When is Noisy State Information at the Encoder as Useless as No Information or as Good as Noise-Free State?

Rui Xu, Jun Chen, *Senior Member, IEEE*, Tsachy Weissman, *Fellow, IEEE*, and Jian-Kang Zhang, *Senior Member, IEEE*

***Abstract*— For any binary-input channel with perfect state information at the decoder, if the mutual information between the noisy state observation at the encoder and the true channel state is below a positive threshold determined solely by the state distribution, then the capacity is the same as that with no encoder side information. A complementary phenomenon is revealed for the generalized probing capacity. Extensions beyond binary-input channels are developed.**

***Index Terms*— Binary-input, channel capacity, erasure channel, probing capacity, state information, stochastically degraded.**

I. Introduction

Consider a memoryless channel $p_{Y|X,S}$ with input X, output Y, and state S. We assume that the channel state S, distributed according to p_S, is provided to the decoder, and a noisy state observation $\tilde{S}$, generated by S through side channel $p_{\tilde{S}|S}$, is available causally at the encoder. Here X, Y, S, and $\tilde{S}$ are defined over finite alphabets $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{S}$, and $\tilde{\mathcal{S}}$, respectively. In this setting (see Fig. 1), Shannon's remarkable result [1] (see also [2, eq. (3)] and [3, Th. 7.2]) implies that the channel capacity is given by

$$C(p_{Y|X,S}, p_S, p_{\tilde{S}|S}) \triangleq \max_{p_U} I(U; Y|S). \tag{1}$$

The auxiliary random variable U is defined over alphabet $\mathcal{U}$ with $|\mathcal{U}| = |\mathcal{X}|^{|\tilde{\mathcal{S}}|}$, whose joint distribution with $(X, Y, S, \tilde{S})$ factors as

$$\begin{aligned} &p_{U,X,Y,S,\tilde{S}}(u, x, y, s, \tilde{s}) \\ &\quad = p_U(u)p_S(s)p_{\tilde{S}|S}(\tilde{s}|s)\mathbb{I}(x = \psi(u, \tilde{s}))p_{Y|X,S}(y|x, s), \\ &\qquad\qquad u \in \mathcal{U}, x \in \mathcal{X}, y \in \mathcal{Y}, s \in \mathcal{S}, \tilde{s} \in \tilde{\mathcal{S}}, \end{aligned} \tag{2}$$

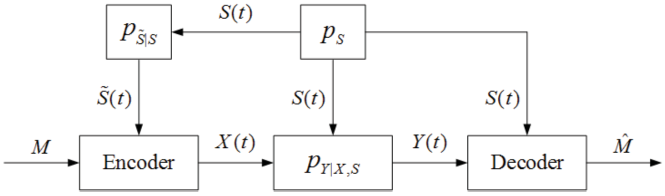

Fig. 1. Channel model.

where $\mathbb{I}(\cdot)$ is the indicator function, and $\psi(u, \cdot)$, $u \in \mathcal{U}$, are $|\mathcal{X}|^{|\tilde{\mathcal{S}}|}$ different mappings from $\tilde{\mathcal{S}}$ to $\mathcal{X}$. Without loss of generality, we set $\mathcal{X} = \{0, 1, \cdots, |\mathcal{X}| - 1\}$, $\mathcal{S} = \{0, 1 \cdots, |\mathcal{S}| - 1\}$, $\mathcal{U} = \{0, 1, \cdots, |\mathcal{X}|^{|\tilde{\mathcal{S}}|} - 1\}$, and order the mappings $\psi(u, \cdot)$, $u \in \mathcal{U}$, in such a way that the first $|\mathcal{X}|$ mappings[1] are

$$\psi(u, \cdot) \equiv u, \quad u \in \mathcal{X}; \tag{3}$$

moreover, we assume that $\rho \triangleq \min_{s \in \mathcal{S}} p_S(s) > 0$. The capacity formula (1) can be simplified in the following two special cases. Specifically, when there is no encoder side information, the channel capacity reduces to [3, eq. (7.2)]

$$\underline{C}(p_{Y|X,S}, p_S) \triangleq \max_{p_X} I(X; Y|S), \tag{4}$$

where $p_{X,Y,S}(x, y, s) = p_X(x)p_S(s)p_{Y|X,S}(y|x, s)$; on the other hand, when perfect state information is available at the encoder (as well as the decoder), the channel capacity becomes [3, eq. (7.3)]

$$\overline{C}(p_{Y|X,S}, p_S) \triangleq \max_{p_{X|S}} I(X; Y|S), \tag{5}$$

where $p_{X,Y,S}(x, y, s) = p_S(s)p_{X|S}(x|s)p_{Y|X,S}(y|x, s)$.

For comparison, consider the following similarly defined quantity

$$C'(p_{Y|X,S}, p_S, p_{\tilde{S}|S}) \triangleq \max_{p_U} I(X; Y|S),$$

where the joint distribution of $(U, X, Y, S, \tilde{S})$ is also given by (2). We shall refer to $C'(p_{Y|X,S}, p_S, p_{\tilde{S}|S})$ as the generalized probing capacity. By the functional representation lemma [3, p. 626] (see also [5, Lemma 1]), $C'(p_{Y|X,S}, p_S, p_{\tilde{S}|S})$ can be defined equivalently as

$$C'(p_{Y|X,S}, p_S, p_{\tilde{S}|S}) \triangleq \max_{p_{X|\tilde{S}}} I(X; Y|S),$$

Manuscript received November 16, 2015; revised July 15, 2016; accepted October 28, 2016. Date of publication November 21, 2016; date of current version January 18, 2017. R. Xu was supported by the Natural Sciences and Engineering Research Council of Canada through a Discovery Grant. T. Weissman was supported by the NSF Center for Science of Information. This paper was presented at the 2016 IEEE International Symposium on Information Theory.

R. Xu, J. Chen, and J.-K. Zhang are with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada (e-mail: xur27@mcmaster.ca; junchen@ece.mcmaster.ca; jkzhang@ece.mcmaster.ca).

T. Weissman is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: tsachy@stanford.edu).

Communicated by S. S. Pradhan, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2016.2631150

[1] These are the mappings that ignore the encoder side information.

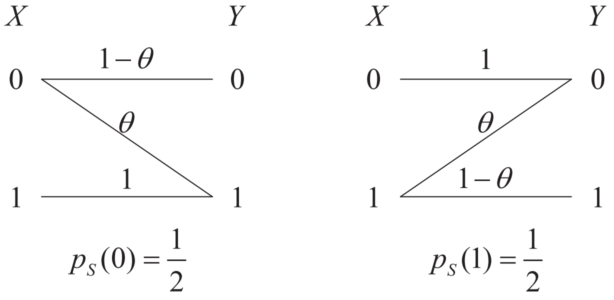

Fig. 2. Illustration of $p_{Y|X,S}$ and p_S given by (9) and (10), respectively.

where

$$p_{X,Y,S,\tilde{S}}(x,y,s,\tilde{s}) = p_S(s)p_{\tilde{S}|S}(\tilde{s}|s)p_{X|\tilde{S}}(x|\tilde{s})p_{Y|X,S}(y|x,s),\quad x\in\mathcal{X}, y\in\mathcal{Y}, s\in\mathcal{S}, \tilde{s}\in\tilde{\mathcal{S}}.$$

Clearly,

$$\begin{aligned}\underline{C}(p_{Y|X,S}, p_S) &\leq C(p_{Y|X,S}, p_S, p_{\tilde{S}|S})\\ &\leq C'(p_{Y|X,S}, p_S, p_{\tilde{S}|S})\\ &\leq \overline{C}(p_{Y|X,S}, p_S). \qquad (6)\end{aligned}$$

Moreover, we have

$$\begin{aligned}C(p_{Y|X,S}, p_S, p_{\tilde{S}|S}) &= C'(p_{Y|X,S}, p_S, p_{\tilde{S}|S})\\ &= \underline{C}(p_{Y|X,S}, p_S) \qquad (7)\end{aligned}$$

if S and $\tilde{S}$ are independent (i.e., $I(S;\tilde{S})=0$), and

$$\begin{aligned}C(p_{Y|X,S}, p_S, p_{\tilde{S}|S}) &= C'(p_{Y|X,S}, p_S, p_{\tilde{S}|S})\\ &= \overline{C}(p_{Y|X,S}, p_S) \qquad (8)\end{aligned}$$

if S is a deterministic function of $\tilde{S}$ (i.e., $H(S|\tilde{S})=0$).

To elucidate the operational meaning of $C'(p_{Y|X,S}, p_S, p_{\tilde{S}|S})$ and its connection with $C(p_{Y|X,S}, p_S, p_{\tilde{S}|S})$, it is instructive to consider the special case where $p_{\tilde{S}|S}$ is a binary erasure channel with erasure probability ϵ (denoted by BEC(ϵ)), which corresponds to the probing channel setup studied in [4]. The probing channel model is essentially the same as the one in Fig. 1 except that, in Fig. 1, the encoder (which, with high probability, observes approximately $n\epsilon$ state symbols out of the whole state sequence of length n when n is large enough) has no control of the exact positions of these $n\epsilon$ symbols whereas, in the probing channel model, the encoder has the freedom to specify the positions of these $n\epsilon$ symbols according to the message to be sent. It is shown in [4] that this additional freedom increases the achievable rate from $C(p_{Y|X,S}, p_S, \mathrm{BEC}(\epsilon))$ to $C'(p_{Y|X,S}, p_S, \mathrm{BEC}(\epsilon))$. Now consider an example (see also Fig. 2) where

$$p_{Y|X,S}(y|x,s) = \begin{cases} 1-\theta, & (x,y,s)=(0,0,0) \text{ or } (1,1,1),\\ \theta, & (x,y,s)=(0,1,0) \text{ or } (1,0,1),\\ 0, & (x,y,s)=(1,0,0) \text{ or } (0,1,1),\\ 1, & (x,y,s)=(1,1,0) \text{ or } (0,0,1),\end{cases} \qquad (9)$$

$$p_S(0)=p_S(1)=\frac{1}{2}. \qquad (10)$$

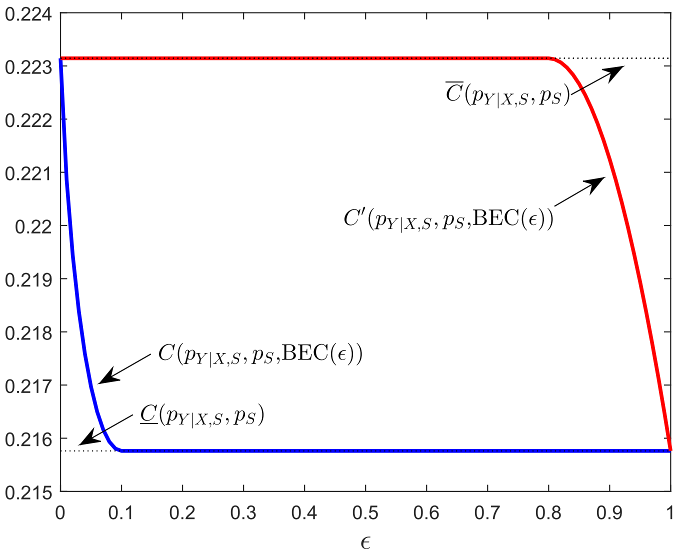

Fig. 3. Plots of $C(p_{Y|X,S}, p_S, \mathrm{BEC}(\epsilon))$ and $C'(p_{Y|X,S}, p_S, \mathrm{BEC}(\epsilon))$ against ϵ for $\epsilon\in[0,1]$, where $p_{Y|X,S}$ and p_S are given by (9) with $\theta=\frac{1}{2}$ and (10), respectively.

For this example, it can be verified that

$$\underline{C}(p_{Y|X,S}, p_S) = \begin{cases}\log 2, & \theta=0,\\ \frac{1}{2}\left((1-\theta)\log 2+\log\frac{2}{1+\theta}+\theta\log\frac{2\theta}{1+\theta}\right), & \theta\in(0,1),\\ 0, & \theta=1,\end{cases}$$

$$\overline{C}(p_{Y|X,S}, p_S) = \begin{cases}\log 2, & \theta=0,\\ \log\left(1+(1-\theta)\theta^{\frac{\theta}{1-\theta}}\right), & \theta\in(0,1),\\ 0, & \theta=1.\end{cases}$$

Note that $\overline{C}(p_{Y|X,S}, p_S)$ is strictly greater than $\underline{C}(p_{Y|X,S}, p_S)$ unless $\theta=0$ or $\theta=1$. It follows by (7) and (8) that

$$\begin{aligned}C(p_{Y|X,S}, p_S, \mathrm{BEC}(\epsilon))\big|_{\epsilon=1} &= C'(p_{Y|X,S}, p_S, \mathrm{BEC}(\epsilon))\big|_{\epsilon=1}\\ &= \underline{C}(p_{Y|X,S}, p_S),\\ C(p_{Y|X,S}, p_S, \mathrm{BEC}(\epsilon))\big|_{\epsilon=0} &= C'(p_{Y|X,S}, p_S, \mathrm{BEC}(\epsilon))\big|_{\epsilon=0}\\ &= \overline{C}(p_{Y|X,S}, p_S).\end{aligned}$$

To gain a better understanding, we plot $C(p_{Y|X,S}, p_S, \mathrm{BEC}(\epsilon))$ and $C'(p_{Y|X,S}, p_S, \mathrm{BEC}(\epsilon))$ against ϵ for $\epsilon\in[0,1]$ in Fig. 3. It turns out that, somewhat counterintuitively, $C(p_{Y|X,S}, p_S, \mathrm{BEC}(\epsilon))$ coincides with $\underline{C}(p_{Y|X,S}, p_S)$ way before ϵ reaches 1. That is to say, when ϵ is above a certain threshold strictly less than 1, the noisy state observation $\tilde{S}$ is useless and can be ignored (as far as the channel capacity is concerned). On the the hand, it can be seen that $C'(p_{Y|X,S}, p_S, \mathrm{BEC}(\epsilon))$ is equal to $\overline{C}(p_{Y|X,S}, p_S)$ for a large range of ϵ strictly greater than 0. Hence, in terms of the probing capacity, the noisy state observation can be as good as the perfect one. As shown in Fig. 4, the same phenomena arise if we choose $p_{\tilde{S}|S}$ to be a binary symmetric channel with crossover probability q (denoted by BSC(q)).

