On The Dispersion of Dirty Paper Coding

Jonathan Scarlett

***Abstract*—This paper studies the second-order asymptotics of the coding rate for a given error probability in the setting of dirty paper coding (Costa, 1983) with an almost-sure power constraint. It is shown that the dispersion is the same as if the state sequence were absent, thus strengthening the analogous capacity result. The result holds under mild technical conditions on the state sequence, and is not limited to the ergodic case.**

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 regained significant attention following the works of Polyanskiy *et al.* [2] and Hayashi [3]. For the additive white Gaussian noise (AWGN) channel with maximal power P, the highest number of codewords $M^*(n,\epsilon)$ of length n yielding an average error probability $\epsilon \in (0,1)$ satisfies [2]

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

where $C = \frac{1}{2}\log(1+P)$ is the channel capacity, $\mathsf{Q}^{-1}(\cdot)$ is the inverse of the Q-function, and $V = \frac{P(2+P)}{2(1+P)^2}$ is known as the *channel dispersion*.

In this paper, we consider the problem of dirty paper coding [4], in which the channel is described by

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{S} + \boldsymbol{Z}, \tag{2}$$

where $\boldsymbol{Z}$ is an i.i.d. noise sequence with $Z_i \sim N(0,1)$, $\boldsymbol{S}$ is a random state sequence known non-causally at the encoder, and the input is subject to a power constraint with power P:

$$\|\boldsymbol{X}\|^2 \le nP \qquad \text{a.s.}, \tag{3}$$

where n is the block length. It is well known that the capacity is equal to that of the AWGN capacity (i.e. $C = \frac{1}{2}\log(1+P)$) [5]. Our main result extends this result to second-order asymptotics, showing that the expansion (1) holds. Moreover, we consider state sequences $\boldsymbol{S}$ which need not be ergodic. Our analysis can be considered an extension of that of [6], which studies the discrete memoryless counterpart of the present problem (i.e. the Gel'fand-Pinsker channel [7]).

A similar second-order result was given in [8] using lattice coding with common randomness in the form of a dither, and a power constraint averaged over the dither. In this setting, one can handle *arbitrary* state distributions [9]. Our setting is stricter, in that we assume an almost-sure power constraint with no common randomness. The removal of these assumptions for lattice coding may be possible, but in any case, our techniques are significantly different and provide a new approach. For other related works, see [10]–[12].

Capacity-Achieving Parameters with Gaussian State : Here we provide an outline of the capacity results for the case that $P_{\boldsymbol{S}}$ is i.i.d. on $\pi \sim N(0,P_\pi)$ for some $P_\pi > 0$; see [4] for details. Although we consider significantly more general state sequences, the choices of parameters given here will play a major role in our analysis.

Analogously to the Gel'fand-Pinsker channel [7], the capacity can be written as

$$C = \max_{\mathcal{U}, Q_{U|S}, \phi(\cdot,\cdot)} I(U;Y) - I(U;S) \tag{4}$$

$$P_{SUY}(s,u,y) = \pi(s) Q_{U|S}(u|s) W(y|\phi(u,s),s), \tag{5}$$

where $W(y|x,s) \triangleq \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-x-s)^2}{2}}$, and the maximization is subject to the constraint $\mathbb{E}[\phi(U,S)^2] \le P$. We fix $\alpha > 0$ and choose $Q_{U|S}(\cdot|s) \sim N(-\alpha s, P)$ and $\phi(u,s) = u - \alpha s$, which can be equivalently be written as

$$U = X + \alpha S \tag{6}$$

$$X \sim N(0,P), \tag{7}$$

where X is independent of S. Under these parameters, the mutual informations in (4) are given by

$$I^{(P_\pi)}(U;Y) \triangleq \frac{1}{2}\log\left(\frac{(P+P_\pi+1)(P+\alpha^2 P_\pi)}{PP_\pi(1-\alpha)^2 + (P+\alpha^2 P_\pi)}\right) \tag{8}$$

$$I^{(P_\pi)}(U;S) \triangleq \frac{1}{2}\log\left(\frac{P+\alpha^2 P_\pi}{P}\right). \tag{9}$$

Furthermore, the optimal choice of α is given by $\alpha = \frac{P}{1+P}$, and yields $C = \frac{1}{2}\log(1+P)$, as desired.

