ISIT'2013 1569716885

A General Formula for Capacity of Channels with Action-Dependent States

Tetsunao Matsuta*, Tomohiko Uyematsu†
Department of Communications and Computer Engineering
Tokyo Institute of Technology
Ookayama, Meguro-ku, Tokyo, 152-8550, Japan
Email:tetsu@it.ce.titech.ac.jp*, uyematsu@ieee.org†

***Abstract*—Weissman introduced a channel coding problem for channels with action-dependent states. In this coding problem, there are two encoders and a decoder. One encoder outputs an action that affects states of the channel. Then, the other encoder encodes a message by using the channel state, and its codeword is fed into the channel. The decoder receives a noisy observation of the codeword, and reconstructs the message. For this coding problem, Weissman showed the capacity when states and the channel are stationary memoryless. In this paper, we show a general formula of the capacity when states and the channel may not be stationary memoryless, which is expressed by mutual information spectrum-sup/inf proposed by Verdú and Han. Our general formula coincides with the capacity derived by Tan when actions cannot affect states of channels. We also show that the capacity for nonstationary memoryless channels can be expressed by using ordinary mutual information.**

I. Introduction

In many practical situations, states of a communication channel change from moment to moment. To analyze the performance of coding systems under such a situation, many researchers have been studied coding problems for channels with states (cf., e.g., [1, Chapter 7]). In these coding problems, researchers considered the situation where states of the channel change by nature. Thus, the coding system can neither control nor affect states of the channel.

On the other hand, Weissman [2] introduced a coding problem for *channels with action-dependent states*, in which an *action* of the coding system affects states of the channel. More precisely, he considered the following coding problem

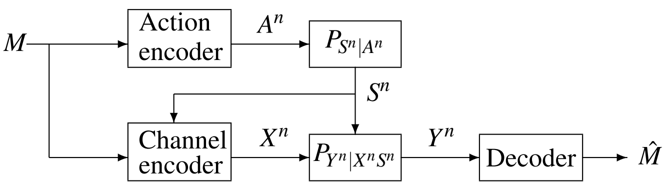

Fig. 1. Channels with action dependent states

(see Fig. 1): To send a message to the receiver via a channel with states, the coding system uses an action encoder and a channel encoder. The action encoder outputs an action corresponding to the message, and the action affects states of the channel. Then, the channel encoder receives the state of the channel, and outputs a codeword of the message into the channel. The receiver receives a noisy observation of the codeword, and reconstructs the message by using a decoder. For this coding problem, Weissman [2] showed the capacity when states and the channel are stationary memoryless, where the capacity is the supremum of rates of the code such that the decoding error probability vanishes as the block length tends to infinity. In [2], he also studied various applications of channels with action-dependent states, and showed the capacity for these applications. Since he introduced the coding problem for channels with action-dependent states, a lot of extensions of this coding problem have been reported [3], [4], [5], [6].

In the coding problem for channels with action-dependent states, when actions cannot affect states of channels, the coding problem coincides with the Gel'fand-Pinsker coding problem [7]. For this coding problem, Gel'fand and Pinsker [7] showed the capacity when states and the channel are stationary memoryless. Recently, Tan [8] showed the capacity for the Gel'fand-Pinsker coding problem when states and the channel may not be stationary memoryless. On the other hand, it has not yet been clarified the capacity of channels with action-dependent states when states and the channel may not be stationary memoryless. In this paper, we show a general formula for the capacity of channels with action-dependent states, which is expressed by mutual information spectrum-sup/inf proposed by Verdú and Han [9]. The proof is based on information-spectrum methods [10], and proofs in [8]. Our general formula coincides with the capacity derived by Tan [8] when actions cannot affect states of channels. We also show that the capacity for nonstationary memoryless channels can be expressed by using ordinary mutual information.

II. Preliminaries

In this section, we provide a precise formulation of the coding problem for channels with action-dependent states.

We will denote an n-length sequence of symbols $(a_1, a_2, \cdots, a_n)$ by a^n. We will denote random variables (RVs) by capital letters $X, Y, Z, \cdots$, the values they can take by lowercase letters $x, y, z, \cdots$, and the set of these values by calligraphic letters $\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \cdots$. The probability distribution of a random variable X will be denoted by P_X. The conditional distribution of X given Y will be denoted by $P_{X|Y}$. We will denote the set of all probability distributions over $\mathcal{X}$ by $\mathscr{P}(\mathcal{X})$, and the set of all conditional distributions from $\mathcal{Y}$ to $\mathcal{X}$ by $\mathscr{W}(\mathcal{X}|\mathcal{Y})$.

