Second-Order Rate of Constant-Composition Codes for the Gel'fand-Pinsker Channel

Jonathan Scarlett

***Abstract*—This paper presents an achievable second-order coding rate for the discrete memoryless Gel'fand-Pinkser channel. The result is obtained using constant-composition random coding, and by using an asymptotically negligible fraction of the block to transmit the type of the state sequence.**

I. Introduction

In this paper, we present an achievable second-order coding rate [1]–[3] for channel coding with a random state known non-causally at the encoder, as studied by Gel'fand and Pinsker [4]. The alphabets of the input, output and state are denoted by $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{S}$ respectively, and each are assumed to be finite. The channel transition law is given by $W^n(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{s}) \triangleq \prod_{i=1}^n W(y_i|x_i,s_i)$, where n is the block length. The state sequence $\boldsymbol{S} = (S_1,\cdots,S_n)$ is assumed to be independent and identically distributed (i.i.d.) according to a distribution $\pi(s)$. The capacity is given by [4]

$$C = \max_{\mathcal{U},Q_{U|S},\phi(\cdot,\cdot)} I(U;Y) - I(U;S), \tag{1}$$

where the mutual informations are with respect to

$$P_{SUY}(s,u,y) = \pi(s)Q_{U|S}(u|s)W(y|\phi(u,s),s) \tag{2}$$

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

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 $M^*(n,\epsilon) \triangleq \max\big\{M : (n,M,\epsilon) \text{ is achievable}\big\}$. Letting $P_{Y|U}$, P_Y, etc. denote the marginals of (2), we define the information densities

$$i(u,s) \triangleq \log\frac{Q_{U|S}(u|s)}{P_U(u)} \tag{3}$$

$$i(u,y) \triangleq \log\frac{P_{Y|U}(y|u)}{P_Y(y)} \tag{4}$$

with a slight abuse of notation.

Theorem 1. *Let $\mathcal{U}$, $Q_{U|S}$ and $\phi(\cdot,\cdot)$ by any set of capacity-achieving parameters in* (1)*, and let P_{SUY}, $i(u,s)$ and $i(u,y)$ be as given in* (2)–(4) *under these parameters. If $\mathbb{E}\big[\mathrm{Var}[i(U,Y)\,|\,U,S]\big] > 0$, then*

$$\log M^*(n,\epsilon) \ge nC - \sqrt{nV}\mathsf{Q}^{-1}(\epsilon) + O(\log n), \tag{5}$$

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

$$V \triangleq \mathbb{E}\big[\mathrm{Var}[i(U,Y)\,|\,U,S]\big] + \mathrm{Var}\big[\mathbb{E}[i(U,Y)-i(U,S)\,|\,S]\big] \tag{6}$$

$$= \mathrm{Var}[i(U,Y)-i(U,S)]. \tag{7}$$

Proof: We provide a number of preliminary results in Section II, and present the proof in Section III. ∎

It should be noted that the equality in (7) holds under the capacity-achieving parameters, but more generally (7) is at least as high as (6), with strict inequality possible for suboptimal choices of $Q_{U|S}$.

To our knowledge, the only previous result on the second-order asymptotics for the present problem is that of Watanabe *et al.* [5] and Yassaee *et al.* [6], who used i.i.d. random coding. In [7], we show that for $\epsilon < \frac{1}{2}$ our second-order term is at least as good as that of [5], [6], with strict improvement possible. Furthermore, we show in [7] that Theorem 1 recovers, as a special case, the dispersion for channels with i.i.d. state known at both the encoder and decoder, which was derived in [8].

Notation: Bold symbols are used for vectors and matrices (e.g. $\boldsymbol{x}$), and the corresponding i-th entry of a vector is denoted with a subscript (e.g. x_i). The marginals of a joint distribution P_{XY} are denoted by P_X and P_Y. The empirical distribution (i.e. type [9, Ch. 2]) of a vector $\boldsymbol{x}$ is denoted by $\hat{P}_{\boldsymbol{x}}$. 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. We make use of the standard asymptotic notations $O(\cdot)$ and $o(\cdot)$.

II. Preliminary Results

In this section, we present a number of preliminary results which will prove useful in the proof of Theorem 1. We assume that $\mathcal{U}$, $Q_{U|S}$ and $\phi(\cdot,\cdot)$ achieve the capacity in (1).

