymbol{U},\boldsymbol{Y}) \le \gamma^{(P_S)}\Big] + ML^{(P_S)}\mathbb{P}\Big[i_n^{(P_S)}(\overline{\boldsymbol{U}},\boldsymbol{Y}) > \gamma^{(P_S)}\Big], \tag{65}$$

where $\overline{\boldsymbol{U}} \sim P_{\boldsymbol{U}}^{(P_S)}$ independently of $(\boldsymbol{S},\boldsymbol{U},\boldsymbol{Y})$. In order to upper bound the second probability, it will prove useful to upper bound the output distribution $P_{\boldsymbol{Y}}^{(P_S)}(\boldsymbol{y}) \triangleq \sum_{\boldsymbol{s},\boldsymbol{u}} P_{\boldsymbol{SU}}^{(P_S)}(\boldsymbol{s},\boldsymbol{u}) W^n(\boldsymbol{y}|\phi^n(\boldsymbol{u},\boldsymbol{s}),\boldsymbol{s})$ as follows:

$$P_{\boldsymbol{Y}}^{(P_S)}(\boldsymbol{y}) \le p_1(n) \sum_{\boldsymbol{s},\boldsymbol{u}} \prod_{i=1}^n P_S(s_i) Q_{U|S,n}^{(P_S)}(u_i|s_i) W(y_i|\phi^n(u_i,s_i),s_i) \tag{66}$$

$$= p_1(n) \prod_{i=1}^n \bigg(\sum_{s,u} P_S(s) Q_{U|S,n}^{(P_S)}(u|s) W(y_i|\phi(u,s),s) \bigg) \tag{67}$$

$$= p_1(n) \prod_{i=1}^n \bigg(P_Y^{(P_S)}(y_i) \Big(1 + O\Big(\frac{1}{n}\Big)\Big)\bigg) \tag{68}$$

$$\le p_2(n) \prod_{i=1}^n P_Y^{(P_S)}(y_i), \tag{69}$$

where (66) holds for some polynomial $p_1(n)$ by a standard change of measure from uniform on the type class to i.i.d. [24, Eq. (2.4)], (68) follows from (44), and (69) follows for some polynomial $p_2(n)$ and sufficiently large n

since $(1+\frac{c}{n})^n \to e^c$, which is a constant. Using the definition of $i_n^{(P_S)}$ in (52), it follows that

$$\mathbb{P}\Big[i_n^{(P_S)}(\overline{\boldsymbol{U}},\boldsymbol{Y}) > \gamma^{(P_S)}\Big] = \sum_{\overline{\boldsymbol{u}},\boldsymbol{y}} P_{\boldsymbol{U}}^{(P_S)}(\overline{\boldsymbol{u}}) P_{\boldsymbol{Y}}^{(P_S)}(\boldsymbol{y}) \mathbb{1}\bigg\{\prod_{i=1}^n \frac{P_{Y|U}^{(P_S)}(y_i|\overline{u}_i)}{P_Y^{(P_S)}(y_i)} > e^{\gamma^{(P_S)}}\bigg\} \tag{70}$$

$$\leq \sum_{\overline{\boldsymbol{u}},\boldsymbol{y}} P_{\boldsymbol{U}}^{(P_S)}(\overline{\boldsymbol{u}}) P_{\boldsymbol{Y}}^{(P_S)}(\boldsymbol{y}) \prod_{i=1}^n \frac{P_{Y|U}^{(P_S)}(y_i|\overline{u}_i)}{P_Y^{(P_S)}(y_i)} e^{-\gamma^{(P_S)}} \tag{71}$$

$$\leq p_2(n) \sum_{\overline{\boldsymbol{u}},\boldsymbol{y}} P_{\boldsymbol{U}}^{(P_S)}(\overline{\boldsymbol{u}}) \prod_{i=1}^n P_{Y|U}^{(P_S)}(y_i|\overline{u}_i) e^{-\gamma^{(P_S)}} \tag{72}$$

$$= p_2(n) e^{-\gamma^{(P_S)}}. \tag{73}$$

We fix a constant K_3 and choose

$$\gamma^{(P_S)} = \log ML^{(P_S)} + K_3 \log n \tag{74}$$

which, when combined with (61), yields

$$\gamma^{(P_S)} = \log M + nI^{(P_S)}(U;S) + K_4 \log n, \tag{75}$$

where $K_4 \triangleq K_2 + K_3$. Setting K_3 to be one higher than the degree of $p_2(n)$, we obtain from (65), (73) and (74) that

$$ML^{(P_S)} \mathbb{P}\Big[i_n^{(P_S)}(\overline{\boldsymbol{U}},\boldsymbol{Y}) > \gamma^{(P_S)}\Big] = O\Big(\frac{1}{n}\Big), \tag{76}$$

and hence

$$\mathbb{P}\big[\mathcal{E}_2 \,|\, P_S, \mathcal{E}_1^c\big] \leq \mathbb{P}\Big[i_n^{(P_S)}(\boldsymbol{U},\boldsymbol{Y}) \leq \log M + nI^{(P_S)}(U;S) + K_4 \log n\Big] + O\Big(\frac{1}{n}\Big), \tag{77}$$

where the remainder term is uniform in $P_S \in \tilde{\mathcal{P}}_n$.

6) Application of the Berry-Esseen Theorem: Combining (62) and (77), we have for all $P_S \in \tilde{\mathcal{P}}_n$ that

$$\mathbb{P}\big[\mathcal{E}_1 \cup \mathcal{E}_2 \,|\, P_S\big] \leq \mathbb{P}\Big[i_n^{(P_S)}(\boldsymbol{U},\boldsymbol{Y}) \leq \log M + nI^{(P_S)}(U;S) + K_4 \log n\Big] + O\Big(\frac{1}{n}\Big). \tag{78}$$

In order to apply the Berry-Esseen theorem [25, Sec. XVI.5] to the right-hand side of (78), we first compute the mean and variance of $i_n^{(P_S)}(\boldsymbol{U},\boldsymbol{Y})$, defined according to (52) and (63). The relevant third moment can easily be uniformly bounded in terms of the alphabet sizes [2, Lemma 46], [26, Appendix D]. We will use the fact that, by the symmetry of the constant-composition distribution in (64), the statistics of $i_n^{(P_S)}(\boldsymbol{U},\boldsymbol{Y})$ are unchanged upon conditioning on $(\boldsymbol{S},\boldsymbol{U}) = (\boldsymbol{s},\boldsymbol{u})$ for some $(\boldsymbol{s},\boldsymbol{u}) \in T^n(P_{SU,n}^{(P_S)})$. Using the joint distribution $P_{SUY,n}^{(P_S)}$ defined in (43), we have

$$\mathbb{E}\big[i_n^{(P_S)}(\boldsymbol{u},\boldsymbol{Y}) \,|\, \boldsymbol{s},\boldsymbol{u}\big] = \mathbb{E}\bigg[\sum_{i=1}^n i^{(P_S)}(u_i,Y_i) \,\bigg|\, s_i,u_i\bigg] \tag{79}$$

$$= n \sum_{u,y} P_{UY,n}^{(P_S)}(u,y) i^{(P_S)}(u,y) \tag{80}$$

$$= nI^{(P_S)}(U;Y) + O(1), \tag{81}$$

where (81) follows from (44) and the definitions of $i^{(P_S)}(u,y)$ and $I^{(P_S)}(U;Y)$ (see (53) and Section III-A4). Similarly, we have

$$\mathrm{Var}\big[i_n^{(P_S)}(\boldsymbol{u},\boldsymbol{Y})\,|\,\boldsymbol{s},\boldsymbol{u}\big] = \mathrm{Var}\bigg[\sum_{i=1}^{n} i^{(P_S)}(u_i,Y_i)\,\bigg|\,s_i,u_i\bigg] \tag{82}$$

$$= n\sum_{s,u} P_{SU,n}^{(P_S)}(s,u)\mathrm{Var}\big[i^{(P_S)}(u,Y)\,|\,s,u\big] \tag{83}$$

$$= n\mathbb{E}\Big[\mathrm{Var}\big[i^{(P_S)}(U,Y)\,|\,S,U\big]\Big] + O(1) \tag{84}$$

$$\triangleq nV(P_S) + O(1). \tag{85}$$

It should be noted that $V(P_S)$ is bounded away for zero for $P_S \in \tilde{\mathcal{P}}_n$ and sufficiently large n, since $V(\pi) > 0$ by assumption in Theorem 1. Furthermore, the $O(1)$ terms in (81) and (85) are uniform in $P_S \in \tilde{\mathcal{P}}_n$, due to the uniformity of (44).

Using the definition of $I(P_S)$ in (45), we choose

$$\log M = nI(\pi) - K_4 \log n - \beta_n, \tag{86}$$

where β_n will be specified later, and will behave as $O(\sqrt{n})$. Combining (78), (81), (85) and (86), we have

$$\mathbb{P}\big[\mathcal{E}_1 \cup \mathcal{E}_2\,|\,P_S\big] \le \mathbb{P}\Big[i_n^{(P_S)}(\boldsymbol{U},\boldsymbol{Y}) \le nI(\pi) + nI^{(P_S)}(U;S) - \beta_n\Big] + O\Big(\frac{1}{n}\Big). \tag{87}$$

$$= \mathbb{P}\Big[i_n^{(P_S)}(\boldsymbol{u},\boldsymbol{Y}) \le nI(\pi) + nI^{(P_S)}(U;S) - \beta_n\,\big|\,\boldsymbol{s},\boldsymbol{u}\Big] + O\Big(\frac{1}{n}\Big) \tag{88}$$

$$\le \mathsf{Q}\bigg(\frac{\beta_n + nI(P_S) - nI(\pi) + K_5}{\sqrt{nV(P_S) + K_6}}\bigg) + O\bigg(\frac{1}{\sqrt{n}}\bigg) \tag{89}$$

where (88) holds for any $(\boldsymbol{s},\boldsymbol{u}) \in T^n(P_{SU,n}^{(P_S)})$ by symmetry, and (89) follows from the Berry-Esseen theorem for independent and non-identically distributed variables [25, Sec. XVI.5], and by introducing the constants K_5 and K_6 to represent the uniform $O(1)$ terms in (81) and (85).

7) Averaging Over the State Type: Substituting (89) into (56), we have

$$\overline{p}_e \le \sum_{P_S \in \tilde{\mathcal{P}}_n} \mathbb{P}\big[\hat{P}_{\boldsymbol{S}} = P_S\big]\mathsf{Q}\bigg(\frac{\beta_n + nI(P_S) - nI(\pi) + K_5}{\sqrt{nV(P_S) + K_6}}\bigg) + O\Big(\frac{1}{\sqrt{n}}\Big) \tag{90}$$

$$\le \sum_{P_S \in \tilde{\mathcal{P}}_n} \mathbb{P}\big[\hat{P}_{\boldsymbol{S}} = P_S\big]\mathsf{Q}\bigg(\frac{\beta_n + nI(P_S) - nI(\pi)}{\sqrt{nV(P_S)}}\bigg) + O\Big(\frac{1}{\sqrt{n}}\Big), \tag{91}$$

where (91) holds for any $\beta_n = O(\sqrt{n})$ using standard inequalities based on Taylor expansions; see Appendix D for details. Analogously to [19, Lemmas 17-18], we simplify (91) using two lemmas.