Let $\mathscr{X}$, $\mathscr{Y}$, $\mathscr{A}$ and $\mathscr{S}$ be finite sets. A channel with action-dependent states is characterized by sequences of channels $\mathbf{P_{S|A}} = \{P_{S^n|A^n}\}_{n=1}^{\infty}$ and $\mathbf{P_{Y|XS}} = \{P_{Y^n|X^nS^n}\}_{n=1}^{\infty}$, where $P_{S^n|A^n} \in \mathscr{W}(\mathscr{S}^n|\mathscr{A}^n)$ and $P_{Y^n|X^nS^n} \in \mathscr{W}(\mathscr{Y}^n|\mathscr{X}^n \times \mathscr{S}^n)$. Note that when channels are stationary memoryless, we can simply express channels $\mathbf{P_{S|A}}$ and $\mathbf{P_{Y|XS}}$ as conditional distributions $P_{S|A} \in \mathscr{W}(\mathscr{S}|\mathscr{A})$ and $P_{Y|XS} \in \mathscr{W}(\mathscr{Y}|\mathscr{X} \times \mathscr{S})$, respectively. For a message $m \in \mathscr{M}_n \triangleq \{1, \cdots, M_n\}$, the action encoder f_{A}^n outputs an action $a^n \in \mathscr{A}^n$ corresponding to the message m. Thus, the action encoder is defined as

$$f_{\mathrm{A}}^n : \mathscr{M}_n \to \mathscr{A}^n.$$

The action a^n affects states $s^n \in \mathscr{S}^n$ according to $P_{S^n|A^n}(s^n|a^n)$. On the other hand, the channel encoder f_{C}^n receives the state s^n, and outputs a codeword $x^n \in \mathscr{X}^n$ corresponding to the message m. Thus, the channel encoder is defined as

$$f_{\mathrm{C}}^n : \mathscr{M}_n \times \mathscr{S}^n \to \mathscr{X}^n.$$

The channel encoder sends the codeword to a decoder via the channel $P_{Y^n|X^nS^n}$. Then, the decoder φ_n receives a channel output $y^n \in \mathscr{Y}^n$ that is drawn from $P_{Y^n|X^nS^n}(y^n|x^n,s^n)$, and reconstructs the message m. Thus, the decoder is defined as

$$\varphi_n : \mathscr{Y}^n \to \mathscr{M}_n.$$

The rate R_n of a code $(f_{\mathrm{A}}^n, f_{\mathrm{C}}^n, \varphi_n)$ is defined as

$$R_n \triangleq \frac{1}{n}\log M_n,$$

and the error probability of the code is defined as

$$\varepsilon_n(f_{\mathrm{A}}^n, f_{\mathrm{C}}^n, \varphi_n) \triangleq \frac{1}{M_n}\sum_{m\in\mathscr{M}_n} \Pr\{\varphi_n(Y^n) \neq m | m \text{ is sent}\}.$$

We say $R \geq 0$ is achievable if there exists a sequence of codes $\{(f_{\mathrm{A}}^n, f_{\mathrm{C}}^n, \varphi_n)\}$ such that

$$\liminf_{n\to\infty} R_n \geq R$$

and

$$\lim_{n\to\infty} \varepsilon_n(f_{\mathrm{A}}^n, f_{\mathrm{C}}^n, \varphi_n) = 0.$$

Then, the capacity C of the channel with action-dependent states defined as

$$C = \sup\{R : R \text{ is achievable.}\}.$$

For channels with action-dependent states, Weissman [2] showed the next theorem.

Theorem 1 (Theorem 1 in [2])**.** For stationary memoryless channels $P_{S|A}$ and $P_{Y|XS}$,

$$\begin{aligned} C &= \max_{(A,S,U,X,Y)\in\mathscr{D}_{\mathrm{M}}(P_{S|A},P_{Y|XS})} [I(U;Y) - I(U;S|A)] \\ &= \max_{(A,S,U,X,Y)\in\mathscr{D}_{\mathrm{M}}(P_{S|A},P_{Y|XS})} [I(A,U;Y) - I(U;S|A)], \end{aligned} \quad (1)$$

where $\mathscr{D}_{\mathrm{M}}(P_{S|A}, P_{Y|XS})$ is the set of all RVs (A,S,U,X,Y) with a probability distribution