The contributions of the present work are summarized in the following theorems, which indicate that the aforedescribed

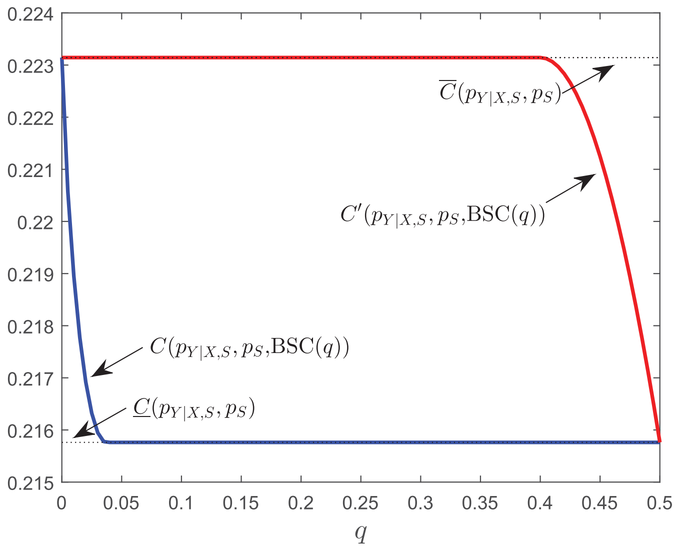

Fig. 4. Plots of $C(p_{Y|X,S}, p_S, \text{BSC}(q))$ and $C'(p_{Y|X,S}, p_S, \text{BSC}(q))$ against q for $q \in [0, \frac{1}{2}]$, where $p_{Y|X,S}$ and p_S are given by (9) with $\theta = \frac{1}{2}$ and (10), respectively.

surprising phenomena can in fact be observed for all binary-input channels.

Theorem 1: For any binary-input channel $p_{Y|X,S}$, state distribution p_S, and side channel $p_{\tilde{S}|S}$,

$$C(p_{Y|X,S}, p_S, p_{\tilde{S}|S}) = \underline{C}(p_{Y|X,S}, p_S)$$

if $I(S; \tilde{S}) \leq \frac{\rho^2}{2e^2}$, where $\rho \triangleq \min_{s\in\mathcal{S}} p_S(s)$.

Theorem 2: For any binary-input channel $p_{Y|X,S}$, state distribution p_S, and side channel $p_{\tilde{S}|S}$,

$$C'(p_{Y|X,S}, p_S, p_{\tilde{S}|S}) = \overline{C}(p_{Y|X,S}, p_S)$$

if $H(S|\tilde{S}) \leq \frac{2\rho\log 2}{(|\mathcal{S}|-1)(e-1)}$, where $\rho \triangleq \min_{s\in\mathcal{S}} p_S(s)$.

On the surface these two results may look rather similar. One might even suspect the existence of a certain duality between them. However, it will be seen that the underlying reasons are actually quite different. The proof of Theorem 1 hinges upon, among other things, a perturbation analysis. In contrast, Theorem 2 is essentially a manifestation of an induced Markov structure.

The conditions in Theorem 1 and Theorem 2 are stated in terms of bounds on $I(S; \tilde{S})$ and $H(S|\tilde{S})$; as a consequence, they depend inevitably on p_S. As shown by Theorem 3 in Section II and Theorem 4 in Section III, it is in fact possible to establish these two results under more general conditions on $p_{\tilde{S}|S}$ that are universal for all binary-input channels and state distributions.

The rest of this paper is organized as follows. We present the proofs of Theorems 1 and 2 in Sections II and III, respectively. The validity of these two results under various modified conditions is discussed in Section IV. Section V contains some concluding remarks. Throughout this paper, all logarithms are base-e.

II. Proof of Theorem 1

First consider the special case where $p_{\tilde{S}|S}$ is a generalized erasure channel (with erasure probability $\epsilon \in [0, 1]$) defined as

$$p_{\tilde{S}^{(\epsilon)}_{GE}|S}(\tilde{s}|s) = \begin{cases} 1-\epsilon, & \tilde{s} = s, \\ \epsilon, & \tilde{s} = *, \\ 0, & \text{otherwise}, \end{cases} \qquad s \in \mathcal{S}, \tilde{s} \in \mathcal{S} \cup \{*\}.$$

Lemma 1: Given any binary-input channel $p_{Y|X,S}$ and state distribution p_S,

$$C(p_{Y|X,S}, p_S, p_{\tilde{S}^{(\epsilon)}_{GE}|S}) = \underline{C}(p_{Y|X,S}, p_S)$$

for $\epsilon \in [1 - e^{-1}, 1]$.

Remark: Lemma 1 provides a universal upper bound[2] on the erasure probability threshold above which the encoder side information is useless. The actual threshold, however, depends on $p_{Y|X,S}$ and p_S (see Section IV-A for a detailed analysis).

Proof: As indicated by (1), the capacity of the channel model in Fig. 1 (i.e., $C(p_{Y|X,S}, p_S, p_{\tilde{S}|S})$) is equal to that of channel $p_{Y,S|U}$, where

$$p_{Y,S|U}(y, s|u) = \sum_{\tilde{s}\in\tilde{\mathcal{S}}} p_S(s)p_{\tilde{S}|S}(\tilde{s}|s)p_{Y|X,S}(y|\psi(u, \tilde{s}), s), \qquad u \in \mathcal{U}, y \in \mathcal{Y}, s \in \mathcal{S}.$$

According to [6, Th. 4.5.1], p_U is a capacity-achieving input distribution of channel $p_{Y,S|U}$ (i.e., p_U is a maximizer of the optimization problem in (1)) if and only if there exists some number C such that

$$D(p_{Y,S|U}(\cdot, \cdot|u)\|p_{Y,S}) = C, \quad u \in \mathcal{U} \text{ with } p_U(u) > 0,$$
$$D(p_{Y,S|U}(\cdot, \cdot|u)\|p_{Y,S}) \leq C, \quad u \in \mathcal{U} \text{ with } p_U(u) = 0;$$

furthermore, the number C is equal to $C(p_{Y|X,S}, p_S, p_{\tilde{S}|S})$. In view of (3), we have

$$p_{Y,S|U}(y, s|u) = p_{Y,S|X}(y, s|u), \quad u \in \mathcal{X}, y \in \mathcal{Y}, s \in \mathcal{S}.$$

Let $p_{\hat{X}}$ be a capacity-achieving input distribution of channel $p_{Y,S|X}$ (i.e, $p_{\hat{X}}$ is a maximizer of the optimization problem in (4)). Define

$$p_{\hat{U}}(u) = \begin{cases} p_{\hat{X}}(u), & u \in \mathcal{X}, \\ 0, & \text{otherwise}. \end{cases} \tag{11}$$

It is clear that $C(p_{Y|X,S}, p_S, p_{\tilde{S}|S}) = \underline{C}(p_{Y|X,S}, p_S)$ if and only if $p_{\hat{U}}$ is a capacity-achieving input distribution of channel $p_{Y,S|U}$.

Now consider the special case where $p_{\tilde{S}|S}$ is a generalized erasure channel with erasure probability ϵ, and define

$$D_{GE}(p_U, \epsilon, u) = D(p_{Y,S|U}(\cdot, \cdot|u)\|p_{Y,S}) \tag{12}$$

to stress the dependence of $D(p_{Y,S|U}(\cdot, \cdot|u)\|p_{Y,S})$ on p_U, ϵ, and u. It can be verified that

$$\begin{aligned} &p_{Y,S|U}(y, s|u) \\ &= \sum_{\tilde{s}\in\mathcal{S}\cup\{*\}} p_S(s)p_{\tilde{S}^{(\epsilon)}|S}(\tilde{s}|s)p_{Y|X,S}(y|\psi(u, \tilde{s}), s) \\ &= p_S(s)\epsilon p_{Y|X,S}(y|\psi(u, *), s) \\ &\quad + p_S(s)(1-\epsilon)p_{Y|X,S}(y|\psi(u, s), s) \\ &= p_S(s)(p_{Y|X,S}(y|\psi(u, s), s) + \epsilon\delta(u, y, s)), \end{aligned} \tag{13}$$

[2] Numerical simulations suggest that this universal upper bound is not tight. Determining the exact universal erasure probability threshold remains an open problem.

where

$$\begin{aligned}\delta(u,y,s) = p_{Y|X,S}(y|\psi(u,*),s) - p_{Y|X,S}(y|\psi(u,s),s),&\\ u\in\mathcal{U}, y\in\mathcal{Y}, s\in\mathcal{S}.&\end{aligned}\tag{14}$$

Since $|\mathcal{X}| = 2$, there is no loss of generality in assuming that [7, Th. 2]

$$p_{\hat{X}}(x) > e^{-1}, \quad x\in\mathcal{X}.\tag{15}$$

To the end of proving Lemma 1, it suffices to show that, for $\epsilon\in[1-e^{-1},1]$,

$$\begin{aligned}D_{GE}(p_{\hat{U}},\epsilon,u) &= \underline{C}(p_{Y|X,S},p_S), \quad u\in\mathcal{X},\\ D_{GE}(p_{\hat{U}},\epsilon,u) &\le \underline{C}(p_{Y|X,S},p_S), \quad \text{otherwise.}\end{aligned}$$

Clearly, $p_{\hat{U}}$ is a capacity-achieving input distribution of channel $p_{Y,S|U}$ when $\epsilon = 1$. Therefore, we have[3]

$$D_{GE}(p_{\hat{U}},1,u) = \underline{C}(p_{Y|X,S},p_S), \quad u\in\mathcal{X},\tag{16}$$

$$D_{GE}(p_{\hat{U}},1,u) \le \underline{C}(p_{Y|X,S},p_S), \quad \text{otherwise.}\tag{17}$$

Note that

$$\begin{aligned}&D_{GE}(p_{\hat{U}},\epsilon,u)\\ &= \sum_{y\in\mathcal{Y},s\in\mathcal{S}} p_{Y,S|U}(y,s|u)\log\frac{p_{Y,S|U}(y,s|u)}{\sum_{u'\in\mathcal{U}} p_{\hat{U}}(u')p_{Y,S|U}(y,s|u')}\\ &= \sum_{y\in\mathcal{Y},s\in\mathcal{S}} p_S(s)(p_{Y|X,S}(y|\psi(u,s),s)+\epsilon\delta(u,y,s))\\ &\quad\times\log\frac{p_{Y|X,S}(y|\psi(u,s),s)+\epsilon\delta(u,y,s)}{\sum_{u'\in\mathcal{U}} p_{\hat{U}}(u')(p_{Y|X,S}(y|\psi(u',s),s)+\epsilon\delta(u',y,s))} \qquad (18)\\ &= \sum_{y\in\mathcal{Y},s\in\mathcal{S}} p_S(s)(p_{Y|X,S}(y|\psi(u,s),s)+\epsilon\delta(u,y,s))\\ &\quad\times\log\frac{p_{Y|X,S}(y|\psi(u,s),s)+\epsilon\delta(u,y,s)}{\sum_{x\in\mathcal{X}} p_{\hat{X}}(x)p_{Y|X,S}(y|x,s)},\\ &\qquad\qquad \epsilon\in[0,1], u\in\mathcal{U}, \qquad (19)\end{aligned}$$

where (18) is due to (13), and (19) is due to (3) and (11). Moreover,

$$\begin{aligned}&\frac{\partial}{\partial\epsilon}D_{GE}(p_{\hat{U}},\epsilon,u)\\ &= \sum_{y\in\mathcal{Y},s\in\mathcal{S}} p_S(s)\delta(u,y,s)\\ &\quad\times\left(\log\frac{p_{Y|X,S}(y|\psi(u,s),s)+\epsilon\delta(u,y,s)}{\sum_{x\in\mathcal{X}} p_{\hat{X}}(x)p_{Y|X,S}(y|x,s)}+1\right)\\ &= \sum_{y\in\mathcal{Y},s\in\mathcal{S}} p_S(s)\delta(u,y,s)\\ &\quad\times\log\frac{p_{Y|X,S}(y|\psi(u,s),s)+\epsilon\delta(u,y,s)}{\sum_{x\in\mathcal{X}} p_{\hat{X}}(x)p_{Y|X,S}(y|x,s)}\\ &\quad+\sum_{s\in\mathcal{S}} p_S(s)\sum_{y\in\mathcal{Y}}\delta(u,y,s)\\ &= \sum_{y\in\mathcal{Y},s\in\mathcal{S}} p_S(s)\delta(u,y,s)\\ &\quad\times\log\frac{p_{Y|X,S}(y|\psi(u,s),s)+\epsilon\delta(u,y,s)}{\sum_{x\in\mathcal{X}} p_{\hat{X}}(x)p_{Y|X,S}(y|x,s)},\\ &\qquad\qquad \epsilon\in[0,1], u\in\mathcal{U}. \qquad (20)\end{aligned}$$

Define

$$\mathcal{G}_\delta = \{u\in\mathcal{U} : \delta(u,y,s) = 0 \text{ for all } y\in\mathcal{Y} \text{ and } s\in\mathcal{S}\}.\tag{21}$$

In light of (19),

$$D_{GE}(p_{\hat{U}},\epsilon,u) = D_{GE}(p_{\hat{U}},1,u), \quad \epsilon\in[0,1], u\in\mathcal{G}_\delta.\tag{22}$$

For any $u\in\mathcal{U}\backslash\mathcal{G}_\delta$, there must exist some $y\in\mathcal{Y}$ and $s\in\mathcal{S}$ such that $\delta(u,y,s)\neq 0$; furthermore, since $|\mathcal{X}| = 2$, we have

$$\begin{aligned}\delta(u,y,s) > 0 \Longrightarrow\ & p_{Y|X,S}(y|\psi(u,s),s)+\epsilon\delta(u,y,s)\\ &= b(y,s)+\epsilon(a(y,s)-b(y,s)), \qquad (23)\\ \delta(u,y,s) < 0 \Longrightarrow\ & p_{Y|X,S}(y|\psi(u,s),s)+\epsilon\delta(u,y,s)\\ &= a(y,s)+\epsilon(b(y,s)-a(y,s)), \qquad (24)\end{aligned}$$

where

$$\begin{aligned}a(y,s) &= \max_{x\in\mathcal{X}} p_{Y|X,S}(y|x,s),\\ b(y,s) &= \min_{x\in\mathcal{X}} p_{Y|X,S}(y|x,s).\end{aligned}$$