Lemma 1. *For any $\beta_n = O(\sqrt{n})$, we have*

$$\sum_{P_S \in \tilde{\mathcal{P}}_n} \mathbb{P}\big[\hat{P}_{\boldsymbol{S}} = P_S\big]\mathsf{Q}\bigg(\frac{\beta_n + nI(P_S) - nI(\pi)}{\sqrt{nV(P_S)}}\bigg) \le \sum_{P_S \in \tilde{\mathcal{P}}_n} \mathbb{P}\big[\hat{P}_{\boldsymbol{S}} = P_S\big]\mathsf{Q}\bigg(\frac{\beta_n + nI(P_S) - nI(\pi)}{\sqrt{nV(\pi)}}\bigg) + O\Big(\frac{\log n}{\sqrt{n}}\Big) \tag{92}$$

Proof: Since $V(P_S)$ is continuously differentiable at $P_S = \pi$ (see Appendix B), a Taylor expansion and the definition of $\tilde{\mathcal{P}}_n$ in (39) yields that the left-hand side of (92) is upper bounded by

$$\sum_{P_S \in \tilde{\mathcal{P}}_n} \mathbb{P}\big[\hat{P}_{\boldsymbol{S}} = P_S\big] \mathsf{Q}\left(\frac{\beta_n + nI(P_S) - nI(\pi)}{\sqrt{n\big(V(\pi) + K_7\sqrt{\frac{\log n}{n}}\big)}}\right) \tag{93}$$

for some constant K_7. In Appendix D, we show that (93) is upper bounded by the right-hand side of (92) using the assumption $\beta_n = O(\sqrt{n})$ along with standard inequalities based on Taylor expansions. ■

Lemma 1 is analogous to [19, Lemma 17], which is proved in a different manner using Hermite polynomials. The proof in [19] is somewhat more involved than that of Lemma 1, but it does not make the assumption that $\beta_n = O(\sqrt{n})$, and it yields a tighter $O\big(\frac{\log n}{n}\big)$ remainder term.

Lemma 2. *For any β_n, we have*

$$\sum_{P_S \in \tilde{\mathcal{P}}_n} \mathbb{P}\big[\hat{P}_{\boldsymbol{S}} = P_S\big] \mathsf{Q}\left(\frac{\beta_n + nI(P_S) - nI(\pi)}{\sqrt{nV(\pi)}}\right) \le \mathsf{Q}\left(\frac{\beta_n}{\sqrt{nV}}\right) + O\left(\frac{\log n}{\sqrt{n}}\right), \tag{94}$$

where V is defined in (24)*.*

Proof: Using the expansion of $I(P_S)$ in terms of $\tilde{I}(P_S)$ and $\Delta(P_S)$ given in (48), we have

$$\sum_{P_S \in \tilde{\mathcal{P}}_n} \mathbb{P}\big[\hat{P}_{\boldsymbol{S}} = P_S\big] \mathsf{Q}\left(\frac{\beta_n + nI(P_S) - nI(\pi)}{\sqrt{nV(\pi)}}\right) \tag{95}$$

$$= \sum_{P_S \in \tilde{\mathcal{P}}_n} \mathbb{P}\big[\hat{P}_{\boldsymbol{S}} = P_S\big] \mathsf{Q}\left(\frac{\beta_n - nI(\pi) + n\tilde{I}(P_S) + n\Delta(P_S)}{\sqrt{nV(\pi)}}\right) \tag{96}$$

$$\le \sum_{P_S \in \tilde{\mathcal{P}}_n} \mathbb{P}\big[\hat{P}_{\boldsymbol{S}} = P_S\big] \mathsf{Q}\left(\frac{\beta_n - nI(\pi) + n\tilde{I}(P_S) - 2K_1 \log n}{\sqrt{nV(\pi)}}\right) \tag{97}$$

$$= \sum_{P_S \in \tilde{\mathcal{P}}_n} \mathbb{P}\big[\hat{P}_{\boldsymbol{S}} = P_S\big] \mathsf{Q}\left(\frac{\beta_n - nI(\pi) + n\tilde{I}(P_S)}{\sqrt{nV(\pi)}}\right) + O\Big(\frac{\log n}{\sqrt{n}}\Big) \tag{98}$$

$$\le \sum_{P_S} \mathbb{P}\big[\hat{P}_{\boldsymbol{S}} = P_S\big] \mathsf{Q}\left(\frac{\beta_n - nI(\pi) + n\tilde{I}(P_S)}{\sqrt{nV(\pi)}}\right) + O\Big(\frac{\log n}{\sqrt{n}}\Big), \tag{99}$$

where (97) follows from (49) and since $\mathsf{Q}(\cdot)$ is decreasing, and (98) follows from the identity $|\mathsf{Q}(z) - \mathsf{Q}(z+a)| \le \frac{|a|}{\sqrt{2\pi}}$. Since $\tilde{I}(P_S)$ is written in the form $\sum_s P_S(s)\psi(s)$, a trivial generalization of [19, Lemma 18] gives

$$\sum_{P_S} \mathbb{P}\big[\hat{P}_{\boldsymbol{S}} = P_S\big] \mathsf{Q}\left(\frac{\beta_n + n\tilde{I}(P_S) - n\tilde{I}(\pi)}{\sqrt{nV(\pi)}}\right) - \mathsf{Q}\left(\frac{\beta_n}{\sqrt{n\big(V(\pi) + V^*(\pi)\big)}}\right) = O\Big(\frac{1}{\sqrt{n}}\Big), \tag{100}$$

where $V^*(\pi) \triangleq \mathrm{Var}_\pi[\psi(S)]$. Using the definition of $V(\cdot)$ in (85) and the fact that $\psi(S) = \mathbb{E}[i(U,Y) - i(U,S)\,|\,S]$ (see (46)), it follows that $V(\pi) + V^*(\pi)$ is equal to V, defined in (24). ■

Using (91) along with Lemmas 1 and 2, we have

$$\overline{p}_e \le \mathsf{Q}\left(\frac{\beta_n}{\sqrt{nV}}\right) + O\Big(\frac{\log n}{\sqrt{n}}\Big). \tag{101}$$

Setting $\overline{p}_e = \epsilon$ and solving for β_n, we obtain

$$\beta_n = \sqrt{nV}\mathsf{Q}^{-1}(\epsilon) + O(\log n). \tag{102}$$

Consistent with the step (91) and the statement of Lemma 1, we have $\beta_n = O(\sqrt{n})$. Finally, substituting (102) into (86), we obtain (23).

8) Equivalent form of V: It remains to show that (25) holds. We have

$$\mathrm{Var}[i(U,Y) - i(U,S)] = \mathbb{E}\big[\mathrm{Var}[i(U,Y) - i(U,S)\,|\,S,U]\big] + \mathrm{Var}\big[\mathbb{E}[i(U,Y) - i(U,S)\,|\,S,U]\big] \tag{103}$$

$$= \mathbb{E}\big[\mathrm{Var}[i(U,Y)\,|\,S,U]\big] + \mathrm{Var}\big[\mathbb{E}[i(U,Y) - i(U,S)\,|\,S,U]\big] \tag{104}$$

$$\geq \mathbb{E}\big[\mathrm{Var}[i(U,Y)\,|\,S,U]\big] + \mathrm{Var}\big[\mathbb{E}[i(U,Y) - i(U,S)\,|\,S]\big], \tag{105}$$

where (103) follows from the law of total variance, and (105) follows by again using the law of total variance to write

$$\mathbb{E}\big[\mathrm{Var}[\cdot|S,U]\big] + \mathrm{Var}\big[\mathbb{E}[\cdot|S,U]\big] = \mathbb{E}\big[\mathrm{Var}[\cdot|S]\big] + \mathrm{Var}\big[\mathbb{E}[\cdot|S]\big], \tag{106}$$

and since $\mathbb{E}\big[\mathrm{Var}[\cdot\,|\,S,U]\big] \leq \mathbb{E}\big[\mathrm{Var}[\cdot\,|\,S]\big]$. We show in Appendix C that whenever $Q_{U|S}$ maximizes the objective in (3), we have for any $s \in \mathcal{S}$ that the quantity $\xi'(s,u) \triangleq \mathbb{E}[i(u,Y) - i(u,s)\,|\,s,u]$ takes a fixed value $\xi(s)$ for all u such that $Q_{U|S}(u|s) > 0$. It follows that $\mathrm{Var}[\xi'(S,U)] = \mathrm{Var}[\xi(S)]$, and hence (105) holds with equality. Since we are considering capacity-achieving parameters, we obtain (25), thus completing the proof of Theorem 1.

IV. Proof for Dirty Paper Coding

In this section, we prove Theorem 2 by adapting the analysis of Section III to the Gaussian setting. To highlight the similarities in the proofs, we use similar or identical notation for analogous quantities.

A. Preliminary Results

1) Power Types: In place of types based on empirical distributions, we make use of *power types* (e.g. see [27]). We fix $\delta_s > 0$, and for each $P_S = \frac{k\delta_s}{n}$ $(k = 0,1,2,\cdots)$, we define the type class

$$T^n(P_S) \triangleq \Big\{\boldsymbol{s} \,:\, nP_S \leq \|\boldsymbol{s}\|^2 < nP_S + \delta_s\Big\}. \tag{107}$$

For each $\boldsymbol{s} \in T^n(P_S)$, we say that P_S is the type of $\boldsymbol{s}$, and we write $\hat{P}_{\boldsymbol{s}} = P_S$. That is, the type of a sequence is its power rounded down to the nearest multiple of $\frac{\delta_s}{n}$. The set of all types is given by $\mathcal{P}_n \triangleq \big\{\frac{k\delta_s}{n} \,:\, k \in \mathbb{Z}\big\}$.

2) A Typical Set of State Types: In general, the type P_S of $\boldsymbol{S}$ can be arbitrarily large with non-zero probability. However, analogously to (39), we can define a typical set of state types as follows:

$$\tilde{\mathcal{P}}_n \triangleq \big\{P_S \in \mathcal{P}_n \,:\, P_S \leq \Pi\big\}. \tag{108}$$

where Π appears in (26). We immediately obtain from (26) that

$$\mathbb{P}\big[\hat{P}_{\boldsymbol{S}} \notin \tilde{\mathcal{P}}_n\big] = O\Big(\frac{\log n}{\sqrt{n}}\Big). \tag{109}$$

Furthermore, the number of state types falling into $\tilde{\mathcal{P}}_n$ grows as $\Theta(n)$.

3) A Genie-Aided Setting: Analogously to the discrete case, we will prove Theorem 4 via the following result for a genie-aided setting.

Theorem 4. *The statement of Theorem 2 holds true in the case that the type P_S of $\boldsymbol{S}$ is known at the decoder.*

We proceed by showing that Theorem 4 implies Theorem 2. The arguments are similar to those following Theorem 3, so we only state the differences. We treat the event $\hat{P}_{\boldsymbol{S}} \notin \tilde{\mathcal{P}}_n$ as an error, thus leaving one of $\Theta(n)$ types to be transmitted to the receiver in $O(\log n)$ channel uses. This can be done provided that we can find a random-coding error exponent which is positive for sufficiently small rates. That is, we wish to show that there exists $\delta > 0$ and $\psi > 0$ such that for sufficiently large n the error probability does does not exceed $e^{-n\psi}$ for $R \leq \delta$.

From [28, Prop. 1], a positive exponent can be achieved for rates below $\frac{1}{2}\log\big(1+\frac{P}{1+P_{\max}}\big)$ even when the state sequence $\boldsymbol{S}$ is unknown at the encoder and arbitrarily varying subject to $\|\boldsymbol{S}\|^2 \leq nP_{\max}$. Since we have treated the event $\hat{P}_{\boldsymbol{S}} \notin \tilde{\mathcal{P}}_n$ as an error, it follows from (108) that the desired exponential decay is achieved for rates below $\frac{1}{2}\log\big(1+\frac{P}{1+\Pi}\big)$.

4) Type-Dependent Distributions: We will consider a decoder which makes use of an information density defined with respect to the joint distribution

$$f_{SUY}^{(P_S)}(s,u,y) = f_S^{(P_S)}(s)Q_{U|S}(u|s)f_{Y|SU}(y|s,u), \tag{110}$$

where in accordance with (5) and (8)–(9), we have

$$f_S^{(P_S)} \sim N(0,P_S) \tag{111}$$

$$Q_{U|S} \sim N(-\alpha S, P) \tag{112}$$

$$f_{Y|SU} \sim N\big(U+(1-\alpha)S, 1\big). \tag{113}$$

The parameter $\alpha > 0$ is assumed to be arbitrary for now. We can think of $f_{SUY}^{(P_S)}$ as being the joint density of (S,U,Y) induced by a Gaussian state $S \sim N(0,P_S)$, the channel W, and the choices of $Q_{U|S}$ and ϕ in (8)–(9). This joint density will play a major role in the analysis even though we are considering a possibly non-Gaussian state sequence. The induced output distribution is given by

$$f_Y^{(P_S)} \sim N(0, P+P_S+1), \tag{114}$$

and similarly to (12)–(13), the corresponding mutual informations are given by

$$I^{(P_S)}(U;Y) = \frac{1}{2}\log\left(\frac{(P+P_S+1)(P+\alpha^2P_S)}{PP_S(1-\alpha)^2+(P+\alpha^2P_S)}\right) \tag{115}$$

$$I^{(P_S)}(U;S) = \frac{1}{2}\log\left(\frac{P+\alpha^2P_S}{P}\right). \tag{116}$$

B. *Proof of Theorem 2*

As stated previously, in order to prove Theorem 2, it suffices to prove Theorem 4. Thus, we henceforth assume that the state type P_S is known at the decoder.