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 $\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle = \sum_i x_{1,i} x_{2,i}$ and the ℓ_2-norm $\|\boldsymbol{x}_1\| = \sqrt{\langle \boldsymbol{x}_1, \boldsymbol{x}_1 \rangle}$. The marginals of a joint distribution P_{XY} are denoted by P_X and P_Y. When the meaning is clear, we will use shorthands such as $\mathbb{P}[\cdot \,|\, x]$ to denote conditioning on events such as $X = x$. We make use of the standard asymptotic notations $O(\cdot)$, $o(\cdot)$ and $\Theta(\cdot)$.

II. Main Result

Theorem 1. *For the dirty paper coding setup with an arbitrary sequence of state distributions $P_{\boldsymbol{S}}$ (indexed by n) such that*

$$\mathbb{P}\big[\|\boldsymbol{S}\|^2 > n\Pi\big] = O\Big(\frac{\log n}{\sqrt{n}}\Big) \tag{10}$$

J. Scarlett is with the Department of Engineering, University of Cambridge, United Kingdom (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.

for some $\Pi < \infty$*, the expansion in* (1) *holds for* $\epsilon \in (0,1)$ *with capacity* $C = \frac{1}{2}\log(1+P)$ *and dispersion* $V = \frac{P(2+P)}{2(1+P)^2}$.

Proof: The converse part follows by revealing the state sequence to the decoder and using the converse result of [2]. The achievability part is proved in Section III. ∎

The assumption in (10) 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. In the case of an i.i.d. state with $S_i \sim \pi$, Chebyshev's inequality reveals that a sufficient condition for (10) to hold is that $\mathbb{E}_\pi[S^4] < \infty$.

The intuition behind allowing for such general state sequence distributions is as follows: We use an asymptotically negligible fraction of the block to inform the decoder 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.

Our random-coding scheme bears some similarities to the Gel'fand-Pinsker coding scheme [7], but uses different codebooks depending on which spherical shell the state sequence lies within, as well as refined notions of "typicality".

III. Proof

Due to space constraints, some details of the proof are omitted, and can be found in [13].

A. Preliminary Definitions and Results

1) Power Types: We make use of *power types* (e.g. see [14]). 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

$$T^n(P_S) \triangleq \Big\{\boldsymbol{s} \,:\, nP_S \le \|\boldsymbol{s}\|^2 < nP_S + \delta_s\Big\}. \tag{11}$$

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, we can define a typical set of state types as follows:

$$\tilde{\mathcal{P}}_n \triangleq \big\{P_S \in \mathcal{P}_n \,:\, P_S \le \Pi\big\}. \tag{12}$$

where Π appears in (10). We obtain from (10) that

$$\mathbb{P}\big[\hat{P}_{\boldsymbol{S}} \notin \tilde{\mathcal{P}}_n\big] = O\Big(\frac{\log n}{\sqrt{n}}\Big). \tag{13}$$

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

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

Theorem 2. *The statement of Theorem 1 holds true in the case that the type* P_S *of* $\boldsymbol{S}$ *is known at the decoder.*

We proceed by showing that Theorem 2 implies Theorem 1. The idea is to treat the event $\hat{P}_{\boldsymbol{S}} \notin \tilde{\mathcal{P}}_n$ as an error and transmit one of the remaining $\Theta(n)$ types to the receiver in $O(\log n)$ channel uses. Using the arguments of [6], this can be done without affecting the second-asymptotics provided that the random-coding error exponent is positive for sufficiently small rates. From [15, 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 $\|\boldsymbol{S}\|^2 \le nP_{\max}$. Since we have treated the event $\hat{P}_{\boldsymbol{S}} \notin \tilde{\mathcal{P}}_n$ as an error, it follows from (12) 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 (e.g. see [2]) defined with respect to the joint distribution

$$f_{SUY}^{(P_S)}(s,u,y) = f_S^{(P_S)}(s) Q_{U|S}(u|s) f_{Y|SU}(y|s,u), \tag{14}$$

where in accordance with the parameters in Section I, we have $f_S^{(P_S)} \sim N(0, P_S)$, $Q_{U|S} \sim N(-\alpha S, P)$ and $f_{Y|SU} \sim N\big(U + (1-\alpha)S, 1\big)$. The parameter $\alpha > 0$ is arbitrary for now. The induced output distribution is given by $f_Y^{(P_S)} \sim N(0, P + P_S + 1)$, and the corresponding mutual informations $I^{(P_S)}(U;Y)$ and $I^{(P_S)}(U;S)$ coincide with (8)–(9).