Continuing from (20),

$$\begin{aligned}\frac{\partial}{\partial\epsilon}D_{GE}(p_{\hat{U}},\epsilon,u) &= \sum_{y\in\mathcal{Y},s\in\mathcal{S}} p_S(s)\delta(u,y,s)\\ &\quad\times\log\frac{p_{Y|X,S}(y|\psi(u,s),s)+\epsilon\delta(u,y,s)}{\sum_{x\in\mathcal{X}} p_{\hat{X}}(x)p_{Y|X,S}(y|x,s)}\\ &\ge \sum_{s\in\mathcal{S}} p_S(s)\sum_{y\in\mathcal{Y}:\delta(u,y,s)>0}\delta(u,y,s)\\ &\quad\times\log\frac{p_{Y|X,S}(y|\psi(u,s),s)+\epsilon\delta(u,y,s)}{(1-e^{-1})a(y,s)+e^{-1}b(y,s)}\\ &\quad+\sum_{s\in\mathcal{S}} p_S(s)\sum_{y\in\mathcal{Y}:\delta(u,y,s)<0}\delta(u,y,s)\\ &\quad\times\log\frac{p_{Y|X,S}(y|\psi(u,s),s)+\epsilon\delta(u,y,s)}{e^{-1}a(y,s)+(1-e^{-1})b(y,s)} \qquad (25)\\ &= \sum_{s\in\mathcal{S}} p_S(s)\sum_{y\in\mathcal{Y}:\delta(u,y,s)>0}\delta(u,y,s)\\ &\quad\times\log\frac{b(y,s)+\epsilon(a(y,s)-b(y,s))}{(1-e^{-1})a(y,s)+e^{-1}b(y,s)}\\ &\quad+\sum_{s\in\mathcal{S}} p_S(s)\sum_{y\in\mathcal{Y}:\delta(u,y,s)<0}\delta(u,y,s)\\ &\quad\times\log\frac{a(y,s)+\epsilon(b(y,s)-a(y,s))}{e^{-1}a(y,s)+(1-e^{-1})b(y,s)} \qquad (26)\\ &\ge \sum_{s\in\mathcal{S}} p_S(s)\sum_{y\in\mathcal{Y}:\delta(u,y,s)>0}\delta(u,y,s)\\ &\quad\times\log\frac{(1-e^{-1})a(y,s)+e^{-1}b(y,s)}{(1-e^{-1})a(y,s)+e^{-1}b(y,s)}\\ &\quad+\sum_{s\in\mathcal{S}} p_S(s)\sum_{y\in\mathcal{Y}:\delta(u,y,s)<0}\delta(u,y,s)\\ &\quad\times\log\frac{e^{-1}a(y,s)+(1-e^{-1})b(y,s)}{e^{-1}a(y,s)+(1-e^{-1})b(y,s)}\\ &= 0, \quad \epsilon\in[1-e^{-1},1], u\in\mathcal{U}, \qquad (27)\end{aligned}$$

[3] The inequality in (17) is in fact an equality.

where (25) is due to (15), and (26) is due to (23) and (24). Combining (16), (17), (22), (27), and the fact $\mathcal{X} \subseteq \mathcal{G}_\delta$ yields the desired result. ■

Recall [3, p. 112] that $p_{\tilde{S}_1|S}$ (with input alphabet $\mathcal{S}$ and output alphabet $\tilde{\mathcal{S}}_1$) is said to be a stochastically degraded version of $p_{\tilde{S}_2|S}$ (with input alphabet $\mathcal{S}$ and output alphabet $\tilde{\mathcal{S}}_2$) if there exists $p_{\tilde{S}_1|\tilde{S}_2}$ satisfying

$$p_{\tilde{S}_1|S}(\tilde{s}_1|s) = \sum_{\tilde{s}_2 \in \tilde{\mathcal{S}}_2} p_{\tilde{S}_2|S}(\tilde{s}_2|s) p_{\tilde{S}_1|\tilde{S}_2}(\tilde{s}_1|\tilde{s}_2), \qquad s \in \mathcal{S}, \tilde{s}_1 \in \tilde{\mathcal{S}}_1. \quad (28)$$

We can write (28) equivalently as

$$p_{\tilde{S}_1|S} = p_{\tilde{S}_2|S} p_{\tilde{S}_1|\tilde{S}_2}$$

by viewing $p_{\tilde{S}_1|S}$, $p_{\tilde{S}_2|S}$, and $p_{\tilde{S}_1|\tilde{S}_2}$ as probability transition matrices.

The following result is obvious and its proof is omitted.

Lemma 2: If $p_{\tilde{S}_1|S}$ is a stochastically degraded version of $p_{\tilde{S}_2|S}$, then

$$C(p_{Y|X,S}, p_S, p_{\tilde{S}_1|S}) \leq C(p_{Y|X,S}, p_S, p_{\tilde{S}_2|S}).$$

Next we extend Lemma 1 to the general case by characterizing the condition under which $p_{\tilde{S}|S}$ is a stochastically degraded version of $p_{\tilde{S}^{(\epsilon)}_{GE}|S}$.

Lemma 3: $p_{\tilde{S}|S}$ is a stochastically degraded version of $p_{\tilde{S}^{(\epsilon)}_{GE}|S}$ if and only if

$$\sum_{\tilde{s} \in \tilde{S}} \min_{s \in \mathcal{S}} p_{\tilde{S}|S}(\tilde{s}|s) \geq \epsilon. \quad (29)$$

Proof: The problem boils down to finding a necessary and sufficient condition for the existence of $p_{\tilde{S}|\tilde{S}^{(\epsilon)}_{GE}}$ such that

$$p_{\tilde{S}|S}(\tilde{s}|s) = \sum_{\tilde{s}' \in \mathcal{S} \cup \{*\}} p_{\tilde{S}^{(\epsilon)}_{GE}|S}(\tilde{s}'|s) p_{\tilde{S}|\tilde{S}^{(\epsilon)}_{GE}}(\tilde{s}|\tilde{s}'), \qquad s \in \mathcal{S}, \tilde{s} \in \tilde{\mathcal{S}}. \quad (30)$$

It suffices to consider the case $\epsilon \in [0, 1)$ since Lemma 3 is trivially true when $\epsilon = 1$. Note that

$$\begin{aligned}&\sum_{\tilde{s}' \in \mathcal{S} \cup \{*\}} p_{\tilde{S}^{(\epsilon)}_{GE}|S}(\tilde{s}'|s) p_{\tilde{S}|\tilde{S}^{(\epsilon)}_{GE}}(\tilde{s}|\tilde{s}') \\ &= (1 - \epsilon) p_{\tilde{S}|\tilde{S}^{(\epsilon)}_{GE}}(\tilde{s}|s) + \epsilon p_{\tilde{S}|\tilde{S}^{(\epsilon)}_{GE}}(\tilde{s}|*), \quad s \in \mathcal{S}, \tilde{s} \in \tilde{\mathcal{S}}. \quad (31)\end{aligned}$$

Combining (30) and (31) gives

$$p_{\tilde{S}|\tilde{S}^{(\epsilon)}_{GE}}(\tilde{s}|s) = \frac{p_{\tilde{S}|S}(\tilde{s}|s) - \epsilon p_{\tilde{S}|\tilde{S}^{(\epsilon)}_{GE}}(\tilde{s}|*)}{1 - \epsilon}, \quad s \in \mathcal{S}, \tilde{s} \in \tilde{\mathcal{S}}. \quad (32)$$

In light of (32),

$$\begin{aligned}&\sum_{\tilde{s} \in \tilde{\mathcal{S}}} p_{\tilde{S}|\tilde{S}^{(\epsilon)}_{GE}}(\tilde{s}|s) = 1, \quad s \in \mathcal{S}, \\ &\quad \Longleftrightarrow \sum_{\tilde{s} \in \tilde{\mathcal{S}}} p_{\tilde{S}|\tilde{S}^{(\epsilon)}_{GE}}(\tilde{s}|*) = 1, \\ &p_{\tilde{S}|\tilde{S}^{(\epsilon)}_{GE}}(\tilde{s}|s) \geq 0, \quad s \in \mathcal{S}, \tilde{s} \in \tilde{\mathcal{S}}, \\ &\quad \Longleftrightarrow \min_{s \in \mathcal{S}} p_{\tilde{S}|S}(\tilde{s}|s) \geq \epsilon p_{\tilde{S}|\tilde{S}^{(\epsilon)}_{GE}}(\tilde{s}|*), \quad \tilde{s} \in \tilde{\mathcal{S}}. \quad (33)\end{aligned}$$

It can be readily seen that the existence of conditional distribution $p_{\tilde{S}|\tilde{S}^{(\epsilon)}_{GE}}$ satisfying (30) is equivalent to the existence of probability vector $(p_{\tilde{S}|\tilde{S}^{(\epsilon)}_{GE}}(\tilde{s}|*))_{\tilde{s} \in \tilde{\mathcal{S}}}$ satisfying (33). Clearly, (29) is a necessary and sufficient condition for the existence of such $(p_{\tilde{S}|\tilde{S}^{(\epsilon)}_{GE}}(\tilde{s}|*))_{\tilde{s} \in \tilde{\mathcal{S}}}$. ■

Theorem 3: For any binary-input channel $p_{Y|X,S}$, state distribution p_S, and side channel $p_{\tilde{S}|S}$,

$$C(p_{Y|X,S}, p_S, p_{\tilde{S}|S}) = \underline{C}(p_{Y|X,S}, p_S)$$

if

$$\sum_{\tilde{s} \in \tilde{S}} \min_{s \in \mathcal{S}} p_{\tilde{S}|S}(\tilde{s}|s) \geq 1 - e^{-1}. \quad (34)$$

Proof: In view of Lemmas 1, 2, and 3, we have

$$C(p_{Y|X,S}, p_S, p_{\tilde{S}|S}) \leq \underline{C}(p_{Y|X,S}, p_S) \quad (35)$$

if (34) is satisfied. Combining (6) and (35) completes the proof of Theorem 3. ■

Now we proceed to prove Theorem 1 by translating (34) (which is a condition on $p_{\tilde{S}|S}$ that is universal for all binary-input channels and state distributions) to an upper bound on $I(S; \tilde{S})$. This upper bound, however, depends inevitably on the state distribution.

For any $p_{\tilde{S}|S}$ violating (34) (i.e, $\sum_{\tilde{s} \in \tilde{S}} \min_{s \in \mathcal{S}} p_{\tilde{S}|S}(\tilde{s}|s) < 1 - e^{-1}$), we have

$$\begin{aligned} I(S; \tilde{S}) &\geq \frac{1}{2}\left(\sum_{s \in \mathcal{S}, \tilde{s} \in \tilde{\mathcal{S}}} p_S(s) \left| p_{\tilde{S}}(\tilde{s}) - p_{\tilde{S}|S}(\tilde{s}|s) \right| \right)^2 \quad (36) \\ &\geq \frac{1}{2}\left(\sum_{\tilde{s} \in \tilde{\mathcal{S}}} p_S(s(\tilde{s})) \left| p_{\tilde{S}}(\tilde{s}) - p_{\tilde{S}|S}(\tilde{s}|s(\tilde{s})) \right| \right)^2 \\ &\geq \frac{1}{2}\left(\rho \sum_{\tilde{s} \in \tilde{\mathcal{S}}} \left| p_{\tilde{S}}(\tilde{s}) - p_{\tilde{S}|S}(\tilde{s}|s(\tilde{s})) \right| \right)^2 \\ &\geq \frac{1}{2}\left(\rho \left| \sum_{\tilde{s} \in \tilde{\mathcal{S}}} p_{\tilde{S}}(\tilde{s}) - \sum_{\tilde{s} \in \tilde{\mathcal{S}}} p_{\tilde{S}|S}(\tilde{s}|s(\tilde{s})) \right| \right)^2 \\ &> \frac{\rho^2}{2e^2}, \end{aligned}$$

where (36) is due to Pinsker's inequality [8, p. 44], and $s(\tilde{s})$ is a minimizer of $\min_{s \in \mathcal{S}} p_{\tilde{S}|S}(\tilde{s}|s)$, $\tilde{s} \in \tilde{\mathcal{S}}$. As a consequence, (34) must hold if $I(S; \tilde{S}) \leq \frac{\rho^2}{2e^2}$. This completes the proof of Theorem 1.

III. Proof of Theorem 2

First consider the special case where $p_{\tilde{S}|S}$ is a generalized symmetric channel (with crossover probability $q \in [0, \frac{1}{|\mathcal{S}|}]$) defined as

$$p_{\tilde{S}^{(q)}_{GS}|S}(\tilde{s}|s) = \begin{cases} 1 - (|\mathcal{S}| - 1)q, & \tilde{s} = s, \\ q, & \text{otherwise}, \end{cases} \quad s \in \mathcal{S}, \tilde{s} \in \mathcal{S}.$$

Lemma 4: $C'(p_{Y|X,S}, p_S, p_{\tilde{S}^{(q)}_{GS}|S}) = \overline{C}(p_{Y|X,S}, p_S)$ if and only if

$$\min_{x \in \mathcal{X}_+, s \in \mathcal{S}} \frac{p_{\hat{X}|S}(x|s)}{\sum_{s' \in \mathcal{S}} p_{\hat{X}|S}(x|s')} \geq q \quad (37)$$

for some $p_{\hat{X}|S} \in \mathcal{P}$, where $\mathcal{P}$ denotes the set of maximizers of the optimization problem in (5), and $\mathcal{X}_+ = \{x \in \mathcal{X} : \sum_{s\in\mathcal{S}} p_{\hat{X}|S}(x|s) > 0\}$.

Proof: Clearly, $C'(p_{Y|X,S}, p_S, p_{\tilde{S}^{(q)}_{GS}|S}) = \overline{C}(p_{Y|X,S}, p_S)$ if and only if there exists $p_{\hat{X}|S} \in \mathcal{P}$ that is a stochastically degraded version of $p_{\tilde{S}^{(q)}_{GS}|S}$. When $q = \frac{1}{|\mathcal{S}|}$, (37) is equivalent to the desired condition that $\hat{X}$ needs to be independent of S. When $q \in [0, \frac{1}{|\mathcal{S}|})$, $p_{\tilde{S}^{(q)}_{GS}|S}$ is invertible and

$$p^{-1}_{\tilde{S}^{(q)}_{GS}|S} = \begin{pmatrix} \frac{q-1}{|\mathcal{S}|q-1} & \frac{q}{|\mathcal{S}|q-1} & \cdots & \frac{q}{|\mathcal{S}|q-1} \\ \frac{q}{|\mathcal{S}|q-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{q}{|\mathcal{S}|q-1} \\ \frac{q}{|\mathcal{S}|q-1} & \cdots & \frac{q}{|\mathcal{S}|q-1} & \frac{q-1}{|\mathcal{S}|q-1} \end{pmatrix}. \tag{38}$$

The problem boils down to finding a necessary and sufficient condition under which $p^{-1}_{\tilde{S}^{(q)}_{GS}|S} p_{\hat{X}|S}$ is a valid probability transition matrix (i.e., all entries are non-negative and the sum of each row vector is equal to 1). Note that

$$\begin{aligned} p^{-1}_{\tilde{S}^{(q)}_{GS}|S} p_{\hat{X}|S} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} &= p^{-1}_{\tilde{S}^{(q)}_{GS}|S} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \\ &= p^{-1}_{\tilde{S}^{(q)}_{GS}|S} p_{\tilde{S}^{(q)}_{GS}|S} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}. \end{aligned} \tag{39}$$

Moreover, all entries of $p^{-1}_{\tilde{S}^{(q)}_{GS}|S} p_{\hat{X}|S}$ are non-negative if and only if

$$\frac{-p_{\hat{X}|S}(x|s) + q\sum_{s'\in\mathcal{S}} p_{\hat{X}|S}(x|s')}{|\mathcal{S}|q-1} \geq 0, \quad x \in \mathcal{S}, s \in \mathcal{S},$$

which is equivalent to (37). ■

The following result is obvious and its proof is omitted.