A. A Genie-Aided Setting

We prove Theorem 1 by first proving the following result for a genie-aided setting.

Theorem 2. *Theorem 1 holds true in the case that the empirical distribution $\hat{P}_{\boldsymbol{S}}$ of $\boldsymbol{S}$ is known at the decoder.*

To see that Theorem 2 implies Theorem 1, we use a technique which was proposed in [10]. We use the first $g(n) = K_0\log(n+1)$ symbols of the block to transmit the

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.

type of the remaining $\tilde{n} = n - g(n)$ symbols. Using Gallager's random-coding bound [11, Sec. 5.6] and the fact that the number of such types is upper bounded by $(n+1)^{|\mathcal{S}|-1}$, it is easily shown that there exists a choice of K_0 such that the decoder estimates the state type correctly with probability $O\big(\frac{1}{n}\big)$. Thus, $\big(n - O(\log n), M, \epsilon - O\big(\frac{1}{n}\big)\big)$-achievability in the genie-aided setting implies (n, M, ϵ)-achievability in the absence of the genie. By performing a Taylor expansion of the square root and $\mathsf{Q}^{-1}(\cdot)$ function in (5), we obtain the desired result.

B. A Typical Set

We define a typical set of state types given by

$$\tilde{\mathcal{P}}_n = \bigg\{ P_S \in \mathcal{P}_n(\mathcal{S}) \,:\, \|P_S - \pi\|_\infty \le \sqrt{\frac{\log n}{n}} \bigg\}. \quad (8)$$

We will see the second-order performance is unaffected by types falling outside $\tilde{\mathcal{P}}_n$, due to the fact that [8, Lemma 22]

$$\mathbb{P}\big[\hat{P}_{\boldsymbol{S}} \notin \tilde{\mathcal{P}}_n\big] = O\Big(\frac{1}{n^2}\Big). \quad (9)$$

C. Approximations of Distributions

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 $Q_{U|S}$ in terms of L_∞ norm, and such that $Q_{U|S,n}^{(P_S)}(u|s) = 0$ wherever $Q_{U|S}(u|s) = 0$. 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 (8) that $\min_s nP_S(s)$ grows linearly in n for all $P_S \in \tilde{\mathcal{P}}_n$. Thus,

$$\Big| Q_{U|S}(u|s) - Q_{U|S,n}^{(P_S)}(u|s) \Big| = O\Big(\frac{1}{n}\Big) \quad (10)$$

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

We will make use of the following joint distributions:

$$P_{SUY}^{(P_S)}(s,u,y) \triangleq P_S(s) Q_{U|S}(u|s) W(y|\phi(u,s),s) \quad (11)$$
$$P_{SUY,n}^{(P_S)}(s,u,y) \triangleq P_S(s) Q_{U|S,n}^{(P_S)}(u|s) W(y|\phi(u,s),s). \quad (12)$$

Using (10), we immediately obtain that

$$\Big| P_{SUY}^{(P_S)}(s,u,y) - P_{SUY,n}^{(P_S)}(s,u,y) \Big| = O\Big(\frac{1}{n}\Big) \quad (13)$$

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

D. A Taylor Expansion of the Mutual Information

Let $I^{(P_S)}(U;S)$ and $I^{(P_S)}(U;Y)$ denote mutual informations under the joint distribution $P_{USY}^{(P_S)}$ in (11), and define

$$I(P_S) \triangleq I^{(P_S)}(U;Y) - I^{(P_S)}(U;S). \quad (14)$$

We observe from (1) that $C = I(\pi)$. The following Taylor expansion (about $P_S = \pi$) is proved in [7]:

$$I(P_S) = \tilde{I}(P_S) + \Delta(P_S), \quad (15)$$

where

$$\tilde{I}(P_S) \triangleq \sum_s P_S(s) \sum_u Q_{U|S}(u|s) \times \bigg(\sum_y W(y|\phi(u,s),s) \log \frac{P_{Y|U}^{(\pi)}(y|u)}{P_Y^{(\pi)}(y)} - \log \frac{Q_{U|S}(u|s)}{P_U^{(\pi)}(u)} \bigg), \quad (16)$$

and

$$\max_{P_S \in \tilde{\mathcal{P}}_n} |\Delta(P_S)| \le \frac{K_1 \log n}{n} \quad (17)$$

for some constant K_1.