B. Proof of Theorem 1

As stated previously, in order to prove Theorem 1, it suffices to prove Theorem 2. Thus, we henceforth assume that the state type P_S is known at the decoder.

1) Codebook Generation: 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 $L^{(P_S)}$. 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 $n(P + \alpha^2 P_S)$:

$$f_{\boldsymbol{U}}^{(P_S)}(\boldsymbol{u}) = \frac{\delta\big(\|\boldsymbol{u}\|^2 - n(P+\alpha^2 P_S)\big)}{S_n\big(\sqrt{n(P+\alpha^2 P_S)}\big)}, \tag{15}$$

where $\delta(\cdot)$ is the Dirac delta function, and $S_n(r) = \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} r^{n-1}$ is the surface area of a sphere of radius r in n dimensions.

2) Encoding and Decoding: Given the state sequence $\boldsymbol{S} \in T^n(P_S)$ and message m, the encoder sends $\boldsymbol{X} = \boldsymbol{U} - \alpha\boldsymbol{S}$, 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

$$\mathcal{D}_n \triangleq \Big\{\boldsymbol{x} \,:\, nP - \delta_x \le \|\boldsymbol{x}\|^2 \le nP\Big\} \tag{16}$$

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 (3) is satisfied.

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

$$i_n^{(P_S)}(\boldsymbol{u}, \boldsymbol{y}) \triangleq \sum_{i=1}^n i^{(P_S)}(u_i, y_i) \tag{17}$$

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

$$i^{(P_S)}(u,y) \triangleq \log \frac{f_{Y|U}^{(P_S)}(y|u)}{f_Y^{(P_S)}(y)} \tag{18}$$

with $f_{SUY}^{(P_S)}$ defined in (14). We consider the events

$$\mathcal{E}_1 \triangleq \big\{\text{No } l \text{ exists with } \boldsymbol{U}^{(P_S)}(m,l) - \alpha \boldsymbol{S} \in \mathcal{D}_n\big\} \tag{19}$$

$$\mathcal{E}_2 \triangleq \big\{\text{Decoder chooses a message } \tilde{m} \neq m\big\}. \tag{20}$$

It follows from these definitions and (13) that the overall random-coding error probability $\overline{p}_e$ satisfies

$$\overline{p}_e \leq \sum_{P_S \in \tilde{\mathcal{P}}_n} \mathbb{P}\big[P_S\big]\Big(\mathbb{P}\big[\mathcal{E}_1 \,|\, P_S\big] + \mathbb{P}\big[\mathcal{E}_2 \,|\, P_S, \mathcal{E}_1^c\big]\Big) + O\Big(\frac{\log n}{\sqrt{n}}\Big). \tag{21}$$

3) 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 (9). 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

$$R_L^{(P_S)} = I^{(P_S)}(U;S) + K_1 \frac{\log n}{n} \tag{22}$$

while achieving

$$\mathbb{P}\big[\mathcal{E}_1 \,|\, P_S\big] \leq e^{-\psi n} \tag{23}$$

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

Lemma 1. *Fix $P_S \in \tilde{\mathcal{P}}_n$, and let $\boldsymbol{U}$ have density $f_{\boldsymbol{U}}^{(P_S)}$ (see (15)). For all $\boldsymbol{s} \in T^n(P_S)$ and sufficiently large n, we have*

$$\mathbb{P}\big[\boldsymbol{U} - \alpha \boldsymbol{s} \in \mathcal{D}_n\big] \geq \frac{1}{p_0(n)} e^{-I^{(P_S)}(U;Y)}, \tag{24}$$

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

Proof: See Appendix A. ∎

We obtain (22)–(23) using Lemma 1 and following identical steps to the discrete case [6].

4) 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. $(\boldsymbol{S},\boldsymbol{U},\boldsymbol{Y}) \sim f_{\boldsymbol{SU}}^{(P_S)}(\boldsymbol{s},\boldsymbol{u}) W^n(\boldsymbol{y}|\boldsymbol{u}-\alpha\boldsymbol{s},\boldsymbol{s})$, where W^n is i.i.d. on $W(y|x,s)$ (defined following (5)). We make use of the following standard threshold-based bound (e.g. see [2]):

$$\begin{aligned}\mathbb{P}\big[\mathcal{E}_2 \,|\, P_S, \mathcal{E}_1^c\big] \leq{}& \mathbb{P}\Big[i_n^{(P_S)}(\boldsymbol{U},\boldsymbol{Y}) \leq \gamma^{(P_S)}\Big] \\ &+ ML^{(P_S)}\mathbb{P}\Big[i_n^{(P_S)}(\overline{\boldsymbol{U}},\boldsymbol{Y}) > \gamma^{(P_S)}\Big]\end{aligned} \tag{25}$$

for any $\gamma^{(P_S)}$, where $\overline{\boldsymbol{U}} \sim f_{\boldsymbol{U}}^{(P_S)}$ is independent of $(\boldsymbol{S},\boldsymbol{U},\boldsymbol{Y})$.