Lemma 5: If $p_{\tilde{S}_1|S}$ is a stochastically degraded version of $p_{\tilde{S}_2|S}$, then

$$C'(p_{Y|X,S}, p_S, p_{\tilde{S}_1|S}) \leq C'(p_{Y|X,S}, p_S, p_{\tilde{S}_2|S}).$$

Lemma 6: $p_{\tilde{S}^{(q)}_{GS}|S}$ is a stochastically degraded version of $p_{\tilde{S}|S}$ if

$$\max_{s\in\mathcal{S},\hat{s}\in\mathcal{S}_+:s\neq\hat{s}} \frac{p_{\hat{S}|S}(\hat{s}|s)}{\sum_{s'\in\mathcal{S}} p_{\hat{S}|S}(\hat{s}|s')} \leq q, \tag{40}$$

where $\hat{S}$ is the maximum likelihood estimate of S based on $\tilde{S}$, and $\mathcal{S}_+ = \{\hat{s} \in \mathcal{S} : \sum_{s\in\mathcal{S}} p_{\hat{S}|S}(\hat{s}|s) > 0\}$.

Proof: The case $q = \frac{1}{|\mathcal{S}|}$ is trivial. When $q \in [0, \frac{1}{|\mathcal{S}|})$, $p_{\tilde{S}^{(q)}_{GS}|S}$ is invertible and $p^{-1}_{\tilde{S}^{(q)}_{GS}|S}$ is given by (38). It can be shown (see the derivation of (39)) that the sum of each row of $p^{-1}_{\tilde{S}^{(q)}_{GS}|S} p_{\hat{S}|S}$ is equal to 1; moreover, the off-diagonal entries of $p^{-1}_{\tilde{S}^{(q)}_{GS}|S} p_{\hat{S}|S}$ are non-positive if and only if

$$\frac{-p_{\hat{S}|S}(\hat{s}|s) + q\sum_{s'\in\mathcal{S}} p_{\hat{S}|S}(\hat{s}|s')}{|\mathcal{S}|q-1} \leq 0, \qquad s \in \mathcal{S}, \hat{s} \in \mathcal{S}_+ : s \neq \hat{s},$$

which is equivalent to (40). Therefore, (40) ensures that $p^{-1}_{\tilde{S}^{(q)}_{GS}|S} p_{\hat{S}|S}$ is a non-singular M-matrix, which in turn ensures that $p^{-1}_{\hat{S}|S} p_{\tilde{S}^{(q)}_{GS}|S}$ exists and is a non-negative matrix [9]. Hence, if (40) is satisfied, then $p^{-1}_{\hat{S}|S} p_{\tilde{S}^{(q)}_{GS}|S}$ is a valid probability transition matrix (the requirement that the entries in each row of $p^{-1}_{\hat{S}|S} p_{\tilde{S}^{(q)}_{GS}|S}$ add up to 1 is automatically satisfied), which implies that $p_{\tilde{S}^{(q)}_{GS}|S}$ is a stochastically degraded version of $p_{\hat{S}|S}$ (and consequently a stochastically degraded version of $p_{\tilde{S}|S}$). ■

Theorem 4: For any binary-input channel $p_{Y|X,S}$, state distribution p_S, and side channel $p_{\tilde{S}|S}$,

$$C'(p_{Y|X,S}, p_S, p_{\tilde{S}|S}) = \overline{C}(p_{Y|X,S}, p_S)$$

if

$$\max_{s\in\mathcal{S},\hat{s}\in\mathcal{S}_+:s\neq\hat{s}} \frac{p_{\hat{S}|S}(\hat{s}|s)}{\sum_{s'\in\mathcal{S}} p_{\hat{S}|S}(\hat{s}|s')} \leq \frac{1}{(|\mathcal{S}|-1)e - |\mathcal{S}| + 2}, \tag{41}$$

where $\hat{S}$ is the maximum likelihood estimate of S based on $\tilde{S}$.

Proof: Since $|\mathcal{X}| = 2$, it follows from [7, Th. 2] that there exists $p_{\hat{X}|S} \in \mathcal{P}$ satisfying

$$p_{\hat{X}|S}(x|s) > e^{-1}, \quad x \in \mathcal{X}, s \in \mathcal{S}.$$

For such $p_{\hat{X}|S}$,

$$\begin{aligned} \min_{x\in\mathcal{X}_+,s\in\mathcal{S}} \frac{p_{\hat{X}|S}(x|s)}{\sum_{s'\in\mathcal{S}} p_{\hat{X}|S}(x|s')} &\geq \frac{e^{-1}}{e^{-1} + (|\mathcal{S}|-1)(1-e^{-1})} \\ &= \frac{1}{(|\mathcal{S}|-1)e - |\mathcal{S}| + 2}. \end{aligned}$$

In view of of Lemmas 4, 5, and 6, we have

$$C'(p_{Y|X,S}, p_S, p_{\tilde{S}|S}) \geq \overline{C}(p_{Y|X,S}, p_S) \tag{42}$$

if (41) is satisfied. Combining (6) and (42) completes the proof of Theorem 4. ■

Now we are in a position to prove Theorem 2. Let $\hat{S}$ and $\hat{S}'$ denote respectively the maximum likelihood estimate and the maximum *a posteriori* estimate of S based on $\tilde{S}$. According to [10, Th. 11],

$$\mathbb{P}(S \neq \hat{S}') \leq \frac{H(S|\tilde{S})}{2\log 2}. \tag{43}$$

It can be verified that

$$\begin{aligned}\sum_{s,\hat{s}\in\mathcal{S}:s\neq\hat{s}} p_{\hat{S}|S}(\hat{s}|s) &\le \sum_{s,\hat{s}\in\mathcal{S}:s\neq\hat{s}} p_{\hat{S}'|S}(\hat{s}|s)\\ &\le \frac{1}{\rho}\sum_{s,\hat{s}\in\mathcal{S}:s\neq\hat{s}} p_S(s)p_{\hat{S}'|S}(\hat{s}|s)\\ &= \frac{\mathbb{P}(S\neq\hat{S}')}{\rho}. \end{aligned}\tag{44}$$

Substituting (43) into (44) yields

$$\sum_{s,\hat{s}\in\mathcal{S}:s\neq\hat{s}} p_{\hat{S}|S}(\hat{s}|s) \le \hbar \triangleq \frac{H(S|\tilde{S})}{2\rho\log 2}. \tag{45}$$

Note that

$$\max_{s\in\mathcal{S},\hat{s}\in\mathcal{S}_+:s\neq\hat{s}} \frac{p_{\hat{S}|S}(\hat{s}|s)}{\sum_{s'\in\mathcal{S}} p_{\hat{S}|S}(\hat{s}|s')} \le \frac{\hbar}{\hbar+\mathbb{I}(\hbar\le 1)}. \tag{46}$$

Indeed, (46) is trivially true when $\hbar > 1$; moreover, when $\hbar \le 1$,

$$\begin{aligned}&\max_{s\in\mathcal{S},\hat{s}\in\mathcal{S}_+:s\neq\hat{s}} \frac{p_{\hat{S}|S}(\hat{s}|s)}{\sum_{s'\in\mathcal{S}} p_{\hat{S}|S}(\hat{s}|s')}\\ &\le \max_{s\in\mathcal{S},\hat{s}\in\mathcal{S}_+:s\neq\hat{s}} \frac{p_{\hat{S}|S}(\hat{s}|s)}{p_{\hat{S}|S}(\hat{s}|s)+p_{\hat{S}|S}(\hat{s}|\hat{s})}\\ &= \max_{s\in\mathcal{S},\hat{s}\in\mathcal{S}_+:s\neq\hat{s}} \frac{p_{\hat{S}|S}(\hat{s}|s)}{p_{\hat{S}|S}(\hat{s}|s)+1-\sum_{\hat{s}'\in\mathcal{S}:\hat{s}'\neq\hat{s}} p_{\hat{S}|S}(\hat{s}'|\hat{s})}\\ &\le \max_{s\in\mathcal{S},\hat{s}\in\mathcal{S}_+:s\neq\hat{s}} \frac{p_{\hat{S}|S}(\hat{s}|s)}{2p_{\hat{S}|S}(\hat{s}|s)+1-\hbar} &(47)\\ &\le \frac{\hbar}{\hbar+1}, &(48)\end{aligned}$$

where (47) and (48) are due to (45). In view of Theorem 4, It suffices to have

$$\frac{\hbar}{\hbar+\mathbb{I}(\hbar\le 1)} \le \frac{1}{(|\mathcal{S}|-1)e-|\mathcal{S}|+2}. \tag{49}$$

Note that (49) is not satisfied when $\hbar > 1$ since its left-hand side is equal to 1 whereas its right-hand side is strictly less than 1 ($\hbar > 1$ implies $|\mathcal{S}| \ge 2$). When $\hbar \le 1$, we can rewrite (49) as[4]

$$\hbar \le \frac{1}{(|\mathcal{S}|-1)(e-1)},$$

which is exactly the desired result. This completes the proof of Theorem 2.

In Appendix A, we give an alternative proof of Theorem 2 with a different threshold on $H(S|\tilde{S})$.

IV. Extension and Discussion

A. Extension of Theorem 1

It is interesting to know to what extent Theorem 1 can be extended beyond the binary-input case. This subsection is largely devoted to answering this question. For any $p_{Y|X,S}$ and p_S, define

$$\begin{aligned}\underline{\epsilon}(p_{Y|X,S},p_S) &= \min\{\epsilon\in[0,1]: C(p_{Y|X,S},p_S,p_{\tilde{S}^{(\epsilon)}_{GE}|S})\\ &\qquad = \underline{C}(p_{Y|X,S},p_S)\},\\ \underline{q}(p_{Y|X,S},p_S) &= \min\{q\in[0,\frac{1}{|\mathcal{S}|}]: C(p_{Y|X,S},p_S,p_{\tilde{S}^{(q)}_{GS}|S})\\ &\qquad = \underline{C}(p_{Y|X,S},p_S)\}.\end{aligned}$$

Proposition 1: 1) There exists $\alpha(p_{Y|X,S},p_S) > 0$ such that $C(p_{Y|X,S},p_S,p_{\tilde{S}|S}) = \underline{C}(p_{Y|X,S},p_S)$ for all $p_{\tilde{S}|S}$ satisfying $I(S;\tilde{S}) \le \alpha(p_{Y|X,S},p_S)$ if and only if $\underline{\epsilon}(p_{Y|X,S},p_S) < 1$.

2) $\underline{\epsilon}(p_{Y|X,S},p_S) < 1$ if and only if

$$\begin{aligned}&\sum_{y\in\mathcal{Y},s\in\mathcal{S}} p_S(s)\delta(u,y,s)\\ &\qquad\times\log\frac{p_{Y|X,S}(y|\psi(u,*),s)}{\sum_{x\in\mathcal{X}} p_{\hat{X}}(x)p_{Y|X,S}(y|x,s)} > 0,\\ &\qquad u\in\mathcal{U}_+\backslash\mathcal{G}_\delta,\end{aligned}\tag{50}$$

where $\delta(u,y,s)$ and $\mathcal{G}_\delta$ are defined in (14) and (21), respectively, $p_{\hat{X}}$ is an arbitrary maximizer of the optimization problem in (4), and

$$\begin{aligned}\mathcal{U}_+ = \Bigg\{u\in\mathcal{U}: &\sum_{y\in\mathcal{Y},s\in\mathcal{S}} p_S(s)p_{Y|X,S}(y|\psi(u,*),s)\\ &\times\log\frac{p_{Y|X,S}(y|\psi(u,*),s)}{\sum_{x\in\mathcal{X}} p_{\hat{X}}(x)p_{Y|X,S}(y|x,s)} = \underline{C}(p_{Y|X,S},p_S)\Bigg\}.\end{aligned}$$

Remark: All maximizers of the optimization problem in (4) give rise to the same $\sum_{x\in\mathcal{X}} p_{\hat{X}}(x)p_{Y|X,S}(y|x,s)$, $y\in\mathcal{Y}$, $s\in\mathcal{S}$ [6, p. 96, corollary 2].

Proof: The first statement can be easily extracted from the proof of Theorem 1.

Now we proceed to prove the second statement. First recall the definitions of $D_{GE}(p_U,\epsilon,u)$ and $p_{\hat{U}}$ in (12) and (11), respectively. Since $p_{\hat{U}}$ is a capacity-achieving input distribution of channel $p_{Y,S|U}$ when $\epsilon = 1$, we must have

$$\begin{aligned}D_{GE}(p_{\hat{U}},1,u) &= \underline{C}(p_{Y|X,S},p_S),\quad u\in\mathcal{U}\text{ with }p_{\hat{U}}(u)>0,\\ D_{GE}(p_{\hat{U}},1,u) &\le \underline{C}(p_{Y|X,S},p_S),\quad u\in\mathcal{U}\text{ with }p_{\hat{U}}(u)=0,\end{aligned}$$

which, together with the fact $\mathcal{U}_+ = \{u\in\mathcal{U}: D_{GE}(p_{\hat{U}},1,u) = \underline{C}(p_{Y|X,S},p_S)\}$, implies

$$\begin{aligned}&\{u\in\mathcal{U}: p_{\hat{U}}(u)>0\}\subseteq\mathcal{U}_+, &(51)\\ &D_{GE}(p_{\hat{U}},1,u) = \underline{C}(p_{Y|X,S},p_S),\quad u\in\mathcal{U}_+, &(52)\\ &D_{GE}(p_{\hat{U}},1,u) < \underline{C}(p_{Y|X,S},p_S),\quad \text{otherwise}. &(53)\end{aligned}$$

It can be verified that

$$D_{GE}(p_{\hat{U}},\epsilon,u) = D_{GE}(p_{\hat{U}},1,u),\quad \epsilon\in[0,1], u\in\mathcal{G}_\delta. \tag{54}$$

Moreover, in view of (20), we can write (50) equivalently as

$$\frac{\partial}{\partial\epsilon}D_{GE}(p_{\hat{U}},\epsilon,u)\bigg|_{\epsilon=1} > 0,\quad u\in\mathcal{U}_+\backslash\mathcal{G}_\delta. \tag{55}$$

[4] Note that $\hbar \le \frac{1}{(|\mathcal{S}|-1)(e-1)}$ implies $\hbar \le 1$ when $|\mathcal{S}| \ge 2$. The case $|\mathcal{S}| = 1$ is trivial since $\hbar$ can only take the value 0.