III. Proof of Theorem 1

As stated above, it suffices to prove Theorem 2. Thus, we assume that the state type P_S is known at the decoder.

1) Random-Coding Parameters: 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 $L^{(P_S)}$ for each state type $P_S \in \mathcal{P}_n(\mathcal{S})$. 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, we randomly generate an auxiliary codebook $\{\boldsymbol{U}^{(P_S)}(m,l)\}_{l=1}^{L^{(P_S)}}$, where each codeword is drawn independently according to the uniform distribution on the type class $T^n(P_{U,n}^{(P_S)})$ (see (12)). 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

$$\phi^n(\boldsymbol{U},\boldsymbol{S}) \triangleq \big(\phi(U_1,S_1), \cdots, \phi(U_n,S_n)\big), \quad (18)$$

where $\boldsymbol{U}$ is an auxiliary codeword $\boldsymbol{U}^{(P_S)}(m,l)$ with l chosen such that $(\boldsymbol{S},\boldsymbol{U}) \in T^n(P_{SU,n}^{(P_S)})$, with an error declared if no such auxiliary codeword exists. Given $\boldsymbol{y}$ 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

$$i_n^{(P_S)}(\boldsymbol{u},\boldsymbol{y}) \triangleq \sum_{i=1}^n i^{(P_S)}(u_i,y_i), \quad (19)$$

where

$$i^{(P_S)}(u_i,y_i) \triangleq \log \frac{P_{Y|U}^{(P_S)}(y|u)}{P_Y^{(P_S)}(y)} \quad (20)$$

with $P_{SUY}^{(P_S)}$ defined in (11). It should be noted that $P_{SUY}^{(\pi)}$ coincides with the distribution in (2), and hence $i^{(\pi)}(u,y)$ coincides with (4).

We consider the events

$$\mathcal{E}_1 \triangleq \Big\{ \text{No } l \text{ yields } (\boldsymbol{S}, \boldsymbol{U}^{(P_S)}(m,l)) \in T^n(P_{SU,n}^{(P_S)}) \Big\} \quad (21)$$
$$\mathcal{E}_2 \triangleq \Big\{ \text{Decoder chooses a message } \tilde{m} \neq m \Big\}. \quad (22)$$

It follows from these definitions and (9) that the overall random-coding error probability $\overline{p}_e$ satisfies

$$\overline{p}_e \le \sum_{P_S \in \tilde{\mathcal{P}}_n} \mathbb{P}\big[\hat{P}_{\boldsymbol{S}} = P_S\big] \Big(\mathbb{P}\big[\mathcal{E}_1 \,|\, \hat{P}_{\boldsymbol{S}} = P_S\big] + \mathbb{P}\big[\mathcal{E}_2 \,|\, \hat{P}_{\boldsymbol{S}} = P_S, \mathcal{E}_1^c\big] \Big) + O\Big(\frac{1}{n^2}\Big). \quad (23)$$

4) Analysis of $\mathcal{E}_1$: We study the probability of $\mathcal{E}_1$ conditioned on $\boldsymbol{S}$ having a given type $P_S \in \tilde{\mathcal{P}}_n$. Combining (13) with a standard property of types [12, Eq. (18)], each of the auxiliary codewords induces the joint type $P_{SU,n}^{(P_S)}$ with probability at least $p_0(n)^{-1}e^{-nI^{(P_S)}(U;S)}$, where $I^{(P_S)}(U;S)$ is defined in Section II-D, and $p_0(n)$ is polynomial in n. Since the codewords are independent, we have

$$\mathbb{P}\big[\mathcal{E}_1 \,|\, \hat{P}_{\boldsymbol{S}} = P_S\big] \le \big(1 - p_0(n)^{-1}e^{-nI^{(P_S)}(U;S)}\big)^{L^{(P_S)}} \tag{24}$$

$$\le \exp\Big(-p_0(n)^{-1}e^{-n\big(I^{(P_S)}(U;S)-R_L^{(P_S)}\big)}\Big), \tag{25}$$

where (25) follows using $1-\alpha \le e^{-\alpha}$ and defining

$$R_L^{(P_S)} \triangleq \frac{1}{n}\log L^{(P_S)}. \tag{26}$$

Choosing

$$R_L^{(P_S)} = I^{(P_S)}(U;S) + K_2\frac{\log n}{n} \tag{27}$$

with K_2 equal to one plus the degree of the polynomial $p_0(n)$, we obtain from (25) that