Lemma 2. *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*

$$(\boldsymbol{X}' \,|\, \boldsymbol{s},\boldsymbol{u}) \sim \frac{\delta\big(\|\boldsymbol{x}'\| - \|\boldsymbol{u} + (1-\alpha)\boldsymbol{s}\|\big)}{S_n\big(\|\boldsymbol{u} + (1-\alpha)\boldsymbol{s}\|\big)} \tag{26}$$

and $\boldsymbol{Y}' = \boldsymbol{X}' + \boldsymbol{Z}$, where $\delta(\cdot)$ and $S_n(\cdot)$ are defined following (15), and $\boldsymbol{Z}$ is the additive noise in (2). For $\overline{\boldsymbol{U}} \sim f_{\boldsymbol{U}}^{(P_S)}$ independent of $(\boldsymbol{S},\boldsymbol{U},\boldsymbol{Y},\boldsymbol{X}',\boldsymbol{Y}')$, we have

$$\mathbb{P}\big[i_n^{(P_S)}(\overline{\boldsymbol{U}},\boldsymbol{Y}) > \gamma^{(P_S)}\big|\boldsymbol{s},\boldsymbol{u}\big] = \mathbb{P}\big[i_n^{(P_S)}(\overline{\boldsymbol{U}},\boldsymbol{Y}') > \gamma^{(P_S)}\big|\boldsymbol{s},\boldsymbol{u}\big]. \tag{27}$$

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

$$\mathbb{P}\Big[\big|\|\boldsymbol{Y}'\|^2 - n(P+P_S+1)\big| > n\epsilon \,\Big|\, \boldsymbol{s},\boldsymbol{u}\Big] \leq e^{-\psi n} \tag{28}$$

$$\min_{\boldsymbol{y}' \,:\, \big|\|\boldsymbol{y}'\|^2 - n(P+P_S+1)\big| \leq n\epsilon} \frac{f_{\boldsymbol{Y}'|\boldsymbol{SU}}^{(P_S)}(\boldsymbol{y}'|\boldsymbol{s},\boldsymbol{u})}{\prod_{i=1}^n f_Y^{(P_S)}(y_i')} \leq K_2 \tag{29}$$

for sufficiently large n and constants $\psi > 0$ and K_2 not depending on P_S.

Proof: See Appendix B. ∎

Using Lemma 2, we can apply a standard change of measure to the i.i.d. output distribution (e.g. see [3]) to deduce that

$$\mathbb{P}\Big[i_n^{(P_S)}(\overline{\boldsymbol{U}},\boldsymbol{Y}) > \gamma^{(P_S)} \,\Big|\, \boldsymbol{s},\boldsymbol{u}\Big] \leq K_2 e^{-\gamma^{(P_S)}} + e^{-\psi n}. \tag{30}$$

We choose $\gamma^{(P_S)} = \log M + nI^{(P_S)}(U;S) + \log n$, which, when combined with (22), yields

$$\gamma^{(P_S)} = \log M + nI^{(P_S)}(U;S) + K_3 \log n \tag{31}$$

with $K_3 \triangleq K_1 + 1$. Combining (25) and (30) with this choice of $\gamma^{(P_S)}$, we conclude that

$$\begin{aligned}\mathbb{P}\big[\mathcal{E}_2 \,|\, P_S, \mathcal{E}_1^c\big] \leq{}& \mathbb{P}\Big[i_n^{(P_S)}(\boldsymbol{u},\boldsymbol{Y}) \leq \log M \\ &+ nI^{(P_S)}(U;S) + K_3 \log n \,\Big|\, \boldsymbol{s},\boldsymbol{u}\Big] + O\Big(\frac{1}{n}\Big)\end{aligned} \tag{32}$$