1) Random-Coding Parameters: The random coding parameters are the constant $\alpha > 0$ and the number of auxiliary codewords for each state type $P_S \in \mathcal{P}_n$, denoted by $L^{(P_S)}$. We will perform the analysis for an arbitrary choice of $\alpha > 0$, and then substitute $\alpha = \frac{P}{1+P}$ in accordance with (14).

2) Codebook Generation: For each state type $P_S \in \mathcal{P}_n$ and each message m, we randomly generate an auxiliary codebook $\mathcal{C}_U^{(P_S)}$ containing $ML^{(P_S)}$ auxiliary codewords $\{\boldsymbol{U}^{(P_S)}(m,l)\}_{l=1}^{L^{(P_S)}}$, where each codeword is independently distributed according to the uniform distribution on the sphere of power $n(P+\alpha^2 P_S)$, namely

$$f_{\boldsymbol{U}}^{(P_S)}(\boldsymbol{u}) = \frac{\delta\big(\|\boldsymbol{u}\|^2 - n(P+\alpha^2 P_S)\big)}{S_n\big(\sqrt{n(P+\alpha^2 P_S)}\big)}, \tag{117}$$

where $\delta(\cdot)$ is the Dirac delta function, and

$$S_n(r) = \frac{2\pi^{\frac{n}{2}}}{\Gamma\big(\frac{n}{2}\big)} r^{n-1} \tag{118}$$

is the surface area of a sphere of radius r in n-dimensional space. Each auxiliary codebook is revealed to the encoder and decoder.

3) Encoding and Decoding: Given the state sequence $\boldsymbol{S} \in T^n(P_S)$ and message m, the encoder sends

$$\boldsymbol{X} = \boldsymbol{U} - \alpha\boldsymbol{S}, \tag{119}$$

where $\boldsymbol{U}$ is an auxiliary codeword $\boldsymbol{U}^{(P_S)}(m,l)$ in $\mathcal{C}_U^{(P_S)}$, with l chosen such that $\boldsymbol{X} \in \mathcal{D}_n$, where

$$\mathcal{D}_n \triangleq \Big\{\boldsymbol{x} \,:\, nP - \delta_x \le \|\boldsymbol{x}\|^2 \le nP\Big\} \tag{120}$$

for some $\delta_x > 0$. If multiple such auxiliary codewords exist, one of them is chosen arbitrarily. An error is declared if no such auxiliary codeword exists. By construction, the power constraint in (7) is satisfied with probability one.

Given the received vector $\boldsymbol{y}$ and the state type P_S, the decoder estimates m according to the pair $(\tilde{m},\tilde{l})$ whose corresponding sequence $\boldsymbol{U}^{(P_S)}(\tilde{m},\tilde{l})$ maximizes

$$i_n^{(P_S)}(\boldsymbol{u},\boldsymbol{y}) \triangleq \sum_{i=1}^n i^{(P_S)}(u_i, y_i) \tag{121}$$

among the auxiliary codewords in $\mathcal{C}_U^{(P_S)}$, where

$$i^{(P_S)}(u,y) \triangleq \log \frac{f_{Y|U}^{(P_S)}(y|u)}{f_Y^{(P_S)}(y)} \tag{122}$$

with $f_{SUY}^{(P_S)}$ defined in (110).

We consider the events

$$\mathcal{E}_1 \triangleq \Big\{\text{No } l \text{ exists such that } \boldsymbol{U}^{(P_S)}(m,l) - \alpha\boldsymbol{S} \in \mathcal{D}_n\Big\} \tag{123}$$

$$\mathcal{E}_2 \triangleq \Big\{\text{Decoder chooses a message } \tilde{m} \neq m\Big\}. \tag{124}$$

It follows from these definitions and (109) that the overall random-coding error probability $\overline{p}_e$ satisfies

$$\overline{p}_e \le \sum_{P_S \in \tilde{\mathcal{P}}_n} \mathbb{P}\big[\hat{P}_{\boldsymbol{S}} = P_S\big]\Big(\mathbb{P}\big[\mathcal{E}_1 \,|\, P_S\big] + \mathbb{P}\big[\mathcal{E}_2 \,|\, P_S, \mathcal{E}_1^c\big]\Big) + O\Big(\frac{\log n}{\sqrt{n}}\Big). \tag{125}$$

4) Analysis of $\mathcal{E}_1$: We study the probability of $\mathcal{E}_1$ conditioned on $\boldsymbol{S}$ having a given type $P_S \in \tilde{\mathcal{P}}_n$. Recall the definition of $I^{(P_S)}(U;S)$ in (116). We claim that there exists a constant K_1 such that the rate $R_L^{(P_S)} \triangleq \frac{1}{n}\log L^{(P_S)}$ can be set to

$$R_L^{(P_S)} = I^{(P_S)}(U;S) + K_1 \frac{\log n}{n} \tag{126}$$

while achieving

$$\mathbb{P}\big[\mathcal{E}_1 \,|\, P_S\big] \le e^{-\psi n} \tag{127}$$

for some $\psi > 0$ and sufficiently large n. The key result in proving this claim is the following.

Lemma 3. *Fix $P_S \in \tilde{\mathcal{P}}_n$, and let $\boldsymbol{U}$ have density $f_{\boldsymbol{U}}^{(P_S)}$ (see (117)). For all $\boldsymbol{s} \in T^n(P_S)$ and sufficiently large n, we have*

$$\mathbb{P}\big[\boldsymbol{U} - \alpha\boldsymbol{s} \in \mathcal{D}_n\big] \ge \frac{1}{p_0(n)} e^{-I^{(P_S)}(U;Y)}, \tag{128}$$

for some polynomial $p_0(n)$ not depending on P_S.

Proof: See Appendix E. ■

We obtain (127) using Lemma 3 and following identical steps to those given in Section III-B4; the remaining details are omitted to avoid repetition.

5) Analysis of $\mathcal{E}_2$: We study the probability of $\mathcal{E}_2$ conditioned on $\boldsymbol{S}$ having a given type $P_S \in \tilde{\mathcal{P}}_n$, and also conditioned on $\mathcal{E}_1^c$. Let $f_{\boldsymbol{SU}}^{(P_S)}(\boldsymbol{s},\boldsymbol{u})$ denote the joint density of $(\boldsymbol{S},\boldsymbol{U})$ conditioned on these events, and let $\boldsymbol{Y}$ be the resulting output random variable, i.e.

$$(\boldsymbol{S},\boldsymbol{U},\boldsymbol{Y}) \sim f_{\boldsymbol{SU}}^{(P_S)}(\boldsymbol{s},\boldsymbol{u}) W^n(\boldsymbol{y}|\boldsymbol{u}-\alpha\boldsymbol{s},\boldsymbol{s}). \tag{129}$$

We do not attempt to give an explicit characterization of $f_{\boldsymbol{SU}}^{(P_S)}$. Instead, we will derive properties of the distribution which will be sufficient for performing the analysis; see Lemmas 4 and 5 below.

We again use the threshold-based bound given in (65), which states that

$$\mathbb{P}\big[\mathcal{E}_2 \,|\, P_S, \mathcal{E}_1^c\big] \le \mathbb{P}\Big[i_n^{(P_S)}(\boldsymbol{U},\boldsymbol{Y}) \le \gamma^{(P_S)}\Big] + ML^{(P_S)}\mathbb{P}\Big[i_n^{(P_S)}(\overline{\boldsymbol{U}},\boldsymbol{Y}) > \gamma^{(P_S)}\Big] \tag{130}$$

for any $\gamma^{(P_S)}$, where $\overline{\boldsymbol{U}} \sim f_{\boldsymbol{U}}^{(P_S)}$ is independent of $(\boldsymbol{S},\boldsymbol{U},\boldsymbol{Y})$. We further upper bound (130) by maximizing over $(\boldsymbol{s},\boldsymbol{u})$:

$$\mathbb{P}\big[\mathcal{E}_2 \,|\, P_S, \mathcal{E}_1^c\big] \le \max_{(\boldsymbol{s},\boldsymbol{u})\,:\, f_{\boldsymbol{SU}}^{(P_S)}(\boldsymbol{s},\boldsymbol{u})>0} \mathbb{P}\Big[i_n^{(P_S)}(\boldsymbol{u},\boldsymbol{Y}) \le \gamma^{(P_S)} \,\Big|\, \boldsymbol{s},\boldsymbol{u}\Big] + ML^{(P_S)}\mathbb{P}\Big[i_n^{(P_S)}(\overline{\boldsymbol{U}},\boldsymbol{Y}) > \gamma^{(P_S)} \,\Big|\, \boldsymbol{s},\boldsymbol{u}\Big]. \tag{131}$$

The analysis of the second term in (131) is simplified by the following lemma.

Lemma 4. *Fix $P_S \in \tilde{\mathcal{P}}_n$ and $(\boldsymbol{s},\boldsymbol{u})$ such that $f_{\boldsymbol{SU}}^{(P_S)}(\boldsymbol{s},\boldsymbol{u}) > 0$, and define the random variables*

$$(\boldsymbol{X}' \,|\, \boldsymbol{s},\boldsymbol{u}) \sim \frac{\delta\big(\|\boldsymbol{x}'\| - \|\boldsymbol{u} + (1-\alpha)\boldsymbol{s}\|\big)}{S_n\big(\|\boldsymbol{u} + (1-\alpha)\boldsymbol{s}\|\big)} \tag{132}$$

$$\boldsymbol{Y}' = \boldsymbol{X}' + \boldsymbol{Z}, \tag{133}$$

where S_n is defined in (118)*, and $\boldsymbol{Z}$ is the additive noise in* (5)*. For $\overline{\boldsymbol{U}} \sim f_{\boldsymbol{U}}^{(P_S)}$ independent of $(\boldsymbol{S}, \boldsymbol{U}, \boldsymbol{Y}, \boldsymbol{X}', \boldsymbol{Y}')$, we have*

$$\mathbb{P}\Big[i_n^{(P_S)}(\overline{\boldsymbol{U}}, \boldsymbol{Y}) > \gamma^{(P_S)} \,\Big|\, \boldsymbol{s}, \boldsymbol{u}\Big] = \mathbb{P}\Big[i_n^{(P_S)}(\overline{\boldsymbol{U}}, \boldsymbol{Y}') > \gamma^{(P_S)} \,\Big|\, \boldsymbol{s}, \boldsymbol{u}\Big]. \tag{134}$$

Furthermore, letting $f_{\boldsymbol{Y}'|\boldsymbol{SU}}^{(P_S)}$ denote the density of $\boldsymbol{Y}'$ given $(\boldsymbol{s}, \boldsymbol{u})$, there exists $\epsilon > 0$ such that

$$\mathbb{P}\Big[\big|\|\boldsymbol{Y}'\|^2 - n(P + P_S + 1)\big| > n\epsilon \,\Big|\, \boldsymbol{s}, \boldsymbol{u}\Big] = O\big(e^{-\psi n}\big) \tag{135}$$

$$\min_{\boldsymbol{y}' :\, \big|\|\boldsymbol{y}'\|^2 - n(P+P_S+1)\big| \leq n\epsilon} \frac{f_{\boldsymbol{Y}'|\boldsymbol{SU}}^{(P_S)}(\boldsymbol{y}'|\boldsymbol{s}, \boldsymbol{u})}{\prod_{i=1}^n f_Y^{(P_S)}(y_i')} \leq K_2 \tag{136}$$

for some constants $\psi > 0$ and K_2 not depending on P_S, where $f_Y^{(P_S)}$ is defined in (114)*.*