According to (52)–(55), there exists $\epsilon(p_{Y|X,S},p_S)\in[0,1)$ such that

$$D_{GE}(p_{\hat{U}},\epsilon,u)=\underline{C}(p_{Y|X,S},p_S),\quad u\in\mathcal{U}_+\cap\mathcal{G}_\delta, \tag{56}$$
$$D_{GE}(p_{\hat{U}},\epsilon,u)\leq\underline{C}(p_{Y|X,S},p_S),\quad \text{otherwise} \tag{57}$$

for $\epsilon\geq\epsilon(p_{Y|X,S},p_S)$. In light of (51) and the fact $\{u\in\mathcal{U}:p_{\hat{U}}(u)>0\}\subseteq\mathcal{X}\subseteq\mathcal{G}_\delta$, we have

$$\{u\in\mathcal{U}:p_{\hat{U}}(u)>0\}\subseteq\mathcal{U}_+\cap\mathcal{G}_\delta. \tag{58}$$

Combining (56), (57), and (58) proves the “if" part of the second statement. Next we turn to the “only if" part of the second statement. Assuming the existence of $\epsilon(p_{Y|X,S},p_S)\in[0,1)$ such that $C(p_{Y|X,S},p_S,p_{\tilde{S}^{(\epsilon)}|S})=\underline{C}(p_{Y|X,S},p_S)$ for $\epsilon\geq\epsilon(p_{Y|X,S},p_S)$ (or equivalently $p_{\hat{U}}$ is a capacity-achieving input distribution of channel $p_{Y,S|U}$ for $\epsilon\geq\epsilon(p_{Y|X,S},p_S)$), we must have

$$D_{GE}(p_{\hat{U}},\epsilon,u)\leq\underline{C}(p_{Y|X,S},p_S),\quad \epsilon\geq\epsilon(p_{Y|X,S},p_S),u\in\mathcal{U}. \tag{59}$$

It can be verified that

$$\begin{aligned}&\frac{\partial^2}{\partial\epsilon^2}D_{GE}(p_{\hat{U}},\epsilon,u)\\&\quad=\sum_{y\in\mathcal{Y},s\in\mathcal{S}}\frac{p_S(s)\delta^2(u,y,s)}{p_{Y|X,S}(y|\psi(u,s),s)+\epsilon\delta(u,y,s)}\\&\quad>0,\quad\epsilon\in[0,1],u\in\mathcal{U}\backslash\mathcal{G}_\delta.\end{aligned} \tag{60}$$

Moreover, by the definition of $\mathcal{U}_+$,

$$D_{GE}(p_{\hat{U}},1,u)=\underline{C}(p_{Y|X,S},p_S),\quad u\in\mathcal{U}_+. \tag{61}$$

Note that (59), (60), and (61) hold simultaneously for $u\in\mathcal{U}_+\backslash\mathcal{G}_\delta$, from which (50) (or equivalently (55)) can be readily deduced. This completes the proof of Proposition 1. ■

As shown by the following example, the necessary and sufficient condition (50) is not always satisfied when $|\mathcal{X}|>2$. Let

$$p_{Y|X,S}(y|x,s)=\begin{cases}1, & (x,y,s)=(0,0,0)\text{ or }(1,1,1),\\0, & (x,y,s)=(0,1,0)\text{ or }(1,0,1),\\\frac{2}{5}, & (x,y,s)=(1,0,0)\text{ or }(0,1,1),\\\frac{3}{5}, & (x,y,s)=(1,1,0)\text{ or }(0,0,1),\\\frac{3}{10}, & (x,y,s)=(2,0,0),\\\frac{1}{5}, & (x,y,s)=(2,0,1),\\\frac{7}{10}, & (x,y,s)=(2,1,0),\\\frac{4}{5}, & (x,y,s)=(2,1,1),\end{cases} \tag{62}$$

$$p_S(0)=p_S(1)=\frac{1}{2}. \tag{63}$$

For this example, it can be verified that $\hat{u}\in\mathcal{U}_+\backslash\mathcal{G}_\delta$ and

$$\sum_{y\in\mathcal{Y},s\in\mathcal{S}}p_S(s)\delta(\hat{u},y,s)\log\frac{p_{Y|X,S}(y|\psi(\hat{u},*),s)}{\sum_{x\in\mathcal{X}}p_{\hat{X}}(x)p_{Y|X,S}(y|x,s)}<0,$$

where $\psi(\hat{u},\cdot)$ is given by $\psi(\hat{u},0)=2$, $\psi(\hat{u},1)=1$, and $\psi(\hat{u},*)=1$; indeed, Fig. 5 shows that $C(p_{Y|X,S},p_S,\mathrm{BEC}(\epsilon))>\underline{C}(p_{Y|X,S},p_S)$ for $\epsilon\in[0,1)$.

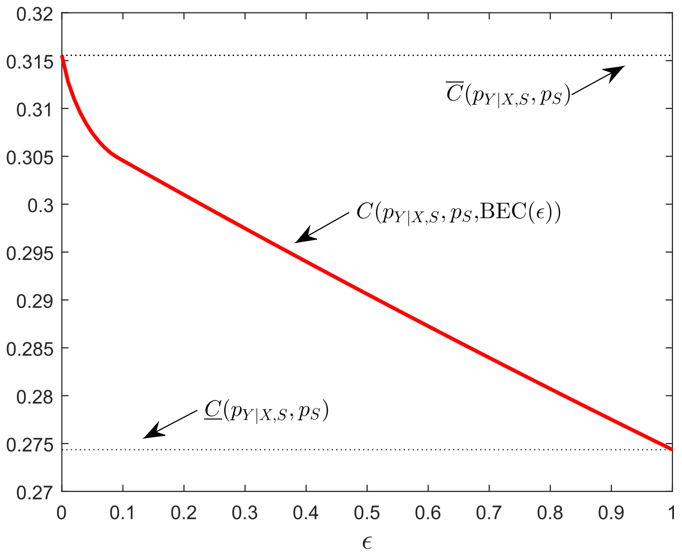

Fig. 5. Plot of $C(p_{Y|X,S},p_S,\mathrm{BEC}(\epsilon))$ against ϵ for $\epsilon\in[0,1]$, where $p_{Y|X,S}$ and p_S are given by (62) and (63), respectively.

The proof of Proposition 1 in fact suggests a strategy for computing $\underline{\epsilon}(p_{Y|X,S},p_S)$. Let $p_{\hat{X}}$ be an arbitrary maximizer of the optimization problem in (4) and define $p_{\hat{U}}$ according to (11). Note that

- $D_{GE}(p_{\hat{U}},1,u)\leq\underline{C}(p_{Y|X,S},p_S)$ for $u\in\mathcal{U}$ (see (52) and (53)),
- $D_{GE}(p_{\hat{U}},\epsilon,u)$ does not depend on ϵ for $u\in\mathcal{G}_\delta$ (see (54)),
- $D_{GE}(p_{\hat{U}},\epsilon,u)$ is a strictly convex function of ϵ for $u\in\mathcal{U}\backslash\mathcal{G}_\delta$ (see (60)).

Hence, for each $u\in\mathcal{U}$, there are three mutually exclusive cases.

1) $D_{GE}(p_{\hat{U}},0,u)\leq\underline{C}(p_{Y|X,S},p_S)$: We have $D_{GE}(p_{\hat{U}},\epsilon,u)\leq\underline{C}(p_{Y|X,S},p_S)$ for $\epsilon\in[\epsilon(u),1]$, where $\epsilon(u)=0$.
2) $D_{GE}(p_{\hat{U}},0,u)>D_{GE}(p_{\hat{U}},1,u)=\underline{C}(p_{Y|X,S},p_S)$ and $\frac{\partial}{\partial\epsilon}D_{GE}(p_{\hat{U}},\epsilon,u)\big|_{\epsilon=1}\leq0$ (this case can arise only when $|\mathcal{X}|>2$): We have $D_{GE}(p_{\hat{U}},0,u)>\underline{C}(p_{Y|X,S},p_S)$ for $\epsilon\in[0,\epsilon(u))$, where $\epsilon(u)=1$.
3) Otherwise: We have $D_{GE}(p_{\hat{U}},\epsilon,u)>\underline{C}(p_{Y|X,S},p_S)$ for $\epsilon\in[0,\epsilon(u))$ and $D_{GE}(p_{\hat{U}},\epsilon,u)\leq\underline{C}(p_{Y|X,S},p_S)$ for $\epsilon\in[\epsilon(u),1]$, where $\epsilon(u)$ is the unique solution of $D_{GE}(p_{\hat{U}},\epsilon,u)=\underline{C}(p_{Y|X,S},p_S)$ for $\epsilon\in(0,1)$.

It can be readily shown that

$$\underline{\epsilon}(p_{Y|X,S},p_S)=\max_{u\in\mathcal{U}}\epsilon(u). \tag{64}$$

We can compute $\underline{q}(p_{Y|X,S},p_S)$ in a similar way. Define

$$D_{GS}(p_U,q,u)=D(p_{Y,S|U}(\cdot,\cdot|u)\|p_{Y,S}),$$

where

$$p_{Y,S|U}(y,s|u)=p_S(s)(p_{Y|X,S}(y|\psi(u,s),s)+q\omega(u,y,s))$$

with

$$\begin{aligned}\omega(u, y, s) = &\sum_{\tilde{s}\in\mathcal{S}:\tilde{s}\neq s} p_{Y|X,S}(y|\psi(u, \tilde{s}), s)\\ &-(|\mathcal{S}| - 1)p_{Y|X,S}(y|\psi(u, s), s),\\ &\qquad u \in \mathcal{U}, y \in \mathcal{Y}, s \in \mathcal{S}.\end{aligned}$$

Again, let $p_{\hat{U}}$ be defined[5] according to (11). It can be verified that

$$\begin{aligned}&D_{GS}(p_{\hat{U}}, q, u)\\ &= \sum_{y\in\mathcal{Y},s\in\mathcal{S}} p_S(s)(p_{Y|X,S}(y|\psi(u, s), s) + q\omega(u, y, s))\\ &\quad\times \log \frac{p_{Y|X,S}(y|\psi(u, s), s) + q\omega(u, y, s)}{\sum_{x\in\mathcal{X}} p_{\hat{X}}(x)p_{Y|X,S}(y|x, s)},\\ &\qquad q \in [0, \frac{1}{|\mathcal{S}|}], u \in \mathcal{U},\\ &\frac{\partial}{\partial q} D_{GS}(p_{\hat{U}}, q, u)\\ &= \sum_{y\in\mathcal{Y},s\in\mathcal{S}} p_S(s)\delta(u, y, s)\\ &\quad\times \log \frac{p_{Y|X,S}(y|\psi(u, s), s) + q\omega(u, y, s)}{\sum_{x\in\mathcal{X}} p_{\hat{X}}(x)p_{Y|X,S}(y|x, s)},\\ &\qquad q \in [0, \frac{1}{|\mathcal{S}|}], u \in \mathcal{U},\\ &\frac{\partial^2}{\partial q^2} D_{GS}(p_{\hat{U}}, q, u)\\ &= \sum_{y\in\mathcal{Y},s\in\mathcal{S}} \frac{p_S(s)\delta^2(u, y, s)}{p_{Y|X,S}(y|\psi(u, s), s) + q\omega(u, y, s)} > 0,\\ &\qquad q \in [0, \frac{1}{|\mathcal{S}|}], u \in \mathcal{U}\backslash\mathcal{G}_\omega,\end{aligned}$$

where

$$\mathcal{G}_\omega = \{u \in \mathcal{U} : \omega(u, y, s) = 0 \text{ for all } y \in \mathcal{Y} \text{ and } s \in \mathcal{S}\}.$$

Clearly,

- $D_{GS}(p_{\hat{U}}, \frac{1}{|\mathcal{S}|}, u) \leq \underline{C}(p_{Y|X,S}, p_S)$ for $u \in \mathcal{U}$,
- $D_{GS}(p_{\hat{U}}, q, u)$ does not depend on q for $u \in \mathcal{G}_\omega$,
- $D_{GS}(p_{\hat{U}}, q, u)$ is a strictly convex function of q for $u \in \mathcal{U}\backslash\mathcal{G}_\omega$.

Hence, for each $u \in \mathcal{U}$, there are also three mutually exclusive cases.

1) $D_{GS}(p_{\hat{U}}, 0, u) \leq \underline{C}(p_{Y|X,S}, p_S)$: We have $D_{GS}(p_{\hat{U}}, q, u) \leq \underline{C}(p_{Y|X,S}, p_S)$ for $q \in [q(u), 1]$, where $q(u) = 0$.
2) $D_{GS}(p_{\hat{U}}, 0, u) > D_{GS}(p_{\hat{U}}, \frac{1}{|\mathcal{S}|}, u) = \underline{C}(p_{Y|X,S}, p_S)$ and $\frac{\partial}{\partial q} D_{GS}(p_{\hat{U}}, q, u)\Big|_{q=\frac{1}{|\mathcal{S}|}} \leq 0$ (this case can arise only when $|\mathcal{X}| > 2$): We have $D_{GS}(p_{\hat{U}}, 0, u) > \underline{C}(p_{Y|X,S}, p_S)$ for $q \in [0, q(u))$, where $q(u) = \frac{1}{|\mathcal{S}|}$.
3) Otherwise: We have $D_{GS}(p_{\hat{U}}, q, u) > \underline{C}(p_{Y|X,S}, p_S)$ for $q \in [0, q(u))$ and $D_{GS}(p_{\hat{U}}, q, u) \leq \underline{C}(p_{Y|X,S}, p_S)$ for $q \in [q(u), \frac{1}{|\mathcal{S}|}]$, where $q(u)$ is the unique solution of $D_{GS}(p_{\hat{U}}, q, u) = \underline{C}(p_{Y|X,S}, p_S)$ for $q \in (0, \frac{1}{|\mathcal{S}|})$.