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

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

$$\mathbb{E}\big[i_n^{(P_S)}(\boldsymbol{u},\boldsymbol{Y}) \,|\, \boldsymbol{s},\boldsymbol{u}\big] = nI^{(P_S)}(U;Y) + O(1), \tag{33}$$

and under the choice $\alpha = \frac{P}{1+P}$, we have

$$\mathrm{Var}\big[i_n^{(P_S)}(\boldsymbol{u},\boldsymbol{Y}) \,|\, \boldsymbol{s},\boldsymbol{u}\big] = nV + O(1), \tag{34}$$

where $V = \frac{P(2+P)}{2(1+P)^2}$. Furthermore, there exists $(\boldsymbol{s}',\boldsymbol{u}')$ 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

$$\sum_{i=1}^n \mathbb{E}\Big[\Big|i^{(P_S)}(u_i',Y_i) - \mathbb{E}\big[i^{(P_S)}(u_i',Y_i)\big]\Big|^3 \,\Big|\, s_i',u_i'\Big] = O(n). \tag{35}$$

The remainder terms in (33)–(35) are uniform in $P_S \in \tilde{\mathcal{P}}_n$.

Proof: See Appendix B. ∎

Combining (23) and (32), we have for some $(\boldsymbol{s},\boldsymbol{u})$ that

$$\mathbb{P}\big[\mathcal{E}_1 \cup \mathcal{E}_2 \,|\, P_S\big] \le \mathbb{P}\Big[i_n^{(P_S)}(\boldsymbol{u},\boldsymbol{Y}) \le \log M + nI^{(P_S)}(U;S) + K_3 \log n \,\Big|\, \boldsymbol{s},\boldsymbol{u}\Big] + O\Big(\frac{1}{n}\Big) \quad (36)$$

As stated in Section I, setting $\alpha = \frac{P}{1+P}$ yields $I^{(P_S)}(U;Y) - I^{(P_S)}(U;S) = C$ for all P_S. Thus, using Lemma 3 and applying the Berry-Esseen theorem [16, Sec. XVI.5] to (36) (after replacing $(\boldsymbol{s},\boldsymbol{u})$ by $(\boldsymbol{s}',\boldsymbol{u}')$ given in the lemma statement if necessary), we obtain for all $P_S \in \tilde{\mathcal{P}}_n$ that

$$\mathbb{P}\big[\mathcal{E}_1 \cup \mathcal{E}_2 | P_S\big] \le \mathsf{Q}\bigg(\frac{\log M - nC + K_3 \log n + K_4}{\sqrt{nV + K_5}}\bigg) + O\Big(\frac{1}{\sqrt{n}}\Big) \quad (37)$$

for some constants K_4 and K_5. Substituting (37) into (21) and inverting the relationship between the error probability and the number of messages, we obtain the desired result.

APPENDIX

A. Proof of Lemma 1

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 $\|\boldsymbol{u} - \alpha\boldsymbol{s}\|^2 = \|\boldsymbol{u}\|^2 - 2\alpha\langle \boldsymbol{s},\boldsymbol{u}\rangle + \alpha^2\|\boldsymbol{s}\|^2$ and using (16), it follows that $\boldsymbol{u} - \alpha\boldsymbol{s} \in \mathcal{D}_n$ if and only if

$$\frac{n\alpha P_S}{2} \le \langle \boldsymbol{s},\boldsymbol{u}\rangle - \frac{\alpha}{2}\|\boldsymbol{s}\|^2 \le \frac{n\alpha P_S}{2} + \frac{\delta_x}{2\alpha}. \quad (38)$$

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 $\boldsymbol{s} = (\|\boldsymbol{s}\|, 0, \cdots, 0)$. Under this choice, (38) becomes[1]

$$\frac{n\alpha P_S}{2\|\boldsymbol{s}\|} + \frac{\alpha}{2}\|\boldsymbol{s}\| \le u_1 \le \frac{n\alpha P_S}{2\|\boldsymbol{s}\|} + \frac{\alpha}{2}\|\boldsymbol{s}\| + \frac{\delta_x}{2\alpha\|\boldsymbol{s}\|}. \quad (39)$$