Proof: See Appendix F. ∎

Using Lemma 4, we can bound the second probability in (131) as follows:

$$\begin{aligned}
&\mathbb{P}\Big[i_n^{(P_S)}(\overline{\boldsymbol{U}}, \boldsymbol{Y}) > \gamma^{(P_S)} \,\big|\, \boldsymbol{s}, \boldsymbol{u}\Big] \\
&\quad = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f_{\boldsymbol{U}}^{(P_S)}(\overline{\boldsymbol{u}}) f_{\boldsymbol{Y}'|\boldsymbol{SU}}^{(P_S)}(\boldsymbol{y}'|\boldsymbol{s}, \boldsymbol{u}) \mathbb{1}\big\{i_n^{(P_S)}(\overline{\boldsymbol{u}}, \boldsymbol{y}') > \gamma^{(P_S)}\big\} d\overline{\boldsymbol{u}} d\boldsymbol{y}' &(137)\\
&\quad = \int_{\big|\|\boldsymbol{y}'\|^2 - n(P+P_S+1)\big| \leq n\epsilon} \int_{\mathbb{R}^n} f_{\boldsymbol{U}}^{(P_S)}(\overline{\boldsymbol{u}}) f_{\boldsymbol{Y}'|\boldsymbol{SU}}^{(P_S)}(\boldsymbol{y}'|\boldsymbol{s}, \boldsymbol{u}) \mathbb{1}\big\{i_n^{(P_S)}(\overline{\boldsymbol{u}}, \boldsymbol{y}') > \gamma^{(P_S)}\big\} d\overline{\boldsymbol{u}} d\boldsymbol{y}' + O\big(e^{-\psi n}\big) &(138)\\
&\quad \leq K_2 e^{-\gamma^{(P_S)}} + e^{-\psi n}, &(139)
\end{aligned}$$

where (138) follows from (135), and (139) follows by upper bounding $f_{\boldsymbol{Y}'|\boldsymbol{SU}}^{(P_S)}$ using (136) and following the steps in (70)–(73).

We choose

$$\gamma^{(P_S)} = \log M + nI^{(P_S)}(U; S) + \log n, \tag{140}$$

which, when combined with (126), yields

$$\gamma^{(P_S)} = \log M + nI^{(P_S)}(U; S) + K_3 \log n \tag{141}$$

with $K_3 \triangleq K_1 + 1$. Combining (131) and (139) with this choice of $\gamma^{(P_S)}$, we conclude that

$$\mathbb{P}\big[\mathcal{E}_2 \,|\, P_S, \mathcal{E}_1^c\big] \leq \mathbb{P}\Big[i_n^{(P_S)}(\boldsymbol{u}, \boldsymbol{Y}) \leq \log M + nI^{(P_S)}(U; S) + K_3 \log n \,\Big|\, \boldsymbol{s}, \boldsymbol{u}\Big] + O\Big(\frac{1}{n}\Big) \tag{142}$$

for some $(\boldsymbol{s}, \boldsymbol{u})$ such that $f_{\boldsymbol{SU}}^{(P_S)}(\boldsymbol{s}, \boldsymbol{u}) > 0$.

6) Application of the Berry-Esseen Theorem: The moments associated with $i_n^{(P_S)}(\boldsymbol{u}, \boldsymbol{Y})$ required to apply the Berry-Esseen theorem are characterized in the following lemma.

Lemma 5. *Fix $P_S \in \tilde{\mathcal{P}}_n$ and $(\boldsymbol{s}, \boldsymbol{u})$ such that $f_{\boldsymbol{SU}}^{(P_S)}(\boldsymbol{s}, \boldsymbol{u}) > 0$. We have*

$$\mathbb{E}\big[i_n^{(P_S)}(\boldsymbol{u}, \boldsymbol{Y}) \,|\, \boldsymbol{s}, \boldsymbol{u}\big] = nI^{(P_S)}(U; Y) + O(1), \tag{143}$$

and for $\alpha = \frac{P}{1+P}$ we have

$$\mathrm{Var}\big[i_n^{(P_S)}(\boldsymbol{u}, \boldsymbol{Y}) \,|\, \boldsymbol{s}, \boldsymbol{u}\big] = nV + O(1), \tag{144}$$

where V is defined in (28)*. Furthermore, there exists a pair $(\boldsymbol{s}', \boldsymbol{u}')$ such that $i_n^{(P_S)}(\boldsymbol{U}, \boldsymbol{Y})$ has the same distribution whether conditioned on $(\boldsymbol{S}, \boldsymbol{U}) = (\boldsymbol{s}', \boldsymbol{u}')$* or *$(\boldsymbol{S}, \boldsymbol{U}) = (\boldsymbol{s}, \boldsymbol{u})$, and such that*

$$\sum_{i=1}^{n} \mathbb{E}\Big[\Big|i^{(P_S)}(u_i', Y_i) - \mathbb{E}\big[i^{(P_S)}(u_i', Y_i)\big]\Big|^3 \,\Big|\, s_i', u_i'\Big] = O(n). \tag{145}$$

The remainder terms in (143)–(145) *are uniform in $P_S \in \tilde{\mathcal{P}}_n$.*

Proof: See Appendix F. ∎

Combining (127) and (142), we have some $(\boldsymbol{s}, \boldsymbol{u})$ that

$$\mathbb{P}\big[\mathcal{E}_1 \cup \mathcal{E}_2 \,|\, P_S\big] \le \mathbb{P}\Big[i_n^{(P_S)}(\boldsymbol{u}, \boldsymbol{Y}) \le \log M + nI^{(P_S)}(U;S) + K_3 \log n \,\Big|\, \boldsymbol{s}, \boldsymbol{u}\Big] + O\Big(\frac{1}{n}\Big) \tag{146}$$

$$\le \mathbb{P}\Big[i_n^{(P_S)}(\boldsymbol{u}, \boldsymbol{Y}) - \mathbb{E}\big[i_n^{(P_S)}(\boldsymbol{u}, \boldsymbol{Y})\big] \le \log M - nI(P_S) + K_3 \log n + K_4 \,\Big|\, \boldsymbol{s}, \boldsymbol{u}\Big] + O\Big(\frac{1}{n}\Big), \tag{147}$$

where (147) follows from (143) and by defining

$$I(P_S) \triangleq I^{(P_S)}(U;Y) - I^{(P_S)}(U;S). \tag{148}$$

The constant K_4 in (147) represents the uniform $O(1)$ term in (143).

The mutual informations $I^{(P_S)}(U;Y)$ and $I^{(P_S)}(U;S)$ are given in (115)–(116), and similarly to (15), setting $\alpha = \frac{P}{1+P}$ yields $I(P_S) = C$ for all P_S. Thus, applying the Berry-Esseen theorem [25, Sec. XVI.5] to (147) (after replacing $(\boldsymbol{s}, \boldsymbol{u})$ by $(\boldsymbol{s}', \boldsymbol{u}')$ given in Lemma 5 if necessary), we obtain for all $P_S \in \tilde{\mathcal{P}}_n$ that

$$\mathbb{P}\big[\mathcal{E}_1 \cup \mathcal{E}_2 \,|\, P_S\big] \le \mathsf{Q}\Big(\frac{\log M - nC + K_3 \log n + K_4}{\sqrt{nV + K_5}}\Big) + O\Big(\frac{1}{\sqrt{n}}\Big) \tag{149}$$

for some constant K_5. Substituting (149) into (125) yields

$$\overline{p}_e \le \mathsf{Q}\Big(\frac{\log M - nC + K_3 \log n + K_4}{\sqrt{nV + K_5}}\Big) + O\Big(\frac{\log n}{\sqrt{n}}\Big), \tag{150}$$

and the proof of Theorem 2 is concluded by inverting the relationship between the error probability and the number of messages.

V. Concluding Remarks

We have presented an achievable second-order coding rate for the discrete memoryless Gel'fand-Pinsker channel, and a conclusive characterization of the second-order asymptotics for dirty paper coding. Possible areas of further research include non-asymptotic bounds and their comparison to the normal approximations obtained by omitting the higher-order terms in (23) and (27), and second-order converse results for the discrete case. For the latter, the techniques used in proving the strong converse [29], [30] may prove useful.

The assumption $\mathbb{E}\big[\mathrm{Var}[i(U,Y) \,|\, S, U]\big] > 0$ in Theorem 23 is analogous to similar assumptions of positive dispersions in other settings (e.g. see [19], [31]). When $\mathbb{E}\big[\mathrm{Var}[i(U,Y) \,|\, S, U]\big] = 0$, the analysis in Section III remains valid until (88), but there are several difficulties in generalizing the subsequent analysis. First, it does not necessarily follow that that $\mathbb{E}\big[\mathrm{Var}[i^{(P_S)}(U,Y) \,|\, S, U]\big] = 0$ under $P_{SUY}^{(P_S)}$ (see (42)) for all $P_S \in \tilde{\mathcal{P}}_n$, and hence we may still need to consider variances of up to $O\Big(\sqrt{\frac{\log n}{n}}\Big)$. Second, the behavior of the probability in (88) varies depending on whether $I(\pi) > I(P_S)$ or $I(\pi) < I(P_S)$, both of which can occur with differences of up to $O\left(\sqrt{\frac{\log n}{n}}\right)$. Finally, [19, Lemma 18] (which is used in the proof of Lemma 2) is based on properties of the Q-function, and thus may be difficult to extend if an alternative bound (e.g. Chebyshev's inequality) is used in place of the Berry-Esseen theorem following (88).

A by-product of our analysis for the Gaussian case (dirty paper coding) is an alternative viewpoint as to why similar performance is achieved for Gaussian or non-Gaussian state sequences: By using a small fraction of the block to send a quantized version of the state power, we can make the decoder aware that the sequence lies within a thin spherical shell. Since all sequences within that shell are essentially equally difficult to handle, the precise statistics of the state sequence are not important.

Appendix

A. Proof of Corollary 1

We apply Theorem 1 with (Y,S) in place of Y. The capacity-achieving parameters are $\mathcal{U}=\mathcal{X}$, $\phi(u,s)=u$ (i.e. $x=u$) and $Q_{U|S}(\cdot|s)=Q(\cdot|s)$, where $Q(\cdot|s)$ achieves the capacity-dispersion pair (C_s,V_s) for the channel $W(\cdot|\cdot,s)$. To avoid ambiguity, we denote the resulting information densities in (17)–(18) by $i_1(x,s)$ and $i_2(x,y,s)$ respectively. Defining

$$P_{SXY}(s,x,y)\triangleq\pi(s)Q(x|s)W(y|x,s), \tag{151}$$

we have

$$i_1(x,s)=\log\frac{P_{SX}(s,x)}{P_S(s)P_X(x)} \tag{152}$$

$$i_2(x,y,s)=\log\frac{P_{SXY}(s,x,y)}{P_X(x)P_{SY}(s,y)}. \tag{153}$$

It follows that

$$\begin{aligned} i_2(x,y,s)-i_1(x,s)&=\log\frac{P_{SXY}(s,x,y)P_S(s)}{P_{SY}(s,y)P_{SX}(s,x)} &(154)\\ &=\log\frac{P_{XY|S}(x,y|s)}{P_{Y|S}(y|s)P_{X|S}(x|s)} &(155)\\ &\triangleq i_3(x,y,s). &(156)\end{aligned}$$

We observe that $i_3(\cdot,\cdot,s)$ is the information density associated with $W(\cdot|\cdot,s)$, and thus has mean C_s and variance V_s [2]. It follows that V in (25) can be written as

$$\begin{aligned} V&=\mathrm{Var}[i_3(X,Y,S)] &(157)\\ &=\mathbb{E}\big[\mathrm{Var}[i_3(X,Y,S)\,|\,S]\big]+\mathrm{Var}\big[\mathbb{E}[i_3(X,Y,S)\,|\,S]\big] &(158)\\ &=\mathbb{E}[V_S]+\mathrm{Var}[C_S], &(159)\end{aligned}$$

where (158) follows from the law of total variance. We similarly have $C=\mathbb{E}[i_3(X,Y,S)]=\mathbb{E}[C_S]$, thus completing the proof.