$$\mathbb{P}\big[\mathcal{E}_1 \,|\, P_S\big] \le e^{-\psi n} \tag{28}$$

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 $\mathcal{E}_1^c$. By symmetry, all $(\boldsymbol{s},\boldsymbol{u}) \in T^n(P_{SU,n}^{(P_S)})$ are equally likely, and hence the conditional distribution given $\hat{P}_{\boldsymbol{S}} = P_S$ and $\mathcal{E}_1^c$ of the state sequence $\boldsymbol{S}$, auxiliary codeword $\boldsymbol{U}$, and received sequence $\boldsymbol{Y}$ is given by

$$(\boldsymbol{S},\boldsymbol{U},\boldsymbol{Y}) \sim P_{\boldsymbol{SU}}^{(P_S)}(\boldsymbol{s},\boldsymbol{u})W^n(\boldsymbol{y}|\phi^n(\boldsymbol{u},\boldsymbol{s}),\boldsymbol{s}), \tag{29}$$

where $P_{\boldsymbol{SU}}^{(P_S)}$ is uniform on the type class:

$$P_{\boldsymbol{SU}}^{(P_S)}(\boldsymbol{s},\boldsymbol{u}) \triangleq \frac{1}{\big|T^n(P_{SU,n}^{(P_S)})\big|}\mathbb{1}\Big\{(\boldsymbol{s},\boldsymbol{u}) \in T^n(P_{SU,n}^{(P_S)})\Big\}. \tag{30}$$

Let $P_{\boldsymbol{Y}}^{(P_S)}(\boldsymbol{y}) \triangleq \sum_{\boldsymbol{u},\boldsymbol{s}} P_{\boldsymbol{US}}^{(P_S)}(\boldsymbol{u},\boldsymbol{s})W^n(\boldsymbol{y}|\phi^n(\boldsymbol{u},\boldsymbol{s}),\boldsymbol{s})$ be the corresponding output distribution. Using a standard change of measure from constant-composition to i.i.d. (e.g. see [9, Ch. 2]), we can easily show that

$$P_{\boldsymbol{Y}}^{(P_S)}(\boldsymbol{y}) \le p_1(n)\prod_{i=1}^n P_Y^{(P_S)}(y_i), \tag{31}$$

where $p_1(n)$ is polynomial in n.

Recall that the decoder maximizes $i_n^{(P_S)}$ given in (19). Using a well-known threshold-based non-asymptotic bound [2], we have for any $\gamma^{(P_S)}$ that

$$\mathbb{P}\big[\mathcal{E}_2 \,|\, \hat{P}_{\boldsymbol{S}} = P_S, \mathcal{E}_1^c\big] \le \mathbb{P}\Big[i_n^{(P_S)}(\boldsymbol{U},\boldsymbol{Y}) \le \gamma^{(P_S)}\Big] + ML^{(P_S)}\mathbb{P}\Big[i_n^{(P_S)}(\overline{\boldsymbol{U}},\boldsymbol{Y}) > \gamma^{(P_S)}\Big], \tag{32}$$

where $\overline{\boldsymbol{U}} \sim P_{\boldsymbol{U}}^{(P_S)}$ independently of $(\boldsymbol{S},\boldsymbol{U},\boldsymbol{Y})$. Using the change of measure given in (31), we can apply standard steps (e.g. see [3]) to upper bound the second term in (32) by $p_2(n)ML^{(P_S)}e^{-\gamma^{(P_S)}}$, where $p_2(n)$ is polynomial in n. We can ensure that this term is $O\big(\frac{1}{n}\big)$ by choosing $\gamma^{(P_S)} = \log ML^{(P_S)} + K_3\log n$, where K_3 is one higher than the degree of $p_2(n)$. Under this choice, and defining $K_4 \triangleq K_2 + K_3$, we obtain from (27) and (32) that

$$\mathbb{P}\big[\mathcal{E}_2 \,|\, \hat{P}_{\boldsymbol{S}} = P_S\big] \le \mathbb{P}\Big[i_n^{(P_S)}(\boldsymbol{U},\boldsymbol{Y}) \le \log M + nI^{(P_S)}(U;S) + K_4\log n\Big] + O\Big(\frac{1}{n}\Big). \tag{33}$$