From (12), there exists $P_{\max} < \infty$ such that $\|\boldsymbol{s}\| \le \sqrt{nP_{\max}}$ whenever $P_S \in \tilde{\mathcal{P}}_n$. It follows that (39) remains a sufficient condition for $\boldsymbol{u} - \alpha\boldsymbol{s} \in \mathcal{D}_n$ when the final term is replaced by $\frac{c}{\sqrt{n}}$, where $c \triangleq \frac{\delta_x}{2\alpha\sqrt{P_{\max}}}$. That is, $\mathbb{P}[\boldsymbol{U} - \alpha\boldsymbol{s} \in \mathcal{D}_n]$ 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 $\frac{n\alpha P_S}{2\|\boldsymbol{s}\|} + \frac{\alpha}{2}\|\boldsymbol{s}\|$. The distribution of a given symbol in a length-n random sequence distributed uniformly on the sphere is known [17, Eq. (4)]:

$$f_{U_1}(u_1) = \frac{1}{\sqrt{\pi n(P+\alpha^2 P_S)}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \times \bigg(1 - \frac{u_1^2}{n(P+\alpha^2 P_S)}\bigg)^{\frac{n-3}{2}} \mathbb{1}\big\{u_1^2 \le n(P+\alpha^2 P_S)\big\}. \quad (40)$$

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

$$\mathbb{P}[\boldsymbol{U} - \alpha\boldsymbol{s} \in \mathcal{D}_n] \ge \frac{c}{\sqrt{n}} f_{U_1}\bigg(\frac{n\alpha P_S}{2\|\boldsymbol{s}\|} + \frac{\alpha}{2}\|\boldsymbol{s}\| + \frac{c}{\sqrt{n}}\bigg). \quad (41)$$

Furthermore, we have from (11) that $nP_S \le \|\boldsymbol{s}\|^2$. Combining this with (11) and $\sqrt{1+\alpha} \le 1 + \frac{\alpha}{2}$, it is easily shown that

$$\frac{n\alpha P_S}{2\|\boldsymbol{s}\|} + \frac{\alpha}{2}\|\boldsymbol{s}\| \le \alpha\sqrt{nP_S} + \frac{\delta_s}{2\sqrt{nP_S}}. \quad (42)$$

[1]If $\|\boldsymbol{s}\| = 0$ then it is understood that $\frac{P_S}{\|\boldsymbol{s}\|} = 0$ and $\frac{\delta_x}{2\alpha\|\boldsymbol{s}\|} = \infty$.

Thus, the square of the argument to f_{U_1} in (41) is upper bounded by $\big(\alpha\sqrt{nP_S} + \frac{\delta_s}{2\sqrt{nP_S}} + \frac{c}{\sqrt{n}}\big)^2$, which can further be upper bounded by $n\alpha^2 P_S + c'$ (for any $c' > \alpha\delta_s + 2\alpha\sqrt{P_{\max}}c$) for sufficiently large n by expanding the square and applying simple bounding techniques. Substituting this into (41) and again using the fact that $f_{U_1}(u_1)$ is decreasing in u_1^2, we obtain

$$\mathbb{P}[\boldsymbol{U} - \alpha\boldsymbol{s} \in \mathcal{D}_n] \ge \frac{1}{p_0'(n)}\bigg(1 - \frac{n\alpha^2 P_S + c'}{n(P+\alpha^2 P_S)}\bigg)^{\frac{n-3}{2}} \quad (43)$$

where we define $p_0'(n) \triangleq \Big(\frac{c}{\sqrt{n}} \frac{1}{\sqrt{\pi n(P+\alpha^2 P_{\max})}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}\Big)^{-1}$, which grows at most polynomially fast since $\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}$ grows as $\Theta(\sqrt{n})$. The lemma follows by factoring the constant c' into the polynomial prefactor using $\big(1 + \frac{\alpha}{n}\big)^{n/2} \to \exp\big(\frac{\alpha}{2}\big)$, and then using $1 - \frac{\alpha^2 P_S}{P+\alpha^2 P_S} = \frac{P}{P+\alpha^2 P_S}$ and (9).

B. Proofs of Lemmas 2 and 3

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 (14). Substituting the (Gaussian) marginal distributions $f_{Y|U}^{(P_S)}$ and $f_Y^{(P_S)}$ into (18), it can be shown that

$$i^{(P_S)}(u,y) = \frac{1}{2}\log\frac{(P+P_S+1)(P+\alpha^2 P_S)}{PP_S(1-\alpha)^2 + (P+\alpha^2 P_S)} - \frac{P+\alpha^2 P_S}{2(PP_S(1-\alpha)^2 + (P+\alpha^2 P_S))}\bigg(y - \frac{P+\alpha P_S}{P+\alpha^2 P_S}u\bigg)^2 + \frac{y^2}{2(P+P_S+1)}. \quad (44)$$