B. Continuous Differentiability and Taylor Expansions

In this section, we study the differentiability properties of $I(P_S)$ (see (45)) and $V(P_S)$ (see (85)), and prove the Taylor expansion given (48).

1) Derivatives of $I(P_S)$: Writing

$$I^{(P_S)}(U;S) = \sum_{s,u} P_S(s) Q_{U|S}(u|s) \log \frac{Q_{U|S}(u|s)}{\sum_{\overline{s}} P_S(\overline{s}) Q_{U|S}(u|\overline{s})}, \tag{160}$$

we obtain

$$\frac{\partial I^{(P_S)}(U;S)}{\partial P_S(s')} = \sum_{s \neq s', u} P_S(s) Q_{U|S}(u|s) \frac{\partial}{\partial P_S(s')} \Big(-\log \sum_{\overline{s}} P_S(\overline{s}) Q_{U|S}(u|\overline{s}) \Big) + \frac{\partial}{\partial P_S(s')} \sum_u P_S(s') Q_{U|S}(u|s') \log \frac{Q_{U|S}(u|s')}{\sum_{\overline{s}} P_S(\overline{s}) Q_{U|S}(u|\overline{s})} \tag{161}$$

$$= -\sum_{s,u} P_S(s) Q_{U|S}(u|s) \frac{Q_{U|S}(u|s')}{\sum_{\overline{s}} P_S(\overline{s}) Q_{U|S}(u|\overline{s})} + \sum_u Q_{U|S}(u|s') \log \frac{Q_{U|S}(u|s')}{\sum_{\overline{s}} P_S(\overline{s}) Q_{U|S}(u|\overline{s})} \tag{162}$$

$$= -\sum_{s,u} P_{S|U}(s|u) Q_{U|S}(u|s') + \sum_u Q_{U|S}(u|s') \log \frac{Q_{U|S}(u|s')}{\sum_{\overline{s}} P_S(\overline{s}) Q_{U|S}(u|\overline{s})} \tag{163}$$

$$= -1 + \sum_u Q_{U|S}(u|s') \log \frac{Q_{U|S}(u|s')}{\sum_{\overline{s}} P_S(\overline{s}) Q_{U|S}(u|\overline{s})}, \tag{164}$$

where (163) follows by writing $P_{S|U}(s|u) = \frac{P_S(s) Q_{U|S}(u|s)}{P_U(u)}$.

The derivative of $I^{(P_S)}(U;Y)$ is computed similarly. We have

$$I^{(P_S)}(U;Y) = \sum_{s,u,y} P_S(s) P_{UY|S}(u,y|s) \log \frac{P_{UY}^{(P_S)}(u,y)}{P_U^{(P_S)}(u) P_Y^{(P_S)}(y)} \tag{165}$$

$$= \sum_{s,u,y} P_S(s) P_{UY|S}(u,y|s) \Big(\log P_{UY}^{(P_S)}(u,y) - \log P_U^{(P_S)}(u) - \log P_Y^{(P_S)}(y) \Big), \tag{166}$$

where $P_{UY|S}(u,y|s) = Q_{U|S}(u|s) W(y|\phi(u,s),s)$ does not depend on P_S. We can write

$$P_{UY}^{(P_S)}(u,y) = \sum_s P_S(s) P_{UY|S}(u,y|s) \tag{167}$$

and similarly for P_U and P_Y, yielding the derivatives

$$\frac{\partial P_{UY}^{(P_S)}(u,y)}{\partial P_S(s')} = P_{UY|S}(u,y|s') \tag{168}$$

$$\frac{\partial P_U^{(P_S)}(u)}{\partial P_S(s)} = P_{U|S}(u|s') \tag{169}$$

$$\frac{\partial P_Y^{(P_S)}(y)}{\partial P_S(s')} = P_{Y|S}(y|s'). \tag{170}$$

It follows using the same arguments as (161)–(164) that

$$\frac{\partial I^{(P_S)}(U;Y)}{\partial P_S(s')} = \sum_{s,u,y} P_S(s) P_{UY|S}(u,y|s)\bigg(\frac{P_{UY|S}(u,y|s')}{P_{UY}^{(P_S)}(u,y)} - \frac{P_{U|S}(u|s')}{P_U^{(P_S)}(u)} - \frac{P_{Y|S}(y|s')}{P_Y^{(P_S)}(y)}\bigg)$$
$$+ \sum_{u,y} P_{UY|S}(u,y|s') \log \frac{P_{UY}^{(P_S)}(u,y)}{P_U^{(P_S)}(u) P_Y^{(P_S)}(y)} \tag{171}$$
$$= -1 + \sum_{u,y} P_{UY|S}(u,y|s') \log \frac{P_{UY}^{(P_S)}(u,y)}{P_U^{(P_S)}(u) P_Y^{(P_S)}(y)}. \tag{172}$$

Differentiating (164) and (172) a second time, we obtain

$$\frac{\partial^2 I^{(P_S)}(U;S)}{\partial P_S(s')\partial P_S(s'')} = -\sum_u Q_{U|S}(u|s') \frac{Q_{U|S}(u|s'')}{\sum_{\overline{s}} P_S(\overline{s}) Q_{U|S}(u|\overline{s})} \tag{173}$$

$$\frac{\partial^2 I^{(P_S)}(U;Y)}{\partial P_S(s')\partial P_S(s'')} = \sum_{u,y} P_{UY|S}(u,y|s')\bigg(\frac{P_{UY|S}(u,y|s'')}{P_{UY}^{(P_S)}(u,y)} - \frac{P_{U|S}(u|s'')}{P_U^{(P_S)}(u)} - \frac{P_{Y|S}(y|s'')}{P_Y^{(P_S)}(y)}\bigg). \tag{174}$$

2) Continuous Differentiability : Using (167), we observe the derivatives in (164) and (172)–(174) are continuous in P_S other than a possible divergence as $\min_s P_S(s) \to 0$. Assuming without loss of generality that $\min_s \pi(s) > 0$, it follows that $I(P_S)$ is twice continuously differentiable within $\tilde{\mathcal{P}}_n$ (see (39)) for sufficiently large n.

For $V(P_S)$ (see (85)) we only require that the first derivatives are continuous within $\tilde{\mathcal{P}}_n$. This can be proved by writing $V(P_S)$ in the form

$$V(P_S) = \sum_{s,u} P_S(s) Q_{U|S}(u|s) V^{(P_S)}(u,s), \tag{175}$$

where

$$V^{(P_S)}(u,s) \triangleq \sum_y W(y|\phi(u,s),s)\bigg(\log \frac{P_{UY}^{(P_S)}(u,y)}{P_U^{(P_S)}(u) P_Y^{(P_S)}(y)}\bigg)^2 - \bigg(\sum_y W(y|\phi(u,s),s) \log \frac{P_{UY}^{(P_S)}(u,y)}{P_U^{(P_S)}(u) P_Y^{(P_S)}(y)}\bigg)^2. \tag{176}$$

The subsequent evaluation of the partial derivatives is cumbersome and similar to the analysis following (165), and is thus omitted.

3) Taylor Expansion of $I(P_S)$: The first-order Taylor approximation of $I(P_S) = I^{(P_S)}(U;Y) - I^{(P_S)}(U;S)$ at $P_S = \pi$ is given by

$$I(P_S) = I(\pi) + \sum_s \big(P_S(s) - \pi(s)\big) \frac{\partial I(P_S)}{\partial P_S(s)}\bigg|_{P_S=\pi} + \Delta(P_S), \tag{177}$$

where $\Delta(P_S)$ is the remainder term. From (164) and (172), we see that $\sum_s \pi(s) \frac{\partial I(P_S)}{\partial P_S(s)}\Big|_{P_S=\pi} = I(\pi)$, and hence the right-hand side of (177) equals $\sum_s P_S(s) \frac{\partial I(P_S)}{\partial P_S(s)}\Big|_{P_S=\pi} + \Delta(P_S)$, which in turn equals $\tilde{I}(P_S) + \Delta(P_S)$ (see (47)). The remainder term satisfies (49) since $I(P_S)$ is twice continuously differentiable within $\tilde{\mathcal{P}}_n$, and since the ℓ_2-norm and ℓ_∞-norm coincide to within a constant factor.

C. Necessary Conditions for the Optimal Input Distribution

Here we study the necessary Karush-Kuhn-Tucker (KKT) conditions [32, Sec. 5.5.3] for $Q_{U|S}$ to maximize the objective in (3) when $\mathcal{U}$ and $\phi(\cdot,\cdot)$ are fixed. Introducing the Lagrange multiplier $\lambda(s)$ corresponding to the constraint $\sum_u Q_{U|S}(u|s) = 1$, we see that any optimal $Q_{U|S}$ must satisfy

$$\frac{\partial}{\partial Q_{U|S}(u'|s')}\Big(I(U;Y) - I(U;S)\Big) = \lambda(s') \tag{178}$$

for all (s', u') such that $Q_{U|S}(u'|s') > 0$; we assume without loss of generality that $\min_s \pi(s) > 0$.

Using (160) with $P_S = \pi$, and writing the logarithm as a difference of two logarithms, we obtain

$$\begin{aligned}\frac{\partial I(U;S)}{\partial Q_{U|S}(u'|s')} &= \pi(s') \log Q_{U|S}(u'|s') + \pi(s') - \pi(s') \log \sum_{\overline{s}} \pi(\overline{s}) Q_{U|S}(u'|\overline{s}) \\ &\qquad - \sum_s \pi(s) Q_{U|S}(u'|s) \frac{\pi(s')}{\sum_{\overline{s}} \pi(\overline{s}) Q_{U|S}(u'|\overline{s})} \qquad (179) \\ &= \pi(s') \log \frac{Q_{U|S}(u'|s')}{\sum_{\overline{s}} \pi(\overline{s}) Q_{U|S}(u'|\overline{s})}, \qquad (180)\end{aligned}$$

where (180) follows by applying Bayes' rule to the last term in (179). To evaluate the partial derivatives of $I(U;Y)$, we write (166) (with $P_S = \pi$) as

$$I(U;Y) = \sum_{s,u,y} \pi(s) Q_{U|S}(u|s) W(y|\phi(u,s),s)\Big(\log P_{UY}(u,y) - \log P_U(u) - \log P_Y(y)\Big).$$

We have $P_{UY}(u,y) = \sum_s \pi(s) Q_{U|S}(u|s) W(y|\phi(u,s),s)$ (and similarly for P_U and P_Y), yielding the derivatives

$$\frac{\partial P_{UY}(u,y)}{\partial Q_{U|S}(u'|s')} = \begin{cases} \pi(s') W(y|\phi(u',s'),s') & u = u' \\ 0 & \text{otherwise} \end{cases} \tag{181}$$

$$\frac{\partial P_U(u)}{\partial Q_{U|S}(u'|s')} = \begin{cases} \pi(s') & u = u' \\ 0 & \text{otherwise} \end{cases} \tag{182}$$

$$\frac{\partial P_Y(y)}{\partial Q_{U|S}(u'|s')} = \pi(s') W(y|\phi(u',s'),s'). \tag{183}$$

It follows using a similar argument to (179)–(180) that

$$\frac{\partial I(U;Y)}{\partial Q_{U|S}(u'|s')} = \pi(s')\left(\sum_y W(y|\phi(u',s'),s') \log \frac{P_{UY}(u',y)}{P_U(u')P_Y(y)} - 1\right). \tag{184}$$

Combining (178), (180) and (184), we see that for any optimal $Q_{U|S}$ and any $s \in \mathcal{S}$, the quantity

$$\sum_y W(y|\phi(u,s),s) \log \frac{P_{UY}(u,y)}{P_U(u)P_Y(y)} - \log \frac{Q_{U|S}(u|s)}{\sum_{\overline{s}} \pi(\overline{s}) Q_{U|S}(u|\overline{s})} \tag{185}$$

is the same for all u such that $Q_{U|S}(u|s) > 0$. Note that (185) can be written more compactly as $\mathbb{E}[i(u,Y) - i(u,s) \,|\, s, u]$; see (17)–(18).