6) Application of the Berry-Esseen Theorem: Combining (28) and (33), we have for all $P_S \in \tilde{\mathcal{P}}_n$ that

$$\mathbb{P}\big[\mathcal{E}_1 \cup \mathcal{E}_2 \,|\, \hat{P}_{\boldsymbol{S}} = P_S\big] \le \mathbb{P}\Big[i_n^{(P_S)}(\boldsymbol{U},\boldsymbol{Y}) \le \log M + nI^{(P_S)}(U;S) + K_4\log n\Big] + O\Big(\frac{1}{n}\Big). \tag{34}$$

In order to apply the Berry-Esseen theorem to the right-hand side of (34), we first compute the mean and variance of $i_n^{(P_S)}(\boldsymbol{U},\boldsymbol{Y})$, defined according to (19) and (29). The required third moment can easily be uniformly bounded in terms of the alphabet sizes [13, Appendix D]. We will use the fact that, by the symmetry of the constant-composition distribution in (30), 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 (12), it follows that

$$\mathbb{E}\big[i_n^{(P_S)}(\boldsymbol{U},\boldsymbol{Y})\big] = n\sum_{u,y} P_{UY,n}^{(P_S)}(u,y)i^{(P_S)}(u,y) \tag{35}$$

$$= nI^{(P_S)}(U;Y) + O(1), \tag{36}$$

where (35) follows by expanding the expectation as a sum from 1 to n, and (36) follows from (13) and the definitions of $i^{(P_S)}(u,y)$ and $I^{(P_S)}(U;Y)$. A similar argument yields

$$\mathrm{Var}\big[i_n^{(P_S)}(\boldsymbol{U},\boldsymbol{Y})\big] = n\mathbb{E}\Big[\mathrm{Var}\big[i^{(P_S)}(U,Y) \,|\, U,S\big]\Big] + O(1) \tag{37}$$

$$\triangleq nV(P_S) + O(1). \tag{38}$$

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 (36) and (38) are uniform in $P_S \in \tilde{\mathcal{P}}_n$.

Using the definition of $I(P_S)$ in (14), we choose

$$\log M = nI(\pi) - K_4\log n - \beta_n, \tag{39}$$

where β_n will be specified later, and will behave as $O(\sqrt{n})$. Combining (34), (36), (38) and (39), we have

$$\mathbb{P}\big[\mathcal{E}_1 \cup \mathcal{E}_2 \,|\, \hat{P}_{\boldsymbol{S}} = P_S\big]$$

$$\le \mathbb{P}\Big[i_n^{(P_S)}(\boldsymbol{U},\boldsymbol{Y}) \le nI(\pi) + nI^{(P_S)}(U;S) - \beta_n\Big] + O\Big(\frac{1}{n}\Big). \tag{40}$$

$$\le \mathsf{Q}\bigg(\frac{\beta_n + nI(P_S) - nI(\pi) + K_5}{\sqrt{nV(P_S) + K_6}}\bigg) + O\bigg(\frac{1}{\sqrt{n}}\bigg) \tag{41}$$

where (41) follows by conditioning on $(\boldsymbol{S},\boldsymbol{U}) = (\boldsymbol{s},\boldsymbol{u})$ for some $(\boldsymbol{s},\boldsymbol{u}) \in T^n(P_{SU,n}^{(P_S)})$ (recall that this does not change the statistics of $i_n^{(P_S)}(\boldsymbol{U},\boldsymbol{Y})$), applying the Berry-Esseen theorem for independent and non-identically distributed variables [14, Sec. XVI.5], and introducing the constants K_5 and K_6 to represent the uniform $O(1)$ terms in (36) and (38).

7) Averaging Over the State Type: Substituting (41) into (23), we have

$$\overline{p}_e \le \sum_{P_S\in\tilde{\mathcal{P}}_n} \mathbb{P}\big[\hat{P}_{\boldsymbol{S}} = P_S\big]\mathsf{Q}\bigg(\frac{\beta + nI(P_S) - nI(\pi)}{\sqrt{nV(P_S)}}\bigg) + O\bigg(\frac{1}{\sqrt{n}}\bigg), \quad (42)$$

where we have factored the constants K_5 and K_6 into the remainder term using standard Taylor expansions along with the assumption $\beta_n = O(\sqrt{n})$; see [7] for details. Analogously to [8, Lemmas 17-18], we simplify (42) using two lemmas.