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 (11), and $\|\boldsymbol{x}\|^2 = \|\boldsymbol{u} - \alpha\boldsymbol{s}\|^2$ is bounded according to (16). It will prove useful to show that there exists a constant $\delta_{xs} > 0$ such that

$$n(P+P_S) - \delta_{xs} \le \|\boldsymbol{u} + (1-\alpha)\boldsymbol{s}\|^2 \le n(P+P_S) + \delta_{xs}. \quad (45)$$

To see this, we first combine (11) and (38) to obtain

$$n\alpha P_S \le \langle \boldsymbol{s},\boldsymbol{u}\rangle \le n\alpha P_S + \frac{\delta_x}{2\alpha} + \frac{\alpha\delta_s}{2}. \quad (46)$$

Combining this with (11) and writing $\|\boldsymbol{u} + (1-\alpha)\boldsymbol{s}\|^2 = \|\boldsymbol{u}\|^2 + 2(1-\alpha)\langle \boldsymbol{s},\boldsymbol{u}\rangle + (1-\alpha)^2\|\boldsymbol{s}\|^2$, we obtain (45).

1) Proof of Lemma 2 : To prove (27), 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 (44), summing from $i = 1$ to n, and using the fact that $\overline{\boldsymbol{U}}$ is circularly symmetric, we find that 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 $(\boldsymbol{s},\boldsymbol{u})$, the distribution of $\|\boldsymbol{Y}\|^2$ coincides with that of $\|\boldsymbol{Y}'\|^2$, and we obtain (27).

We now turn to the proof of (28)–(29). 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 $\big|\|\boldsymbol{y}'\|^2 - nP_Y\big| \le n\epsilon$. By definition, $(\boldsymbol{X}'|\boldsymbol{s},\boldsymbol{u})$ is uniform on a shell of power $n(P+P_S)+\eta$ for some $-\delta_{xs} \le \eta \le \delta_{xs}$ (see (26) and (45)). 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 $f_{\boldsymbol{Y}'|\boldsymbol{SU}}^{(P_S)}(\boldsymbol{y}'|\boldsymbol{s},\boldsymbol{u}) \le K_2' \prod_{i=1}^n f_{Y,n}^{(P_S)}(y_i')$ for $\boldsymbol{y}' \in \mathcal{B}_\epsilon$, where K_2' is a constant depending on P_Y. The exponential decay in (28) 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 (29), we show that for $\boldsymbol{y}' \in \mathcal{B}_\epsilon$ we have $\prod_{i=1}^n f_{Y,n}^{(P_S)}(y_i') \le K_2'' \prod_{i=1}^n f_Y^{(P_S)}(y_i')$ for some constant K_2''. Recalling that $f_Y^{(P_S)} \sim N(0, P_Y)$, it is easy to show that

$$\frac{\prod_{i=1}^n f_{Y,n}^{(P_S)}(y_i')}{\prod_{i=1}^n f_Y^{(P_S)}(y_i')} = \sqrt{\frac{1}{1+\frac{\eta}{nP_Y}}} \exp\left(\frac{\|\boldsymbol{y}'\|^2\eta}{2(nP_Y)^2}\left(\frac{1}{1+\frac{\eta}{nP_Y}}\right)\right) \tag{47}$$

The desired result follows in a straightforward fashion using the definition of $\mathcal{B}_\epsilon$ and the fact that $\eta \in [-\delta_{xs}, \delta_{xs}]$. The constants ψ and K_2 in (28)–(29) can be taken as independent of P_S, since P_Y is uniformly bounded for $P_S \in \tilde{\mathcal{P}}_n$.

2) Proof of Lemma 3: The evaluation of the moments of the information density is cumbersome and similar to [18, Appendix A], so we omit some of the details.

We first consider the mean and variance. We write (44) as $i^{(P_S)}(u,y) = c_0 + c_1(y + c_2 u)^2 + c_3 y^2$, where (c_0, c_1, c_2, c_3) are constants. Substituting $y = u + (1-\alpha)s + z$, we obtain

$$i^{(P_S)}(u,y) = c_0 + d_1 s^2 + d_2 u^2 + d_3 z^2 + d_4 su + d_5 sz + d_6 uz, \tag{48}$$

for some constants $(d_1, d_2, d_3, d_4, d_5, d_6)$ which are written in terms of (c_1, c_2, c_3). 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 (48) with respect to $Z \sim N(0,1)$, we obtain