D. Proofs of Steps Involving Taylor Expansions

In this section, we make use of the fact that

$$\min_{P_S \in \tilde{\mathcal{P}}_n} V(P_S) \geq V_{\min} \tag{186}$$

for some $V_{\min} > 0$ and sufficiently large n, which follows from the definition of $V(P_S)$ in (85), the assumption of Theorem 1, and the fact that $P_S \to \pi$ within $\tilde{\mathcal{P}}_n$ (see (39)).

1) Proof of (91)*:* We first eliminate K_5 from (90) by writing

$$\sum_{P_S\in\tilde{\mathcal{P}}_n}\mathbb{P}[P_S]\mathsf{Q}\bigg(\frac{\beta_n+nI(P_S)-nI(\pi)+K_5}{\sqrt{nV(P_S)+K_6}}\bigg)=\sum_{P_S\in\tilde{\mathcal{P}}_n}\mathbb{P}[P_S]\mathsf{Q}\bigg(\frac{\beta_n+nI(P_S)-nI(\pi)}{\sqrt{nV(P_S)+K_6}}\bigg)+O\Big(\frac{1}{\sqrt{n}}\Big),\quad(187)$$

which follows from (186) and the identity $|\mathsf{Q}(z)-\mathsf{Q}(z+a)|\le\frac{|a|}{\sqrt{2\pi}}$. It remains to eliminate K_6. The case $K_6\le 0$ is trivial, so we assume that $K_6>0$. We have

$$\frac{1}{\sqrt{nV(P_S)+K_6}}=\frac{1}{\sqrt{nV(P_S)}\sqrt{1+\frac{K_6}{nV(P_S)}}}\quad(188)$$

$$\ge\frac{1}{\sqrt{nV(P_S)}}\Big(1-\frac{K_6'}{n}\Big),\quad(189)$$

where (189) follows with $K_6'=\frac{K_6}{2V_{\min}}$ using (186) and the identity $\frac{1}{\sqrt{1+z}}\ge 1-\frac{z}{2}$. We thus obtain

$$\sum_{P_S\in\tilde{\mathcal{P}}_n}\mathbb{P}[P_S]\mathsf{Q}\bigg(\frac{\beta_n+nI(P_S)-nI(\pi)}{\sqrt{nV(P_S)+K_6}}\bigg)\quad(190)$$

$$\le\sum_{P_S\in\tilde{\mathcal{P}}_n}\mathbb{P}[P_S]\mathsf{Q}\bigg(\frac{\beta_n+nI(P_S)-nI(\pi)}{\sqrt{nV(P_S)}}\Big(1-\frac{K_6'}{n}\Big)\bigg)\quad(191)$$

$$\le\sum_{P_S\in\tilde{\mathcal{P}}_n}\mathbb{P}[P_S]\mathsf{Q}\bigg(\frac{\beta_n+nI(P_S)-nI(\pi)}{\sqrt{nV(P_S)}}\bigg)+\frac{K_6'}{\sqrt{2\pi}n}\sum_{P_S\in\tilde{\mathcal{P}}_n}\mathbb{P}[P_S]\frac{|\beta_n+nI(P_S)-nI(\pi)|}{\sqrt{nV(P_S)}}\quad(192)$$

$$=\sum_{P_S\in\tilde{\mathcal{P}}_n}\mathbb{P}[P_S]\mathsf{Q}\bigg(\frac{\beta_n+nI(P_S)-nI(\pi)}{\sqrt{nV(P_S)}}\bigg)+O\Big(\frac{1}{\sqrt{n}}\Big),\quad(193)$$

where (192) follows from the identity $|\mathsf{Q}(z)-\mathsf{Q}(z+a)|\le\frac{|a|}{\sqrt{2\pi}}$, and (193) follows since $\beta_n=O(\sqrt{n})$ by assumption.

2) Upper Bound on (93)*:* Using the same argument as the one leading to (192), we have for some K_7' that

$$\sum_{P_S\in\tilde{\mathcal{P}}_n}\mathbb{P}[P_S]\mathsf{Q}\left(\frac{\beta_n+nI(P_S)-nI(\pi)}{\sqrt{n\big(V(\pi)+K_7\sqrt{\frac{\log n}{n}}\big)}}\right)$$

$$\le\sum_{P_S\in\tilde{\mathcal{P}}_n}\mathbb{P}[P_S]\mathsf{Q}\left(\frac{\beta_n+nI(P_S)-nI(\pi)}{\sqrt{nV(\pi)}}\right)+K_7'\sqrt{\frac{\log n}{n}}\sum_{P_S\in\tilde{\mathcal{P}}_n}\mathbb{P}[P_S]\frac{|\beta_n+nI(P_S)-nI(\pi)|}{\sqrt{nV(P_S)}}.\quad(194)$$

We now analyze the growth rate of the second term. Applying (186), we can upper bound this term by

$$\frac{K_7'}{\sqrt{V_{\min}}}\frac{\sqrt{\log n}}{n}\sum_{P_S\in\tilde{\mathcal{P}}_n}\mathbb{P}[P_S]\big|\beta_n+nI(P_S)-nI(\pi)\big|.\quad(195)$$

Since $I(\cdot)$ is continuously differentiable (see Appendix B), we have from (39) that $\max_{P_S\in\tilde{\mathcal{P}}_n}|I(P_S)-I(\pi)|\le K_8\sqrt{\frac{\log n}{n}}$ for some constant K_8. Using this observation along with $\beta_n=O(\sqrt{n})$, we obtain

$$\frac{K_7'}{\sqrt{V_{\min}}}\frac{\sqrt{\log n}}{n}\sum_{P_S\in\tilde{\mathcal{P}}_n}\mathbb{P}[P_S]\big|\beta_n+nI(P_S)-nI(\pi)\big|\le\frac{K_7'}{\sqrt{V_{\min}}}\frac{\sqrt{\log n}}{n}\sum_{P_S\in\tilde{\mathcal{P}}_n}\mathbb{P}[P_S]\Big(|\beta_n|+K_8\sqrt{n\log n}\Big)\quad(196)$$

$$=O\Big(\frac{\log n}{\sqrt{n}}\Big).\quad(197)$$

Substituting (197) into (194), we obtain the desired result.

E. Proof of Lemma 3

Recall that $\|\boldsymbol{U}\|^2 = n(P + \alpha^2 P_S)$ almost surely and $\boldsymbol{s} \in T^n(P_S)$ by assumption, and let $(\boldsymbol{s}, \boldsymbol{u})$ be fixed accordingly. Writing $\|\boldsymbol{u} - \alpha\boldsymbol{s}\|^2 = \|\boldsymbol{u}\|^2 - 2\alpha\langle\boldsymbol{s}, \boldsymbol{u}\rangle + \alpha^2\|\boldsymbol{s}\|^2$, we have from (120) that $\boldsymbol{u} - \alpha\boldsymbol{s} \in \mathcal{D}_n$ if and only if

$$-n\alpha^2 P_S - \delta_x \leq -2\alpha\langle\boldsymbol{s}, \boldsymbol{u}\rangle + \alpha^2\|\boldsymbol{s}\|^2 \leq -n\alpha^2 P_S, \tag{198}$$

or equivalently

$$\frac{n\alpha P_S}{2} \leq \langle\boldsymbol{s}, \boldsymbol{u}\rangle - \frac{\alpha}{2}\|\boldsymbol{s}\|^2 \leq \frac{n\alpha P_S}{2} + \frac{\delta_x}{2\alpha}. \tag{199}$$

By symmetry, the distribution of $\langle\boldsymbol{s}, \boldsymbol{U}\rangle$ depends on $\boldsymbol{s}$ only through its magnitude, and we can thus assume that $\boldsymbol{s} = (\|\boldsymbol{s}\|, 0, \cdots, 0)$. In this case, the condition in (199) becomes

$$\frac{n\alpha P_S}{2} \leq u_1\|\boldsymbol{s}\| - \frac{\alpha}{2}\|\boldsymbol{s}\|^2 \leq \frac{n\alpha P_S}{2} + \frac{\delta_x}{2\alpha}. \tag{200}$$

where u_1 is the first entry of $\boldsymbol{u}$. Adding $\frac{\alpha}{2}\|\boldsymbol{s}\|^2$ and dividing by $\|\boldsymbol{s}\|$, this becomes

$$\frac{n\alpha P_S}{2\|\boldsymbol{s}\|} + \frac{\alpha}{2}\|\boldsymbol{s}\| \leq u_1 \leq \frac{n\alpha P_S}{2\|\boldsymbol{s}\|} + \frac{\alpha}{2}\|\boldsymbol{s}\| + \frac{\delta_x}{2\alpha\|\boldsymbol{s}\|}. \tag{201}$$

From (108), there exists $P_{\max} < \infty$ such that $\|\boldsymbol{s}\| \leq \sqrt{nP_{\max}}$ whenever $P_S \in \tilde{\mathcal{P}}_n$. It follows that $\boldsymbol{u} - \alpha\boldsymbol{s} \in \mathcal{D}_n$ provided that

$$\frac{n\alpha P_S}{2\|\boldsymbol{s}\|} + \frac{\alpha}{2}\|\boldsymbol{s}\| \leq u_1 \leq \frac{n\alpha P_S}{2\|\boldsymbol{s}\|} + \frac{\alpha}{2}\|\boldsymbol{s}\| + \frac{\delta_x}{2\alpha\sqrt{nP_{\max}}}. \tag{202}$$

We conclude that $\mathbb{P}[\boldsymbol{U} - \alpha\boldsymbol{s} \in \mathcal{D}_n]$ is lower bounded by the probability of the first entry U_1 of $\boldsymbol{U}$ falling within an interval of length $\frac{c}{\sqrt{n}}$ starting at $\frac{n\alpha P_S}{2\|\boldsymbol{s}\|} + \frac{\alpha}{2}\|\boldsymbol{s}\|$, where $c \triangleq \frac{\delta_x}{2\alpha\sqrt{P_{\max}}}$. The distribution of a given symbol in a length-n random sequence distributed uniformly on the sphere is known [33, Eq. (4)], and yields

$$f_{U_1}(u_1) = \frac{1}{\sqrt{\pi n(P + \alpha^2 P_S)}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \left(1 - \frac{u_1^2}{n(P + \alpha^2 P_S)}\right)^{\frac{n-3}{2}} \mathbb{1}\big\{u_1^2 \leq n(P + \alpha^2 P_S)\big\}. \tag{203}$$

This density function is decreasing in u_1^2, which implies that

$$\mathbb{P}[\boldsymbol{U} - \alpha\boldsymbol{s} \in \mathcal{D}_n] \geq \frac{c}{\sqrt{n}} f_{U_1}\left(\frac{n\alpha P_S}{2\|\boldsymbol{s}\|} + \frac{\alpha}{2}\|\boldsymbol{s}\| + \frac{c}{\sqrt{n}}\right). \tag{204}$$

Furthermore, we have from (107) that $nP_S \leq \|\boldsymbol{s}\|^2$, and hence

$$\frac{n\alpha P_S}{2\|\boldsymbol{s}\|} + \frac{\alpha}{2}\|\boldsymbol{s}\| \leq \alpha\|\boldsymbol{s}\| \tag{205}$$

$$\leq \alpha\sqrt{nP_S} + \frac{\delta_s}{2\sqrt{nP_S}}, \tag{206}$$

where (206) follows by again using (107), along with the identity $\sqrt{1+\alpha} \le 1+\frac{\alpha}{2}$. Thus, the square of the argument to f_{U_1} in (204) is upper bounded by

$$\begin{aligned}
&\left(\alpha\sqrt{nP_S} + \frac{\delta_s}{2\sqrt{nP_S}} + \frac{c}{\sqrt{n}}\right)^2 \\
&\quad = n\alpha^2 P_S + 2\alpha\sqrt{nP_S}\left(\frac{\delta_s}{2\sqrt{nP_S}} + \frac{c}{\sqrt{n}}\right) + \left(\frac{\delta_s}{2\sqrt{nP_S}} + \frac{c}{\sqrt{n}}\right)^2 && (207)\\
&\quad \le n\alpha^2 P_S + \alpha\delta_s + 2\alpha\sqrt{P_{\max}}c + \left(\frac{\delta_s}{2\sqrt{nP_S}} + \frac{c}{\sqrt{n}}\right)^2 && (208)\\
&\quad \le n\alpha^2 P_S + c', && (209)
\end{aligned}$$

where (209) holds for any $c' > \alpha\delta_s + 2\alpha\sqrt{P_{\max}}c$ and sufficiently large n. Substituting (209) into (204) and again using the fact that $f_{U_1}(u_1)$ is decreasing in u_1^2, we obtain

$$\begin{aligned}
\mathbb{P}[\boldsymbol{U} - \alpha\boldsymbol{s} \in \mathcal{D}_n] &\ge \frac{1}{p_0'(n)}\left(1 - \frac{n\alpha^2 P_S + c'}{n(P+\alpha^2 P_S)}\right)^{\frac{n-3}{2}} && (210)\\
&= \frac{1}{p_0'(n)}\left(\frac{P}{P+\alpha^2 P_S}\left(1 - \frac{c'}{nP}\right)\right)^{\frac{n-3}{2}} && (211)\\
&\ge \frac{1}{p_0(n)}\left(\frac{P}{P+\alpha^2 P_S}\right)^{\frac{n}{2}}, && (212)
\end{aligned}$$

where $p_0'(n) \triangleq \left(\frac{c}{\sqrt{n}}\frac{1}{\sqrt{\pi n(P+\alpha^2 P_{\max})}}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}\right)^{-1}$ (which grows at most polynomially fast since $\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}$ grows as $\Theta(\sqrt{n})$), and (212) holds for some polynomial $p_0(n)$ and sufficiently large n since $\left(1+\frac{c'}{nP}\right)^{n/2} \to \exp\left(\frac{c'}{2P}\right)$. We obtain (128) by combining (212) with the definition of $I^{(P_S)}(U;S)$ in (116).