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

$$\sum_{P_S\in\tilde{\mathcal{P}}_n} \mathbb{P}\big[\hat{P}_{\boldsymbol{S}} = P_S\big]\mathsf{Q}\left(\frac{\beta_n + nI(P_S) - nI(\pi)}{\sqrt{nV(P_S)}}\right) \le \sum_{P_S\in\tilde{\mathcal{P}}_n} \mathbb{P}\big[\hat{P}_{\boldsymbol{S}} = P_S\big]\mathsf{Q}\left(\frac{\beta_n + nI(P_S) - nI(\pi)}{\sqrt{nV(\pi)}}\right) + O\Big(\frac{\log n}{\sqrt{n}}\Big) \quad (43)$$

Proof: This follows using standard Taylor expansions along with the definition of $\tilde{\mathcal{P}}_n$ in (8) and the fact that $V(P_S)$ is continuously differentiable at $P_S = \pi$; see [7]. ∎

Lemma 2. *For any β_n, we have*

$$\sum_{P_S\in\tilde{\mathcal{P}}_n} \mathbb{P}\big[\hat{P}_{\boldsymbol{S}} = P_S\big]\mathsf{Q}\bigg(\frac{\beta_n + nI(P_S) - nI(\pi)}{\sqrt{nV(\pi)}}\bigg) \le \mathsf{Q}\left(\frac{\beta_n}{\sqrt{nV}}\right) + O\Big(\frac{\log n}{\sqrt{n}}\Big), \quad (44)$$

where V is defined in (6).

Proof: Using the expansion of $I(P_S)$ in terms of $\tilde{I}(P_S)$ and $\Delta(P_S)$ given in (15), along with the property given in (17), we can easily show that the left-hand side of (44) is upper bounded by

$$\sum_{P_S\in\tilde{\mathcal{P}}_n} \mathbb{P}\big[\hat{P}_{\boldsymbol{S}} = P_S\big]\mathsf{Q}\left(\frac{\beta_n - nI(\pi) + n\tilde{I}(P_S)}{\sqrt{nV(\pi)}}\right) + O\Big(\frac{\log n}{\sqrt{n}}\Big). \quad (45)$$

Since $\tilde{I}(P_S)$ is written in the form $\sum_s P_S(s)\psi(s)$, a trivial generalization of [8, Lemma 18] gives

$$\sum_{P_S} \mathbb{P}\big[\hat{P}_{\boldsymbol{S}} = P_S\big]\mathsf{Q}\left(\frac{\beta_n + n\tilde{I}(P_S) - n\tilde{I}(\pi)}{\sqrt{nV(\pi)}}\right) = \mathsf{Q}\left(\frac{\beta_n}{\sqrt{n\big(V(\pi)+V^*(\pi)\big)}}\right) + O\bigg(\frac{1}{\sqrt{n}}\bigg), \quad (46)$$

where $V^*(\pi) \triangleq \mathrm{Var}_\pi[\psi(S)]$. Using (16), we see that $\psi(S) = \mathbb{E}[i^{(\pi)}(U,Y) - i^{(\pi)}(U,S)\,|\,S]$, and it follows that $V(\pi)+V^*(\pi)$ is equal to V, defined in (6). The proof is concluded by expanding the summation in (45) to be over all types, and substituting (46). ∎

Using (42) along with Lemmas 1 and 2, we have

$$\overline{p}_e \le \mathsf{Q}\bigg(\frac{\beta_n}{\sqrt{nV}}\bigg) + O\Big(\frac{\log n}{\sqrt{n}}\Big). \quad (47)$$

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

$$\beta_n = \sqrt{nV}\mathsf{Q}^{-1}(\epsilon) + O(\log n). \quad (48)$$

Consistent with (42) and Lemma 1, we have $\beta_n = O(\sqrt{n})$. Substituting (48) into (39) yields the desired result with V of the form given in (6).

By analyzing the Karush-Kuhn-Tucker (KKT) corresponding to the maximization in (1), it can be shown that the equality in (7) holds under any $Q_{U|S}$ which maximizes the objective for a given pair $(\mathcal{U},\phi)$ [7]. Since the parameters are capacity-achieving by assumption, this completes the proof.

Acknowledgment

I would like to thank Vincent Tan for many helpful comments and suggestions.

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] 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.

[6] 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.

[7] 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].

[8] 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].

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

[10] 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.

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

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

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

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