$$\mathbb{E}\big[i^{(P_S)}(u,Y) \,|\, s,u\big] = c_0 + d_1 s^2 + d_2 u^2 + d_3 + d_4 su \tag{49}$$

$$\mathrm{Var}\big[i^{(P_S)}(u,Y) \,|\, s,u\big] = 2d_3^2 + (d_5 s + d_6 u)^2, \tag{50}$$

where we have used $\mathbb{E}[Z] = 0$, $\mathrm{Var}[Z] = 1$, $\mathrm{Var}[Z^2] = 2$ and $\mathrm{Cov}[Z^2, Z] = 0$. By summing from $i = 1$ to n, we obtain analogous expressions for the mean and variance of $i_n^{(P_S)}(\boldsymbol{u},\boldsymbol{Y})$ given $(\boldsymbol{s},\boldsymbol{u})$. Using these expressions and the definitions of the constants c_i and d_i, it can be verified that, for any $(\boldsymbol{s},\boldsymbol{u})$ such that $\|\boldsymbol{s}\|^2 = nP_S$, $\|\boldsymbol{u}\|^2 = n(P+\alpha^2 P_S)$, and $\langle \boldsymbol{s},\boldsymbol{u}\rangle = n\alpha P_S$, we have

$$\mathbb{E}\big[i_n^{(P_S)}(\boldsymbol{u},\boldsymbol{Y}) \,|\, \boldsymbol{s},\boldsymbol{u}\big] = nI^{(P_S)}(U;Y) \tag{51}$$

$$\mathrm{Var}\big[i_n^{(P_S)}(\boldsymbol{u},\boldsymbol{Y}) \,|\, \boldsymbol{s},\boldsymbol{u}\big] = n\frac{1}{2(1+P+P_S)^2\big(PP_S(1-\alpha)^2 + P + \alpha^2 P_S\big)^2} \times \Bigg((P+\alpha P_S)^2\Big(\alpha^2 P_S(2+P_S) + P^2\Big(1 + 2(1-\alpha)^2 P_S\Big) + 2P\Big(1 + \alpha P_S + P_S(1-\alpha)^2(2+P_S)\Big)\Big)\Bigg). \tag{52}$$

The distribution of $(\boldsymbol{S},\boldsymbol{U})$ under consideration does not ensure that the equalities $\|\boldsymbol{u}\|^2 = n(P+\alpha^2 P_S)$ and $\langle \boldsymbol{s},\boldsymbol{u}\rangle = n\alpha P_S$ hold. However, they do hold to within an $O(1)$ term; see (11) and (46). From (49)–(50), the mean and variance of $i_n^{(P_S)}(\boldsymbol{u},\boldsymbol{Y})$ are affine in $\|\boldsymbol{s}\|^2$ and $\langle \boldsymbol{s},\boldsymbol{u}\rangle$, and hence (51) and (52) hold for all $(\boldsymbol{s},\boldsymbol{u})$ such that $f_{\boldsymbol{SU}}^{(P_S)}(\boldsymbol{s},\boldsymbol{u}) > 0$ upon adding $O(1)$ to the right-hand sides. Substituting $\alpha = \frac{P}{1+P}$ reveals that (52) equals $\frac{P(2+P)}{2(1+P)^2}$ for all P_S, thus proving (33)–(34).

Finally, (35) is proved in the same way as in [18, Appendix A] by first showing that 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 $\langle \boldsymbol{s},\boldsymbol{u}\rangle$, and choosing $(\boldsymbol{s}',\boldsymbol{u}')$ to attain the same powers and correlation as $(\boldsymbol{s},\boldsymbol{u})$, while having entries which are uniformly bounded for all n.

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

[5] A. Cohen and A. Lapidoth, "Generalized writing on dirty paper," in *IEEE Int. Symp. Inf. Theory*, 2002.

[6] J. Scarlett, "Second-order rate of constant-composition codes for the Gel'fand-Pinsker channel," in *Int. Zurich Sem. Comms.*, 2014.

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

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

[9] 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.

[10] 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.

[11] M. H. Yassaee, M. R. Aref, and A. Gohari, "A technique for deriving one-shot achievability results in network information theory," 2013, http://arxiv.org/abs/1303.0696.

[12] 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.

[13] J. Scarlett, "On the dispersions of the Gel'fand-Pinsker channel and dirty paper coding," 2013, submitted to *IEEE Trans. Inf. Theory* [arxiv: http://arxiv.org/abs/1309.6200].

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

[15] 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.

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

[17] 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.

[18] 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].