F. Proofs of Lemmas 4 and 5

We first introduce some results which will be used in both proofs. Recall that the information density $i^{(P_S)}(u,y)$ is defined with respect to the joint distribution $f_{SUY}^{(P_S)}$ defined in (110). The corresponding covariance matrix associated with (S,U,Y) is given by

$$\boldsymbol{V}_{SUY} = \begin{bmatrix} P_S & \alpha P_S & P_S \\ \alpha P_S & P+\alpha^2 P_S & P+\alpha P_S \\ P_S & P+\alpha P_S & P+P_S+1 \end{bmatrix}. \tag{213}$$

Substituting the (Gaussian) marginal distributions $f_{Y|U}^{(P_S)}$ and $f_Y^{(P_S)}$ into (122), it can be shown that

$$\begin{aligned}
i^{(P_S)}(u,y) = \frac{1}{2}\log\frac{(P+P_S+1)(P+\alpha^2 P_S)}{PP_S(1-\alpha)^2 + (P+\alpha^2 P_S)} \\
- \frac{P+\alpha^2 P_S}{2(PP_S(1-\alpha)^2 + (P+\alpha^2 P_S))}\left(y - \frac{P+\alpha P_S}{P+\alpha^2 P_S}u\right)^2 + \frac{y^2}{2(P+P_S+1)}.
\end{aligned} \tag{214}$$

Observe that the leading term coincides with the mutual information in (115).

Let $(\boldsymbol{s},\boldsymbol{u})$ be an arbitrary pair on the support of $f_{\boldsymbol{SU}}^{(P_S)}$, and recall that the definition of $f_{\boldsymbol{SU}}^{(P_S)}$ conditions on $\boldsymbol{S} \in T^n(P_S)$ and $\mathcal{E}_1^c$. It follows that $\|\boldsymbol{s}\|^2$ is bounded according to (107), and $\|\boldsymbol{x}\|^2 = \|\boldsymbol{u} - \alpha\boldsymbol{s}\|^2$ is bounded according to (120). It will prove useful to show that there exists a constant $\delta_{xs} > 0$ such that

$$n(P + P_S) - \delta_{xs} \leq \|\boldsymbol{u} + (1-\alpha)\boldsymbol{s}\|^2 \leq n(P + P_S) + \delta_{xs}. \tag{215}$$

To see this, we first combine (107) and (199) (the latter of which was derived using only (120) and the fact that $\|\boldsymbol{u}\|^2 = n(P + \alpha^2 P_S)$) to obtain

$$n\alpha P_S \leq \langle \boldsymbol{s}, \boldsymbol{u} \rangle \leq n\alpha P_S + \frac{\delta_x}{2\alpha} + \frac{\alpha\delta_s}{2}. \tag{216}$$

Writing $\|\boldsymbol{u}+(1-\alpha)\boldsymbol{s}\|^2 = \|\boldsymbol{u}\|^2+2(1-\alpha)\langle \boldsymbol{s},\boldsymbol{u}\rangle+(1-\alpha)^2\|\boldsymbol{s}\|^2$ and applying (107), (216) and $\|\boldsymbol{u}\|^2 = n(P+\alpha^2 P_S)$, we obtain (215).

1) Proof of Lemma 4 : To prove (134), we will show that conditioned on $(\boldsymbol{S}, \boldsymbol{U}) = (\boldsymbol{s}, \boldsymbol{u})$, the distribution of $i_n^{(P_S)}(\overline{\boldsymbol{U}}, \boldsymbol{Y})$ coincides with that of $i_n^{(P_S)}(\overline{\boldsymbol{U}}, \boldsymbol{Y}')$. Substituting $y \leftarrow y_i$ and $u \leftarrow u_i$ into (214) and summing from $i = 1$ to n, we see that $i_n^{(P_S)}(\boldsymbol{u}, \boldsymbol{y})$ depends on $(\boldsymbol{u}, \boldsymbol{y})$ only through $\|\boldsymbol{u}\|^2$, $\|\boldsymbol{y}\|^2$ and $\langle \boldsymbol{u}, \boldsymbol{y} \rangle$. Thus, since $\overline{\boldsymbol{U}}$ is circularly symmetric and has a fixed magnitude, the distribution of $i_n^{(P_S)}(\overline{\boldsymbol{U}}, \boldsymbol{y})$ depends on $\boldsymbol{y}$ only through $\|\boldsymbol{y}\|^2$. Writing $\boldsymbol{Y} = \boldsymbol{u} + (1-\alpha)\boldsymbol{s} + \boldsymbol{Z}$ and $\boldsymbol{Y}' = \boldsymbol{X}' + \boldsymbol{Z}$, we see that conditioned on $(\boldsymbol{s}, \boldsymbol{u})$, both $\boldsymbol{Y}$ and $\boldsymbol{Y}'$ are obtained by adding the i.i.d. Gaussian vector $\boldsymbol{Z}$ to a vector whose power is (almost surely or deterministically) equal to $\|\boldsymbol{u} + (1-\alpha)\boldsymbol{s}\|^2$. Thus, the conditional distribution of $\|\boldsymbol{Y}\|^2$ coincides with that of $\|\boldsymbol{Y}'\|^2$, and we obtain (134).

We now turn to the proof of (135)–(136). For the sake of notational brevity, we define $P_Y \triangleq P + P_S + 1$, and let $\mathcal{B}_\epsilon$ denote the set of sequences $\boldsymbol{y}'$ such that $\big|\|\boldsymbol{y}'\|^2 - nP_Y\big| \leq n\epsilon$. By definition, $(\boldsymbol{Y}'|\boldsymbol{s}, \boldsymbol{u})$ is obtained by adding i.i.d. Gaussian noise to $(\boldsymbol{X}'|\boldsymbol{s}, \boldsymbol{u})$, which in turn is uniform on a shell of power $n(P + P_S) + \eta$ for some $-\delta_{xs} \leq \eta \leq \delta_{xs}$ (see (132) and (215)). Defining $f_{Y,n}^{(P_S)} \sim N\big(0, P_Y + \frac{\eta}{n}\big)$, Step 1 of the proof of [2, Lemma 61] states there exists $\epsilon > 0$ such that[2]

$$f_{\boldsymbol{Y}'|\boldsymbol{S}\boldsymbol{U}}^{(P_S)}(\boldsymbol{y}'|\boldsymbol{s}, \boldsymbol{u}) \leq K_2' \prod_{i=1}^n f_{Y,n}^{(P_S)}(y_i') \tag{217}$$

for $\boldsymbol{y}' \in \mathcal{B}_\epsilon$, where K_2' is a constant depending on P_Y (see also [18, Prop. 2]). As noted in [2], we can choose $\epsilon = 1$ (or any $\epsilon \in (0, 1)$). The exponential decay in (135) follows from the Chernoff bound and the fact that $\mathbb{E}\big[\|\boldsymbol{Y}'\|^2 \,|\, \boldsymbol{s}, \boldsymbol{u}\big] = nP_Y + \eta$ [2, Eq. (417)].

To complete the proof of (136), we show that for $\boldsymbol{y}' \in \mathcal{B}_\epsilon$ we have $\prod_{i=1}^n f_{Y,n}^{(P_S)}(y_i') \leq K_2'' \prod_{i=1}^n f_Y^{(P_S)}(y_i')$ for some constant K_2''. We have from (114) that $f_Y^{(P_S)} \sim N(0, P_Y)$, and hence

$$\frac{\prod_{i=1}^n f_{Y,n}^{(P_S)}(y_i')}{\prod_{i=1}^n f_Y^{(P_S)}(y_i')} = \sqrt{\frac{nP_Y}{nP_Y + \eta}} \exp\left(- \frac{\|\boldsymbol{y}'\|^2}{2}\left(\frac{1}{nP_Y + \eta} - \frac{1}{nP_Y}\right)\right) \tag{218}$$

$$= \sqrt{\frac{1}{1 + \frac{\eta}{nP_Y}}} \exp\left(- \frac{\|\boldsymbol{y}'\|^2}{2nP_Y}\left(\frac{1}{1 + \frac{\eta}{nP_Y}} - 1\right)\right) \tag{219}$$

$$= \sqrt{\frac{1}{1 + \frac{\eta}{nP_Y}}} \exp\left(\frac{\|\boldsymbol{y}'\|^2\eta}{2(nP_Y)^2}\left(\frac{1}{1 + \frac{\eta}{nP_Y}}\right)\right) \tag{220}$$

[2]More precisely, it was shown in [2] that the ratio of the densities of the *norms* is upper bounded by a constant on $\mathcal{B}_\epsilon$. Since we are considering circularly symmetric distributions, this immediately implies that the same is true of the densities of the sequences themselves.

In the case that $\eta \in [-\delta_{xs}, 0)$, the desired result follows since the argument to $\exp(\cdot)$ in (220) is negative, and the subexponential prefactor tends to one. In the case that $\eta \in [0, \delta_{xs}]$, the bound $\|\boldsymbol{y}'\|^2 \le n(P_Y + \epsilon)$ (within $\mathcal{B}_\epsilon$) implies that the argument to $\exp(\cdot)$ in (220) is upper bounded by $\frac{(P_Y+\epsilon)^2\delta_{xs}}{2P_Y^2}$, which is a constant. It follows that (136) holds for all $\eta \in [-\delta_{xs}, \delta_{xs}]$, as desired.

Finally, the constants ψ and K_2 in (135)–(136) can be taken as independent of P_S due to the fact that P_S (and hence P_Y) is uniformly bounded within $\tilde{\mathcal{P}}_n$.

2) Proof of Lemma 5: The evaluation of the moments of the information density is cumbersome and similar to [34, Appendix A], so we omit some of the details.