It can be readily shown that

$$\underline{q}(p_{Y|X,S}, p_S) = \max_{u\in\mathcal{U}} q(u). \tag{65}$$

For $p_{Y|X,S}$ and p_S illustrated in Fig. 2 (see also (9) and (10)), we show in Appendix B that

$$\underline{\epsilon}(p_{Y|X,S}, p_S) = \begin{cases} \hat{\epsilon}(\theta), & \theta \in (0, 1),\\ 0, & \text{otherwise},\end{cases} \tag{66}$$

$$\underline{q}(p_{Y|X,S}, p_S) = \begin{cases} \hat{q}(\theta), & \theta \in (0, 1),\\ 0, & \text{otherwise},\end{cases} \tag{67}$$

where $\hat{\epsilon}(\theta)$ is the unique solution of

$$\begin{aligned}&\epsilon(1 - \theta) \log 2\epsilon + (1 - \epsilon(1 - \theta)) \log \frac{2(1 - \epsilon(1 - \theta))}{1 + \theta}\\ &\quad= (1 - \theta) \log 2 + \theta \log \frac{2\theta}{1 + \theta}\end{aligned}$$

for $\epsilon \in (0, 1)$, and $\hat{q}(\theta)$ is the unique solution of

$$\begin{aligned}&q(1 - \theta) \log 2q + (1 - q(1 - \theta)) \log \frac{2(1 - q(1 - \theta))}{1 + \theta}\\ &\quad= \frac{1}{2}\left((1 - \theta) \log 2 + \log \frac{2}{1 + \theta} + \theta \log \frac{2\theta}{1 + \theta}\right)\end{aligned}$$

for $q \in (0, \frac{1}{2})$. Setting $\theta = \frac{1}{2}$ gives $\underline{\epsilon}(p_{Y|X,S}, p_S) \approx 0.1$ (cf. Fig. 3) and $\underline{q}(p_{Y|X,S}, p_S) \approx 0.037$ (cf. Fig. 4).

B. Extension of Theorem 2

We shall extend Theorem 2 in a similar fashion. For any $p_{Y|X,S}$ and p_S, define

$$\begin{aligned}\overline{\epsilon}(p_{Y|X,S}, p_S) = \max\{&\epsilon \in [0, 1] : C'(p_{Y|X,S}, p_S, p_{\tilde{S}^{(\epsilon)}_{GE}|S})\\ &= \overline{C}(p_{Y|X,S}, p_S)\},\\ \overline{q}(p_{Y|X,S}, p_S) = \max\{&q \in [0, \frac{1}{|\mathcal{S}|}] : C'(p_{Y|X,S}, p_S, p_{\tilde{S}^{(q)}_{GS}|S})\\ &= \overline{C}(p_{Y|X,S}, p_S)\}.\end{aligned}$$

Proposition 2: 1) There exists $\beta(p_{Y|X,S}, p_S) > 0$ such that $C'(p_{Y|X,S}, p_S, p_{\tilde{S}|S}) = \overline{C}(p_{Y|X,S}, p_S)$ for all $p_{\tilde{S}|S}$ satisfying $H(S|\tilde{S}) \leq \beta(p_{Y|X,S}, p_S)$ if and only if $\overline{q}(p_{Y|X,S}, p_S) > 0$.
2) $\overline{q}(p_{Y|X,S}, p_S) > 0$ if and only if there exists $p_{\hat{X}|S} \in \mathcal{P}$ such that

$$\{x \in \mathcal{X} : p_{\hat{X}|S}(x|s) > 0\} = \mathcal{X}_+, \quad s \in \mathcal{S}. \tag{68}$$

Proof: The first statement can be easily extracted from the proof of Theorem 2. The second statement is a consequence of Lemma 4. ■

As shown by the following example, the necessary and sufficient condition (68) is not always satisfied when $|\mathcal{X}| > 2$.

[5] Note that the underlying $\mathcal{U}$ depends on $\tilde{\mathcal{S}}$. In particular, $|\mathcal{U}| = |\mathcal{X}|^{|\mathcal{S}|}$ when $p_{\tilde{S}|S}$ is a generalized symmetric channel whereas $|\mathcal{U}| = |\mathcal{X}|^{|\mathcal{S}|+1}$ when $p_{\tilde{S}|S}$ is a generalized erasure channel.

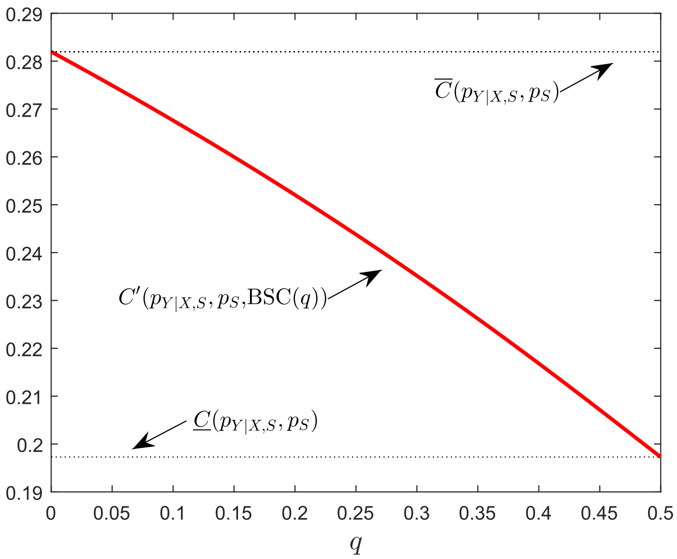

Fig. 6. Plot of $C'(p_{Y|X,S}, p_S, \mathrm{BSC}(q))$ against q for $q \in [0, \frac{1}{2}]$, where $p_{Y|X,S}$ and p_S are given by (69) and (70), respectively.

Let

$$p_{Y|X,S}(y|x,s) = \begin{cases} 1, & (x,y,s) = (0,0,0) \text{ or } (2,1,1), \\ 0, & (x,y,s) = (0,1,0) \text{ or } (2,0,1), \\ \frac{2}{5}, & (x,y,s) = (1,0,0) \text{ or } (0,1,1), \\ \frac{3}{5}, & (x,y,s) = (1,1,0) \text{ or } (0,0,1), \\ \frac{4}{5}, & (x,y,s) = (2,0,0) \text{ or } (1,1,1), \\ \frac{1}{5}, & (x,y,s) = (2,1,0) \text{ or } (1,0,1), \end{cases} \tag{69}$$

$$p_S(0) = p_S(1) = \frac{1}{2}. \tag{70}$$

For this example, it can be verified that the maximizer of the optimization problem in (5), denoted by $p_{\hat{X}|S}$, is unique and

$$\{x \in \mathcal{X} : p_{\hat{X}|S}(x|0) > 0\} = \{0, 1\},$$
$$\{x \in \mathcal{X} : p_{\hat{X}|S}(x|1) > 0\} = \{0, 2\};$$

indeed, Fig. 6 shows that $C'(p_{Y|X,S}, p_S, \mathrm{BSC}(q)) < \overline{C}(p_{Y|X,S}, p_S)$ for $q \in (0, \frac{1}{2}]$.

In view of Lemmas 3 and 4, we have

$$\overline{\epsilon}(p_{Y|X,S}, p_S) = \max_{p_{\hat{X}|S} \in \mathcal{P}} \sum_{x \in \mathcal{X}} \min_{s \in \mathcal{S}} p_{\hat{X}|S}(x|s), \tag{71}$$

$$\overline{q}(p_{Y|X,S}, p_S) = \max_{p_{\hat{X}|S} \in \mathcal{P}} \min_{x \in \mathcal{X}_+, s \in \mathcal{S}} \frac{p_{\hat{X}|S}(x|s)}{\sum_{s' \in \mathcal{S}} p_{\hat{X}|S}(x|s')}. \tag{72}$$

Note that $\mathcal{P}$ does not depend on p_S (under the assumption $\rho > 0$); as a consequence, $\overline{\epsilon}(p_{Y|X,S}, p_S)$ and $\overline{q}(p_{Y|X,S}, p_S)$ do not depend on p_S either. For $p_{Y|X,S}$ and p_S illustrated in Fig. 2 (see also (9) and (10)), we show in Appendix C that

$$\overline{\epsilon}(p_{Y|X,S}, p_S) = \begin{cases} 2\left(1 + (1-\theta)\theta^{\frac{\theta}{1-\theta}}\right)^{-1} \theta^{\frac{\theta}{1-\theta}}, & \theta \in (0,1), \\ 1, & \text{otherwise}, \end{cases} \tag{73}$$

$$\overline{q}(p_{Y|X,S}, p_S) = \begin{cases} \left(1 + (1-\theta)\theta^{\frac{\theta}{1-\theta}}\right)^{-1} \theta^{\frac{\theta}{1-\theta}}, & \theta \in (0,1), \\ \frac{1}{2}, & \text{otherwise}. \end{cases} \tag{74}$$

Setting $\theta = \frac{1}{2}$ gives $\overline{\epsilon}(p_{Y|X,S}, p_S) = \frac{4}{5}$ (cf. Fig. 3) and $\overline{q}(p_{Y|X,S}, p_S) = \frac{2}{5}$ (cf. Fig. 4).

C. Two Implicit Conditions

In this subsection, we shall examine the following two implicit conditions in Theorem 1:

1) perfect state information at the decoder,
2) causal noisy state observation at the encoder.

If no state information is available at the decoder, then the channel capacity is given by

$$\tilde{C}(p_{Y|X,S}, p_S, p_{\tilde{S}|S}) \triangleq \max_{p_U} I(U; Y),$$

where the joint distribution of $(U, X, Y, S, \tilde{S})$ is given by (2). Furthermore, if there is also no state information available at the encoder, then the channel capacity becomes

$$\underline{\tilde{C}}(p_{Y|X,S}, p_S) \triangleq \max_{p_X} I(X; Y), \tag{75}$$

where $p_{X,Y,S}(x, y, s) = p_X(x)p_S(s)p_{Y|X,S}(y|x, s)$. Define

$$\underline{\tilde{\epsilon}}(p_{Y|X,S}, p_S) = \min\{\epsilon \in [0, 1] : \tilde{C}(p_{Y|X,S}, p_S, p_{\tilde{S}^{(\epsilon)}_{GE}|S}) = \underline{\tilde{C}}(p_{Y|X,S}, p_S)\}.$$

The proof of the following result is similar to that of Proposition 1 and is omitted.

Proposition 3: 1) There exists $\tilde{\alpha}(p_{Y|X,S}, p_S) > 0$ such that $\tilde{C}(p_{Y|X,S}, p_S, p_{\tilde{S}|S}) = \underline{\tilde{C}}(p_{Y|X,S}, p_S)$ for all $p_{\tilde{S}|S}$ satisfying $I(S; \tilde{S}) \leq \tilde{\alpha}(p_{Y|X,S}, p_S)$ if and only if $\underline{\tilde{\epsilon}}(p_{Y|X,S}, p_S) < 1$.

2) $\underline{\tilde{\epsilon}}(p_{Y|X,S}, p_S) < 1$ if and only if

$$\sum_{y \in \mathcal{Y}} \left(\sum_{s \in \mathcal{S}} p_S(s)\delta(u, y, s) \right) \times \log \frac{\sum_{s \in \mathcal{S}} p_S(s) p_{Y|X,S}(y|\psi(u, *), s)}{\sum_{x \in \mathcal{X}, s \in \mathcal{S}} p_{\hat{X}}(x) p_S(s) p_{Y|X,S}(y|x, s)} > 0, \quad u \in \tilde{\mathcal{U}}_+ \backslash \tilde{\mathcal{G}}_\delta, \tag{76}$$

where $\delta(u, y, s)$ is defined in (14), $p_{\hat{X}}$ is an arbitrary maximizer of the optimization problem in (75), and

$$\tilde{\mathcal{G}}_\delta = \left\{ u \in \mathcal{U} : \sum_{s \in \mathcal{S}} p_S(s)\delta(u, y, s) = 0 \text{ for all } y \in \mathcal{Y} \right\},$$

$$\tilde{\mathcal{U}}_+ = \left\{ u \in \mathcal{U} : \sum_{y \in \mathcal{Y}} \left(\sum_{s \in \mathcal{S}} p_S(s) p_{Y|X,S}(y|\psi(u, *), s) \right) \times \log \frac{\sum_{s \in \mathcal{S}} p_S(s) p_{Y|X,S}(y|\psi(u, *), s)}{\sum_{x \in \mathcal{X}, s \in \mathcal{S}} p_{\hat{X}}(x) p_S(s) p_{Y|X,S}(y|x, s)} = \underline{\tilde{C}}(p_{Y|X,S}, p_S) \right\}.$$

As shown by the following example, the necessary and sufficient condition (76) is not always satisfied even when $|\mathcal{X}| = 2$. Let

$$Y = X \oplus S, \quad \mathcal{X} = \mathcal{Y} = \mathcal{S} = \{0, 1\}, \tag{77}$$

$$p_S(1) = \mu \in (0, \frac{1}{2}), \tag{78}$$

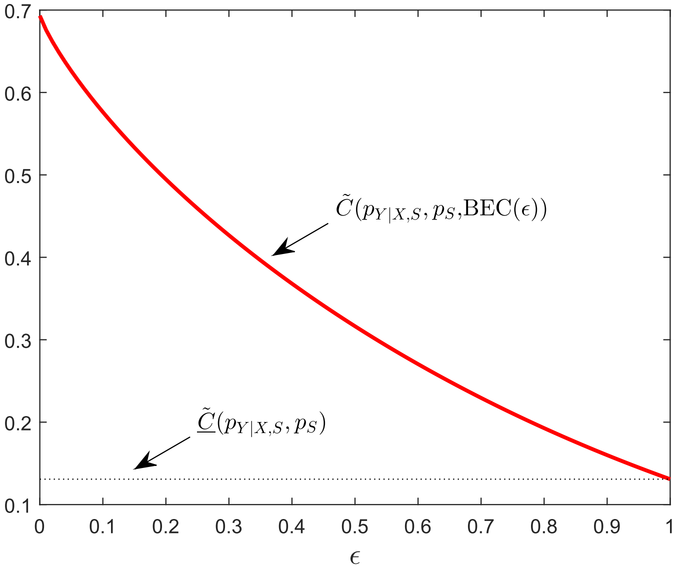

Fig. 7. Plot of $\tilde{C}(p_{Y|X,S}, p_S, \mathrm{BEC}(\epsilon))$ against ϵ for $\epsilon \in [0, 1]$, where $p_{Y|X,S}$ and p_S are given by (77) with $\mu = \frac{1}{4}$ and (78), respectively.

where $\oplus$ is the modulo-2 addition. It can be verified that (76) is not satisfied for this example; indeed, Fig. 7 indicates that

$$\tilde{C}(p_{Y|X,S}, p_S, \mathrm{BEC}(\epsilon)) > \underline{\tilde{C}}(p_{Y|X,S}, p_S), \quad \epsilon \in [0, 1). \quad (79)$$

Here we give an alternative way to prove (79). Write $S = \tilde{S} \oplus \Delta$, where $\tilde{S}$ and Δ are two mutually independent Bernoulli random variables with

$$p_{\tilde{S}}(1) = \nu \in [0, \mu],$$
$$p_{\Delta}(1) = \frac{\mu - \nu}{1 - 2\nu}.$$

It is clear that

$$\begin{aligned}\tilde{C}(p_{Y|X,S}, p_S, p_{\tilde{S}|S}) &= \log 2 - H(\Delta)\\ &> \log 2 - H(S)\\ &= \underline{\tilde{C}}(p_{Y|X,S}, p_S), \quad \nu \in (0, \mu]. \quad (80)\end{aligned}$$

In light of Lemma 3, $p_{\tilde{S}|S}$ is a stochastically degraded version of $\mathrm{BEC}(\epsilon)$ and consequently

$$\tilde{C}(p_{Y|X,S}, p_S, \mathrm{BEC}(\epsilon)) \geq \tilde{C}(p_{Y|X,S}, p_S, p_{\tilde{S}|S}) \quad (81)$$

if $H(S) - H(\Delta) \leq \frac{\mu^2(1-\epsilon)^2}{2}$. Combining (80) and (81) proves (79).