We first consider the mean and variance. We write (214) as

$$i^{(P_S)}(u,y) = c_0 + c_1(y + c_2 u)^2 + c_3 y^2 \tag{221}$$

$$= c_0 + c_1 c_2^2 u^2 + 2c_1 c_2 uy + (c_1 + c_3)y^2, \tag{222}$$

where

$$c_0 \triangleq \frac{1}{2}\log\frac{(P + P_S + 1)(P + \alpha^2 P_S)}{PP_S(1-\alpha)^2 + (P + \alpha^2 P_S)} \tag{223}$$

$$c_1 \triangleq -\frac{P + \alpha^2 P_S}{2(PP_S(1-\alpha)^2 + (P + \alpha^2 P_S))} \tag{224}$$

$$c_2 \triangleq -\frac{P + \alpha P_S}{P + \alpha^2 P_S} \tag{225}$$

$$c_3 \triangleq \frac{1}{2(P + P_S + 1)}. \tag{226}$$

Substituting $y = u + (1-\alpha)s + z$ into (222), we obtain

$$i^{(P_S)}(u,y) = c_0 + d_1 s^2 + d_2 u^2 + d_3 z^2 + d_4 su + d_5 sz + d_6 uz, \tag{227}$$

where

$$d_1 \triangleq (c_1 + c_3)(1-\alpha)^2 \tag{228}$$

$$d_2 \triangleq c_1 c_2^2 + 2c_1 c_2 + (c_1 + c_3) \tag{229}$$

$$d_3 \triangleq c_1 + c_3 \tag{230}$$

$$d_4 \triangleq 2c_1 c_2(1-\alpha) + 2(c_1 + c_3)(1-\alpha) \tag{231}$$

$$d_5 \triangleq 2(c_1 + c_3)(1-\alpha) \tag{232}$$

$$d_6 \triangleq 2c_1 c_2 + 2(c_1 + c_3). \tag{233}$$

Letting $Y = u + (1-\alpha)s + Z$ (which follows by combining $Y = x + s + Z$ and $x = u - \alpha s$) and taking the mean and variance of (227) with respect to $Z \sim N(0,1)$ for fixed (s,u), we obtain

$$\mathbb{E}\Big[i^{(P_S)}(u,Y) \,|\, s,u\Big] = c_0 + d_1 s^2 + d_2 u^2 + d_3 + d_4 su \tag{234}$$

$$\mathrm{Var}\Big[i^{(P_S)}(u,Y) \,|\, s,u\Big] = \mathrm{Var}\Big[d_3 Z^2 + (d_5 s + d_6 u)Z\Big] \tag{235}$$

$$= 2d_3^2 + (d_5 s + d_6 u)^2, \tag{236}$$

where we have used $\mathbb{E}[Z]=0$, $\mathrm{Var}[Z]=1$, $\mathrm{Var}[Z^2]=2$ and $\mathrm{Cov}[Z^2,Z]=0$. It follows that

$$\mathbb{E}\Big[i_n^{(P_S)}(\boldsymbol{u},\boldsymbol{Y})\,|\,\boldsymbol{s},\boldsymbol{u}\Big] = nc_0 + d_1\|\boldsymbol{s}\|^2 + d_2\|\boldsymbol{u}\|^2 + nd_3 + d_4\langle\boldsymbol{s},\boldsymbol{u}\rangle \tag{237}$$

$$\mathrm{Var}\Big[i_n^{(P_S)}(\boldsymbol{u},\boldsymbol{Y})\,|\,\boldsymbol{s},\boldsymbol{u}\Big] = 2nd_3^2 + d_5^2\|\boldsymbol{s}\|^2 + 2d_5d_6\langle\boldsymbol{s},\boldsymbol{u}\rangle + d_6^2\|\boldsymbol{u}\|^2. \tag{238}$$

Using the definitions of the constants c_i and d_i, it can be verified from (237)–(238) that, for any $(\boldsymbol{s},\boldsymbol{u})$ such that $\|\boldsymbol{s}\|^2 = nP_S$, $\|\boldsymbol{u}\|^2 = n(P+\alpha^2P_S)$, and $\langle\boldsymbol{s},\boldsymbol{u}\rangle = n\alpha P_S$, we have

$$\mathbb{E}\big[i_n^{(P_S)}(\boldsymbol{u},\boldsymbol{Y})\,|\,\boldsymbol{s},\boldsymbol{u}\big] = nI^{(P_S)}(U;Y) \tag{239}$$

$$\begin{aligned}\mathrm{Var}\big[i_n^{(P_S)}(\boldsymbol{u},\boldsymbol{Y})\,|\,\boldsymbol{s},\boldsymbol{u}\big] &= \frac{1}{2(1+P+P_S)^2\big(PP_S(1-\alpha)^2+P+\alpha^2P_S\big)^2}\Bigg((P+\alpha P_S)^2\bigg(\alpha^2P_S(2+P_S) \\ &\quad + P^2\Big(1+2(1-\alpha)^2P_S\Big) + 2P\Big(1+\alpha P_S + P_S(1-\alpha)^2(2+P_S)\Big)\bigg)\Bigg) \qquad (240) \\ &\triangleq nV(P_S). \qquad (241)\end{aligned}$$

The distribution of $(\boldsymbol{S},\boldsymbol{U})$ under consideration does not ensure that the equalities $\|\boldsymbol{u}\|^2 = n(P+\alpha^2P_S)$ and $\langle\boldsymbol{s},\boldsymbol{u}\rangle = n\alpha P_S$ hold. However, they do hold to within an additive $O(1)$ term; see (107) and (216). Since the right-hand sides of (237)–(238) are linear in $\|\boldsymbol{s}\|^2$ and $\langle\boldsymbol{s},\boldsymbol{u}\rangle$, we conclude that (239) and (241) hold for all $(\boldsymbol{s},\boldsymbol{u})$ on the support of $f_{\boldsymbol{SU}}^{(P_S)}$ upon adding $O(1)$ to the right-hand sides. A direct substitution of $\alpha=\frac{P}{1+P}$ reveals that $V(P_S)=\frac{P(2+P)}{2(1+P)^2}$ for all P_S, and we have thus proved (143)–(144).

It remains to prove (145). To this end, we use (227) to write $i_n^{(P_S)}(\boldsymbol{u},\boldsymbol{Y})$ (given $(\boldsymbol{s},\boldsymbol{u})$) as

$$i_n^{(P_S)}(u,y) = nd_0 + d_1\|\boldsymbol{s}\|^2 + d_2\|\boldsymbol{u}\|^2 + d_3\|\boldsymbol{Z}\|^2 + d_4\langle\boldsymbol{s},\boldsymbol{u}\rangle + d_5\langle\boldsymbol{s},\boldsymbol{Z}\rangle + d_6\langle\boldsymbol{u},\boldsymbol{Z}\rangle. \tag{242}$$

Since $\boldsymbol{Z}$ is i.i.d. Gaussian, the distributions of the last two terms only depend on $\|\boldsymbol{s}\|^2$ and $\|\boldsymbol{u}\|^2$. Thus, the statistics of $i_n^{(P_S)}(\boldsymbol{U},\boldsymbol{Y})$ given $(\boldsymbol{S},\boldsymbol{U})=(\boldsymbol{s},\boldsymbol{u})$ depends on $(\boldsymbol{s},\boldsymbol{u})$ only through $\|\boldsymbol{u}\|^2$, $\|\boldsymbol{s}\|^2$ and $\langle\boldsymbol{s},\boldsymbol{u}\rangle$. We thus obtain (145) in the same way as [34, Appendix A] by choosing $(\boldsymbol{s}',\boldsymbol{u}')$ to attain the same powers and correlation as $(\boldsymbol{s},\boldsymbol{u})$ (thus yielding the same statistics of $i_n^{(P_S)}$), while having entries which are uniformly bounded for all n.

ACKNOWLEDGMENT

I would like to thank Vincent Tan for many helpful suggestions which greatly helped in making the proofs more rigorous and complete. In particular, Corollary 1 was pointed out by him.

REFERENCES

[1] V. Strassen, "Asymptotische Abschätzungen in Shannon's Informationstheorie," in *Trans. 3rd Prague Conf. on Inf. Theory*, 1962, pp. 689–723, English Translation: http://www.math.wustl.edu/~luthy/strassen.pdf.

[2] Y. Polyanskiy, V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

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

[4] S. I. Gelfand and M. S. Pinsker, "Coding for channel with random parameters," *Prob. Inf. Transm.*, vol. 9, no. 1, pp. 19–31, 1980.

[5] M. Costa, “Writing on dirty paper,” *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, 1983.

[6] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[7] Y. Polyanskiy, “Channel coding: Non-asymptotic fundamental limits,” Ph.D. dissertation, Princeton University, 2010.

[8] U. Erez and R. Zamir, “Achieving 1/2 log (1+SNR) on the AWGN channel with lattice encoding and decoding,” *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2293–2314, 2004.

[9] T. Liu, P. Moulin, and R. Koetter, “On error exponents of modulo lattice additive noise channels,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 454–471, 2006.

[10] J. Jiang and T. Liu, “On dispersion of modulo lattice additive noise channels,” in *Int. Symp. Wireless. Comm. Sys.*, 2011, pp. 241–245.

[11] A. Sarwate and M. Gastpar, “Arbitrarily dirty paper coding and applications,” in *IEEE Int. Symp. Inf. Theory*, 2008, pp. 925–929.

[12] S. Watanabe, S. Kuzuoka, and V. Y. F. Tan, “Non-asymptotic and second-order achievability bounds for coding with side-information,” 2013, http://arxiv.org/abs/1301.6467.

[13] M. H. Yassaee, M. R. Aref, and A. Gohari, “A technique for deriving one-shot achievability results in network information theory,” http://arxiv.org/abs/1303.0696.

[14] A. Somekh-Baruch and N. Merhav, “On the random coding error exponents of the single-user and the multiple-access Gel’fand-Pinsker channels,” in *IEEE Int. Symp. Inf. Theory*, Chicago, IL, June 2004.

[15] P. Moulin and Y. Wang, “Capacity and random-coding exponents for channel coding with side information,” *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1326–1347, 2007.

[16] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, “Second-order rate region of constant-composition codes for the multiple-access channel,” 2013, http://arxiv.org/abs/1303.6167.

[17] M. H. Yassaee, M. R. Aref, and A. Gohari, “Non-asymptotic output statistics of random binning and its applications,” http://arxiv.org/abs/1303.0695.

[18] E. MolavianJazi and J. N. Laneman, “A finite-blocklength perspective on gaussian multi-access channels,” arXiv:1309.2343v1 [cs.IT].

[19] M. Tomamichel and V. Y. F. Tan, “ϵ-capacities and second-order coding rates for channels with general state,” [Online: http://arxiv.org/abs/1305.6789].

[20] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.

[21] R. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.

[22] ——, “Fixed composition arguments and lower bounds to error probability,” http://web.mit.edu/gallager/www/notes/notes5.pdf.

[23] A. Martinez and A. Guillén i Fàbregas, “Random-coding bounds for threshold decoders: Error exponent and saddlepoint approximation,” in *Int. Symp. Inf. Theory*, St. Petersburg, Aug. 2011.

[24] G. Poltyrev, “Random coding bounds for discrete memoryless channels,” *Prob. Inf. Transm.*, vol. 18, no. 1, pp. 9–21, 1982.

[25] W. Feller, *An introduction to probability theory and its applications*, 2nd ed. John Wiley & Sons, 1971, vol. 2.

[26] V. Y. F. Tan and O. Kosut, “On the dispersions of three network information theory problems,” 2012, arXiv:1201.3901v2 [cs.IT].

[27] N. Merhav, “Universal decoding for memoryless gaussian channels with a deterministic interference,” *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1261–1269, 1993.

[28] T. Thomas and B. Hughes, “Exponential error bounds for random codes on Gaussian arbitrarily varying channels,” *IEEE Trans. Inf. Theory*, vol. 37, no. 3, pp. 643–649, 1991.

[29] H. Tyagi and P. Narayan, “The Gelfand-Pinsker channel: Strong converse and upper bound for the reliability function,” in *IEEE Int. Symp. Inf. Theory*, 2009, pp. 1954–1957.

[30] P. Moulin, “Strong converse for Gel’fand-Pinsker channel,” in *IEEE Int. Symp. Inf. Theory*, 2009, pp. 1958–1962.

[31] Y. Polyanskiy, “On dispersion of compound DMCs,” in *Allerton Conf. on Comm., Control and Comp.*, Monticello, IL, 2013.

[32] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[33] A. J. Stam, “Limit theorems for uniform distributions on spheres in high-dimensional euclidean spaces,” *Journal of Applied Probability*, vol. 19, no. 1, pp. 221–228, March 1982.

[34] J. Scarlett and V. Y. F. Tan, “Second-order asymptotics for the Gaussian MAC with degraded message sets,” 2013, submitted to *IEEE Trans. Inf. Theory* [Online: http://arxiv.org/abs/1310.1197].