Now we proceed to examine the second implicit condition. If the noisy state observation is available non-causally at the encoder, the Gelfand-Pinsker theorem [11] (see also [3, Th. 7.3]) indicates that the channel capacity is given by

$$C_{GP}(p_{Y|X,S}, p_S, p_{\tilde{S}|S}) \triangleq \max_{p_{U|\tilde{S}}} I(U; Y, S) - I(U; \tilde{S}),$$

where the joint distribution of $(U, X, Y, S, \tilde{S})$ factors as

$$\begin{aligned}&p_{U,X,Y,S,\tilde{S}}(u, x, y, s, \tilde{s})\\ &\quad = p_S(s)p_{\tilde{S}|S}(\tilde{s}|s)p_{U|\tilde{S}}(u|\tilde{s})\mathbb{I}(x = \psi(u, \tilde{s}))p_{Y|X,S}(y|x, s),\\ &\qquad\qquad u \in \mathcal{U}, x \in \mathcal{X}, y \in \mathcal{Y}, s \in \mathcal{S}, \tilde{s} \in \tilde{\mathcal{S}}.\end{aligned}$$

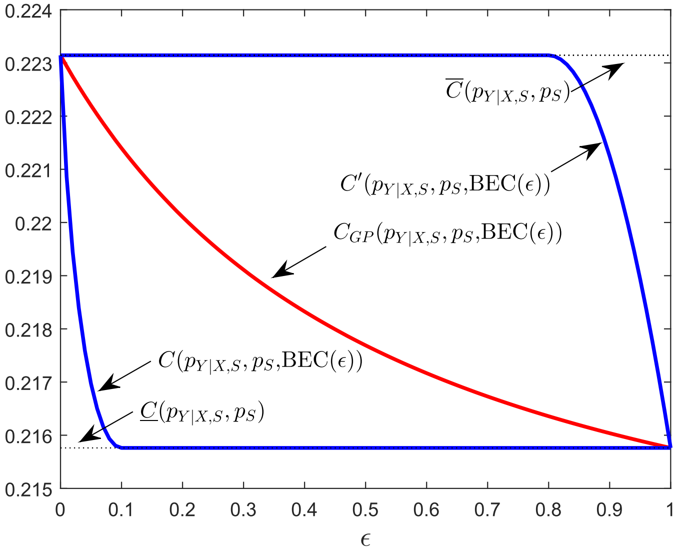

Fig. 8. Plot of $C_{GP}(p_{Y|X,S}, p_S, \mathrm{BEC}(\epsilon))$ against ϵ for $\epsilon \in [0, 1]$, where $p_{Y|X,S}$ and p_S are given by (9) with $\theta = \frac{1}{2}$ and (10), respectively.

It turns out that $C_{GP}(p_{Y|X,S}, p_S, p_{\tilde{S}|S})$ is bounded between $C(p_{Y|X,S}, p_S, p_{\tilde{S}|S})$ and $C'(p_{Y|X,S}, p_S, p_{\tilde{S}|S})$, i.e.,

$$\begin{aligned}C(p_{Y|X,S}, p_S, p_{\tilde{S}|S}) &\leq C_{GP}(p_{Y|X,S}, p_S, p_{\tilde{S}|S})\\ &\leq C'(p_{Y|X,S}, p_S, p_{\tilde{S}|S}).\end{aligned}$$

Indeed, the first inequality is obvious, and the second one holds because

$$\begin{aligned}I(U; Y, S) - I(U; \tilde{S}) &\leq I(U; Y, S) - I(U; S)\\ &= I(U; Y|S)\\ &\leq I(X; Y|S).\end{aligned}$$

In Fig. 8 we plot $C_{GP}(p_{Y|X,S}, p_S, \mathrm{BEC}(\epsilon))$ against ϵ for $\epsilon \in [0, 1]$, where $p_{Y|X,S}$ and p_S are given by (9) with $\theta = \frac{1}{2}$ and (10), respectively; it can be seen that $C_{GP}(p_{Y|X,S}, p_S, \mathrm{BEC}(\epsilon))$ is strictly greater than $\underline{C}(p_{Y|X,S}, p_S)$ except when $\epsilon = 1$. So the causality condition on the noisy state observation at the encoder is not superfluous for Theorem 1.

V. Conclusion

We have shown that the capacity of binary-input[6] channels is very "sensitive" to the quality of the encoder side information whereas the generalized probing capacity is very "robust." Here the words "sensitive" and "robust" should not be understood in a quantitative sense. Indeed, it is known [7] that, when $|\mathcal{X}| = 2$, the ratio of $\underline{C}(p_{Y|X,S}, p_S)$ to $\overline{C}(p_{Y|X,S}, p_S)$ is at least 0.942 and the difference between these two quantities is at most ~0.011 bit; in other words, the gain that can be obtained by exploiting the encoder side information (or the loss that can be incurred by ignoring the encoder side information) is very limited anyway.

[6] In fact, both numerical simulation and theoretical analysis suggest that similar results hold for many (but not all) non-binary-input channels.

Binary signalling is widely used, especially in wideband communications. So our work might have some practical relevance. However, great caution should be exercised in interpreting Theorems 1 and 2. Specifically, both results rely on the assumption that the channel state takes values from a finite set,[7] which is not necessarily satisfied in reality; moreover, the freedom of power control in real communication systems is not captured by our results. Nevertheless, our work can be viewed as an initial step towards a better understanding of the fundamental performance limits of communication systems where the transmitter side information and the receiver side information are not deterministically related.

Finally, it is worth mentioning that our results might have their counterparts in source coding.

APPENDIX A
AN ALTERNATIVE PROOF OF THEOREM 2

We shall show that, for any binary-input channel $p_{Y|X,S}$, state distribution p_S, and side channel $p_{\tilde{S}|S}$,

$$C'(p_{Y|X,S}, p_S, p_{\tilde{S}|S}) = \overline{C}(p_{Y|X,S}, p_S)$$

if

$$H(S|\tilde{S}) \le \frac{4\rho \log 2}{3 + 2(e-1)\sqrt{2|\mathcal{S}|}}. \tag{82}$$

Lemma 7: $p_{\hat{X}|S}$ is a stochastically degraded version of $p_{\tilde{S}|S}$ if

$$H(S|\tilde{S}) \le \frac{4\tau\rho \log 2}{3\tau + 2\sqrt{2|\mathcal{S}|}}, \tag{83}$$

where

$$\tau = \min_{x \in \mathcal{X}_+} \frac{\min_{s\in\mathcal{S}} p_{\hat{X}|S}(x|s)}{\max_{s\in\mathcal{S}} p_{\hat{X}|S}(x|s)}.$$

Proof: Let $\hat{S}$ denote the maximum likelihood estimate of S based on $\tilde{S}$. It suffices to show that $p_{\hat{S}|S}$ is invertible and $p_{\hat{S}|S}^{-1} p_{\hat{X}|S}$ is a valid probability transition matrix if (83) is satisfied.

Let $\sigma_{\min}(p_{\hat{S}|S})$ denote the smallest singular value of $p_{\hat{S}|S}$. It follows from [12, Th. 3] that

$$\sigma_{\min}(p_{\hat{S}|S}) \ge \min_{s\in\mathcal{S}} \frac{1}{2}\Bigg(2p_{\hat{S}|S}(s|s) - \sum_{\hat{s}\in\mathcal{S}:\hat{s}\ne s} p_{\hat{S}|S}(\hat{s}|s) - \sum_{\hat{s}\in\mathcal{S}:\hat{s}\ne s} p_{\hat{S}|S}(s|\hat{s})\Bigg). \tag{84}$$

Clearly,

$$\begin{aligned}
&\min_{s\in\mathcal{S}} \frac{1}{2}\Bigg(2p_{\hat{S}|S}(s|s) - \sum_{\hat{s}\in\mathcal{S}:\hat{s}\ne s} p_{\hat{S}|S}(\hat{s}|s) - \sum_{\hat{s}\in\mathcal{S}:\hat{s}\ne s} p_{\hat{S}|S}(s|\hat{s})\Bigg)\\
&= \min_{s\in\mathcal{S}} \frac{1}{2}\Bigg(2 - 3\sum_{\hat{s}\in\mathcal{S}:\hat{s}\ne s} p_{\hat{S}|S}(\hat{s}|s) - \sum_{\hat{s}\in\mathcal{S}:\hat{s}\ne s} p_{\hat{S}|S}(s|\hat{s})\Bigg)\\
&\ge 1 - \frac{3}{2}\sum_{s,\hat{s}\in\mathcal{S}:s\ne\hat{s}} p_{\hat{S}|S}(\hat{s}|s). \qquad (85)
\end{aligned}$$

Substituting (85) into (84) and invoking (45) gives

$$\sigma_{\min}(p_{\hat{S}|S}) \ge 1 - \frac{3H(S|\tilde{S})}{4\rho \log 2}. \tag{86}$$

Therefore, $p_{\hat{S}|S}$ is invertible if $H(S|\tilde{S}) < \frac{4\rho\log 2}{3}$. Let $\|\cdot\|_\infty$, $\|\cdot\|_2$, and $\|\cdot\|_F$ denote the maximum row sum matrix norm, the spectral norm, and the Frobenius norm, respectively [13]. Note that

$$\begin{aligned}
&\|p_{\hat{S}|S}^{-1} - \mathrm{diag}(1,\cdots,1)\|_\infty\\
&\quad\le \sqrt{|\mathcal{S}|}\,\|p_{\hat{S}|S}^{-1} - \mathrm{diag}(1,\cdots,1)\|_2\\
&\quad\le \sqrt{|\mathcal{S}|}\,\|p_{\hat{S}|S}^{-1}\|_2 \|p_{\hat{S}|S} - \mathrm{diag}(1,\cdots,1)\|_2 \qquad (87)\\
&\quad\le \sqrt{|\mathcal{S}|}\,\|p_{\hat{S}|S}^{-1}\|_2 \|p_{\hat{S}|S} - \mathrm{diag}(1,\cdots,1)\|_F, \qquad (88)
\end{aligned}$$

where (87) follows by the sub-multiplicative property of the spectral norm. We have

$$\begin{aligned}
\|p_{\hat{S}|S}^{-1}\|_2 &= \frac{1}{\sigma_{\min}(p_{\hat{S}|S})}\\
&\le \left(1 - \frac{3H(S|\tilde{S})}{4\rho\log 2}\right)^{-1}, \qquad (89)
\end{aligned}$$

where (89) is due to (86). For $p_{\hat{S}|S} - \mathrm{diag}(1,\cdots,1)$, it is clear that the diagonal entries are non-positive, the off-diagonal entries are non-negative, and the sum of all entries is equal to 0; moreover, the sum of its off-diagonal entries is bounded above by $\frac{H(S|\tilde{S})}{2\rho\log 2}$ (see (45)). Therefore,

$$\begin{aligned}
&\|p_{\hat{S}|S} - \mathrm{diag}(1,\cdots,1)\|_F\\
&= \sqrt{\sum_{s\in\mathcal{S}} (p_{\hat{S}|S}(s|s) - 1)^2 + \sum_{s,\hat{s}\in\mathcal{S}:s\ne\hat{s}} (p_{\hat{S}|S}(\hat{s}|s))^2}\\
&\le \sqrt{\Bigg(\sum_{s\in\mathcal{S}} (p_{\hat{S}|S}(s|s) - 1)\Bigg)^2 + \Bigg(\sum_{s,\hat{s}\in\mathcal{S}:s\ne\hat{s}} p_{\hat{S}|S}(\hat{s}|s)\Bigg)^2}\\
&= \sqrt{2\Bigg(\sum_{s,\hat{s}\in\mathcal{S}:s\ne\hat{s}} p_{\hat{S}|S}(\hat{s}|s)\Bigg)^2}\\
&\le \frac{H(S|\tilde{S})}{\sqrt{2}\rho\log 2}. \qquad (90)
\end{aligned}$$

Substituting (89) and (90) into (88) yields

$$\begin{aligned}
&\|p_{\hat{S}|S}^{-1} - \mathrm{diag}(1,\cdots,1)\|_\infty\\
&\quad\le \frac{\sqrt{|\mathcal{S}|}H(S|\tilde{S})}{\sqrt{2}\rho\log 2}\left(1 - \frac{3H(S|\tilde{S})}{4\rho\log 2}\right)^{-1}. \qquad (91)
\end{aligned}$$

To ensure that all entries of $p_{\hat{S}|S}^{-1} p_{\hat{X}|S}$ are non-negative (or equivalently $(\mathrm{diag}(1,\cdots,1) - p_{\hat{S}|S}^{-1})p_{\hat{X}|S}$ is component-wise dominated by $p_{\hat{X}|S}$), it suffices to have

$$\|p_{\hat{S}|S}^{-1} - \mathrm{diag}(1,\cdots,1)\|_\infty \le \tau. \tag{92}$$

[7] In contrast, the assumption $|\mathcal{Y}| < \infty$ and $|\tilde{\mathcal{S}}| < \infty$ is not essential

TABLE I

SPECIFICATION OF $\psi(\cdot,\cdot)$ FOR $\mathcal{U}=\{0,1,\cdots,7\}$ AND $\tilde{\mathcal{S}}=\{0,1,*\}$

$\psi(u,\tilde{s})$	$\tilde{s}=0$	$\tilde{s}=1$	$\tilde{s}=*$
$u=0$	0	0	0
$u=1$	1	1	1
$u=2$	1	1	0
$u=3$	0	0	1
$u=4$	0	1	0
$u=5$	0	1	1
$u=6$	1	0	0
$u=7$	1	0	1

Combining (91) and (92) shows that $p^{-1}_{\hat{S}|S}p_{\hat{X}|S}$ is a valid probability transition matrix[8] if (83) is satisfied.[9] ■

Since $|\mathcal{X}|=2$, it follows from [7, Th. 2] that there exists $p_{\hat{X}|S}\in\mathcal{P}$ satisfying

$$p_{\hat{X}|S}(x|s)>e^{-1},\quad x\in\mathcal{X},s\in\mathcal{S}.$$

For such $p_{\hat{X}|S}$, we have

$$\tau\geq\frac{1}{e-1}.$$

Invoking Lemma 7 shows that $p_{\hat{X}|S}$ is a stochastically degraded version of $p_{\tilde{S}|S}$ (and consequently $C'(p_{Y|X,S},p_S,p_{\tilde{S}|S})=\overline{C}(p_{Y|X,S},p_S)$) if (82) is satisfied.

APPENDIX B
PROOF OF (66) AND (67)

Lemma 8: For $\theta\in(0,1)$,

$$\eta(\theta)\triangleq(1-\theta)\log(1+\theta)+\theta\log\theta<0.$$

Proof: We have

$$\begin{aligned}\frac{\mathrm{d}^2\eta(\theta)}{\mathrm{d}\theta^2}&=\frac{\mathrm{d}}{\mathrm{d}\theta}\left(-\log(1+\theta)+\frac{1-\theta}{1+\theta}+\log\theta+1\right)\\&=-\frac{1}{1+\theta}-\frac{2}{(1+\theta)^2}+\frac{1}{\theta}\\&=\frac{1-\theta}{\theta(1+\theta)^2}\\&>0,\quad\theta\in(0,1),\end{aligned}$$

which, together with the fact $\eta(0)=\eta(1)=0$, implies the desired result. ■

When $\theta=0$ or $\theta=1$, we have $\underline{C}(p_{Y|X,S},p_S)=\overline{C}(p_{Y|X,S},p_S)$, which implies $\underline{\epsilon}(p_{Y|X,S},p_S)=\underline{q}(p_{Y|X,S},p_S)=0$. When $\theta\in(0,1)$, the maximizer of the optimization problem in (4), denoted by $p_{\hat{X}}$, is unique and is given by

$$p_{\hat{X}}(0)=p_{\hat{X}}(1)=\frac{1}{2}.$$

Now consider $\psi(\cdot,\cdot)$ specified by Table I. It can be verified that

$$\begin{aligned}&D_{GE}(p_{\hat{U}},\epsilon,u)\\&=\frac{1}{2}\left((1-\theta)\log 2+\log\frac{2}{1+\theta}+\theta\log\frac{2\theta}{1+\theta}\right),\quad u=0,1,\\&D_{GE}(p_{\hat{U}},\epsilon,u)\\&=\frac{1}{2}\Bigg(\epsilon(1-\theta)\log 2\epsilon+(\theta+\epsilon(1-\theta))\log\frac{2(\theta+\epsilon(1-\theta))}{1+\theta}\\&\qquad+(1-\epsilon(1-\theta))\log\frac{2(1-\epsilon(1-\theta))}{1+\theta}\\&\qquad+(1-\epsilon)(1-\theta)\log 2(1-\epsilon)\Bigg),\quad u=2,3,\\&D_{GE}(p_{\hat{U}},\epsilon,u)\\&=\frac{1}{2}\Bigg((1-\theta)\log 2+(\theta+\epsilon(1-\theta))\log\frac{2(\theta+\epsilon(1-\theta))}{1+\theta}\\&\qquad+\theta\log\frac{2\theta}{1+\theta}+(1-\epsilon)(1-\theta)\log 2(1-\epsilon)\Bigg),\quad u=4,5,\\&D_{GE}(p_{\hat{U}},\epsilon,u)\\&=\frac{1}{2}\Bigg(\epsilon(1-\theta)\log 2\epsilon+\log\frac{2}{1+\theta}\\&\qquad+(1-\epsilon(1-\theta))\log\frac{2(1-\epsilon(1-\theta))}{1+\theta}\Bigg),\quad u=6,7.\end{aligned}$$

Moreover,

$$\begin{aligned}D_{GE}(p_{\hat{U}},0,u)&=\frac{1}{2}\Bigg((1-\theta)\log 2+\log\frac{2}{1+\theta}\\&\qquad+\theta\log\frac{2\theta}{1+\theta}\Bigg)\\&=\underline{C}(p_{Y|X,S},p_S),\quad u=0,1,2,3,\\D_{GE}(p_{\hat{U}},0,u)&=(1-\theta)\log 2+\theta\log\frac{2\theta}{1+\theta}\\&<\underline{C}(p_{Y|X,S},p_S),\quad u=4,5,\qquad(93)\\D_{GE}(p_{\hat{U}},0,u)&=\log\frac{2}{1+\theta}\\&>\underline{C}(p_{Y|X,S},p_S),\quad u=6,7,\qquad(94)\end{aligned}$$

where (93) and (94) follow from Lemma 8. Therefore, we have

$$\begin{aligned}\epsilon(u)&=0,\quad u=0,1,2,3,4,5,\\\epsilon(u)&=\hat{\epsilon}(\theta),\quad u=6,7,\end{aligned}$$

which, together with (64), proves (66) for $\theta\in(0,1)$. Next consider $\psi(\cdot,\cdot)$ specified by Table II. It can be verified that

$$\begin{aligned}&D_{GS}(p_{\hat{U}},q,u)\\&=\frac{1}{2}\left((1-\theta)\log 2+\log\frac{2}{1+\theta}+\theta\log\frac{2\theta}{1+\theta}\right),\quad u=0,1,\\&D_{GS}(p_{\hat{U}},q,2)\\&=(1-q)(1-\theta)\log 2(1-q)\\&\quad+(\theta+q(1-\theta))\log\frac{2(\theta+q(1-\theta))}{1+\theta},\\&D_{GS}(p_{\hat{U}},q,3)\\&=q(1-\theta)\log 2q+(1-q(1-\theta))\log\frac{2(1-q(1-\theta))}{1+\theta}.\end{aligned}$$

[8] The requirement that the entries in each row of $p^{-1}_{\hat{S}|S}p_{\hat{X}|S}$ add up to 1 is automatically satisfied.

[9] Note that (83) implies $H(S|\tilde{S})<\frac{4\rho\log 2}{3}$, which further implies the existence of $p^{-1}_{\hat{S}|S}$.

TABLE II

SPECIFICATION OF $\psi(\cdot,\cdot)$ FOR $\mathcal{U}=\{0,1,\cdots,3\}$ AND $\tilde{\mathcal{S}}=\{0,1,*\}$

$\psi(u,\tilde{s})$	$\tilde{s}=0$	$\tilde{s}=1$
$u=0$	0	0
$u=1$	1	1
$u=2$	0	1
$u=3$	1	0

Moreover,

$$\begin{aligned} D_{GS}(p_{\hat{U}},0,u) &= \frac{1}{2}\left((1-\theta)\log 2+\log\frac{2}{1+\theta}\right.\\ &\qquad\left.+\theta\log\frac{2\theta}{1+\theta}\right)\\ &= \underline{C}(p_{Y|X,S},p_S),\quad u=0,1,\\ D_{GS}(p_{\hat{U}},0,2) &= (1-\theta)\log 2+\theta\log\frac{2\theta}{1+\theta}\\ &< \underline{C}(p_{Y|X,S},p_S), \qquad (95)\\ D_{GS}(p_{\hat{U}},0,3) &= \log\frac{2}{1+\theta}\\ &> \underline{C}(p_{Y|X,S},p_S), \qquad (96) \end{aligned}$$

where (95) and (96) follow from Lemma 8. Therefore, we have

$$\begin{aligned} q(u) &= 0,\quad u=0,1,2,\\ q(3) &= \hat{q}(\theta), \end{aligned}$$

which, together with (65), proves (67) for $\theta\in(0,1)$.

APPENDIX C
PROOF OF (73) AND (74)

When $\theta=0$ or $\theta=1$, we have $\underline{C}(p_{Y|X,S},p_S)=\overline{C}(p_{Y|X,S},p_S)$, which implies $\overline{\epsilon}(p_{Y|X,S},p_S)=1$ and $\overline{q}(p_{Y|X,S},p_S)=\frac{1}{2}$. When $\theta\in(0,1)$, the maximizer of the optimization problem in (5), denoted by $p_{\hat{X}|S}$, is unique and is given by

$$\begin{aligned} &p_{\hat{X}|S}(x|s)\\ &=\begin{cases} \left(1+(1-\theta)\theta^{\frac{\theta}{1-\theta}}\right)^{-1}\theta^{\frac{\theta}{1-\theta}}, & x=s,\\ \left(1+(1-\theta)\theta^{\frac{\theta}{1-\theta}}\right)^{-1}\left(1-\theta^{\frac{1}{1-\theta}}\right), & \text{otherwise.}\end{cases} \end{aligned}$$

In view of (71) and (72), it suffices to show that

$$\theta^{\frac{\theta}{1-\theta}}<1-\theta^{\frac{1}{1-\theta}},\quad \theta\in(0,1).$$

Indeed, for $\theta\in(0,1)$,

$$\begin{aligned} &\theta^{\frac{\theta}{1-\theta}}<1-\theta^{\frac{1}{1-\theta}}\\ &\Leftrightarrow 1<\theta^{-\frac{\theta}{1-\theta}}-\theta\\ &\Leftrightarrow (1-\theta)\log(1+\theta)+\theta\log\theta<0, \end{aligned}$$

and the last inequality is true according to Lemma 8.

ACKNOWLEDGMENT

The authors wish to thank the associate editor and the anonymous reviewer for their valuable comments and suggestions.

REFERENCES

[1] C. E. Shannon, "Channels with side information at the transmitter," *J. Res. Develop.*, vol. 2, no. 4, pp. 289–293, Oct. 1958.

[2] G. Caire and S. Shamai (Shitz), "On the capacity of some channels with channel state information," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2007–2019, Sep. 1999.

[3] A. El-Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[4] H. Asnani, H. Permuter, and T. Weissman, "Probing capacity," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7317–7332, Nov. 2011.

[5] J. Wang, J. Chen, L. Zhao, P. Cuff, and H. Permuter, "On the role of the refinement layer in multiple description coding and scalable coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1443–1456, Mar. 2011.

[6] R. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[7] N. Shulman and M. Feder, "The uniform distribution as a universal prior," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1356–1362, Jun. 2004.

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

[9] R. J. Plemmons, "*M*-matrices characterizations.I—Nonsingular *M*-matrices," *Linear Algebra Appl.*, vol. 18, no. 2, pp. 175–188, 1977.

[10] S.-W. Ho and S. Verdú, "On the interplay between conditional entropy and error probability," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 5930–5942, Dec. 2010.

[11] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Control Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[12] C. R. Johnson, "A Gersgorin-type lower bound for the smallest singular value," *Linear Algebra Appl.*, vol. 112, no. 1, pp. 1–7, 1989.

[13] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

Rui Xu received the B.Sc. degree in electronic information engineering from Nanjing University of Science and Technology, Nanjing, China, in 2010, and the M.S. degree in electromagnetic field and microwave techniques from Southeast University, Nanjing, China, in 2013. He is currently pursuing the Ph.D. degree in electrical and computer engineering at McMaster University, Hamilton, ON, Canada. His research interests include information and coding theory.

Jun Chen (S'03–M'06–SM'16) received the B.E. degree with honors in communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2001 and the M.S. and Ph.D. degrees in electrical and computer engineering from Cornell University, Ithaca, NY, in 2004 and 2006, respectively.

He was a Postdoctoral Research Associate in the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign, Urbana, IL, from September 2005 to July 2006, and a Postdoctoral Fellow at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, from July 2006 to August 2007. Since September 2007 he has been with the Department of Electrical and Computer Engineering at McMaster University, Hamilton, ON, Canada, where he is currently an Associate Professor and a Joseph Ip Distinguished Engineering Fellow. His research interests include information theory, wireless communications, and signal processing.

He received several awards for his research, including the Josef Raviv Memorial Postdoctoral Fellowship in 2006, the Early Researcher Award from the Province of Ontario in 2010, and the IBM Faculty Award in 2010. He is currently serving as an Associate Editor for Shannon Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY.

Tsachy Weissman (S'99–M'02–SM'07–F'13) graduated *summa cum laude* with a B.Sc. in electrical engineering from the Technion in 1997, and earned his Ph.D. at the same place in 2001. He then worked at Hewlett Packard Laboratories with the information theory group until 2003, when he joined Stanford University, where he is currently Professor of Electrical Engineering and incumbent of the STMicroelectronics chair in the School of Engineering. He has spent leaves at the Technion, and at ETH Zurich. Tsachy's research is focused on information theory, compression, communication, statistical signal processing, the interplay between them, and their applications. He is recipient of several best paper awards, and prizes for excellence in research and teaching. He served on the editorial board of the IEEE TRANSACTIONS ON INFORMATION THEORY from Sept. 2010 to Aug. 2013, and currently serves on the editorial board of Foundations and Trends in Communications and Information Theory. He is Founding Director of the Stanford Compression Forum.

Jian-Kang Zhang (SM'16) received the B.S. degree in information science (math.) from Shaanxi Normal University, Xian, China, the M.S. degree in information and computational science (math.) from Northwest University, Xian, China, and the Ph.D. degree in electrical engineering from Xidian University, Xian, China, in 1983, 1988, and 1999, respectively. He is currently an Associate Professor in the Department of Electrical and Computer Engineering at McMaster University, Hamilton, ON, Canada. He has held research positions at McMaster University and Harvard University.

His research interests are in the general area of signal processing, mainly emphasizing mathematics-based new technology innovation and exploration for variety of signal processing and practical applications. His current research focuses on transceiver designs for multiuser communication systems, coherent and noncoherent space-time signal and receiver designs for MIMO and cooperative relay communications.

Dr. Zhang is the coauthor of the paper that received the IEEE Signal Processing Society Best Young Author Award in 2008. He has served as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS and the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He is currently serving as an Associate Editor for the *Journal of Electrical and Computer Engineering*.