$$\begin{aligned} P_{ASUXY}(a,s,u,x,y) =& P_A(a)P_{S|A}(s|a)P_{U|SA}(u|s,a) \\ &\times 1_{\{f(u,s)\}}(x)P_{Y|XS}(y|x,s) \end{aligned}$$

for some $P_A \in \mathscr{P}(\mathscr{A})$, $P_{U|SA} \in \mathscr{W}(\mathscr{U}|\mathscr{S}\times\mathscr{A})$, $f : \mathscr{U}\times\mathscr{S} \to \mathscr{X}$, $|\mathscr{U}| \leq |\mathscr{A}||\mathscr{S}||\mathscr{X}| + 1$, and $1_{\mathscr{X}}(x)$ denotes the indicator function defined as

$$1_{\mathscr{X}}(x) \triangleq \begin{cases} 1 & \text{if } x \in \mathscr{X}, \\ 0 & \text{if } x \notin \mathscr{X}. \end{cases}$$

Remark 1. The second equality of (1) always holds. This was shown in the beginning of the proof of [2, Theorem 1].

In order to characterize the capacity for general channels, we introduce the limit superior in probability and limit inferior in probability [9].

Definition 1 (Limit superior/inferior in probability)**.** For an arbitrary sequence of real-valued random variables $\{Z_n\}_{n=1}^{\infty}$,

$$\text{p-}\limsup_{n\to\infty} Z_n \triangleq \inf\left\{\alpha : \lim_{n\to\infty}\Pr\{Z_n > \alpha\} = 0\right\},$$

$$\text{p-}\liminf_{n\to\infty} Z_n \triangleq \sup\left\{\beta : \lim_{n\to\infty}\Pr\{Z_n < \beta\} = 0\right\}.$$

By using this notion, we introduce the next information measures [9], [10].

Definition 2 ((Conditional) mutual information spectrum–sup/inf)**.** For a given sequence of random variables $(\mathbf{X},\mathbf{Y},\mathbf{Z}) = \{(X^n,Y^n,Z^n)\}_{n=1}^{\infty}$, we define

$$\overline{I}(\mathbf{X};\mathbf{Y}) \triangleq \text{p-}\limsup_{n\to\infty}\frac{1}{n}\log\frac{P_{Y^n|X^n}(Y^n|X^n)}{P_{Y^n}(Y^n)},$$

$$\underline{I}(\mathbf{X};\mathbf{Y}) \triangleq \text{p-}\liminf_{n\to\infty}\frac{1}{n}\log\frac{P_{Y^n|X^n}(Y^n|X^n)}{P_{Y^n}(Y^n)},$$

$$\overline{I}(\mathbf{X};\mathbf{Y}|\mathbf{Z}) \triangleq \text{p-}\limsup_{n\to\infty}\frac{1}{n}\log\frac{P_{Y^n|X^nZ^n}(Y^n|X^n,Z^n)}{P_{Y^n|Z^n}(Y^n|Z^n)},$$

$$\underline{I}(\mathbf{X};\mathbf{Y}|\mathbf{Z}) \triangleq \text{p-}\liminf_{n\to\infty}\frac{1}{n}\log\frac{P_{Y^n|X^nZ^n}(Y^n|X^n,Z^n)}{P_{Y^n|Z^n}(Y^n|Z^n)}.$$

III. A GENERAL FORMULA FOR THE CAPACITY

In this section, we show a general formula for the capacity of channels with action-dependent states. The next theorem is our main result.

Theorem 2. For channels $\mathbf{P_{S|A}}$ and $\mathbf{P_{Y|XS}}$,

$$\begin{aligned} C &= \sup_{(\mathbf{A},\mathbf{S},\mathbf{U},\mathbf{X},\mathbf{Y})\in\mathscr{D}} [\underline{I}(\mathbf{U};\mathbf{Y}) - \overline{I}(\mathbf{U};\mathbf{S}|\mathbf{A})] \\ &= \sup_{(\mathbf{A},\mathbf{S},\mathbf{U},\mathbf{X},\mathbf{Y})\in\mathscr{D}} [\underline{I}(\mathbf{A},\mathbf{U};\mathbf{Y}) - \overline{I}(\mathbf{U};\mathbf{S}|\mathbf{A})], \end{aligned} \quad (2)$$

where $\mathscr{D}$ is the set of all sequences of RVs $(\mathbf{A},\mathbf{S},\mathbf{U},\mathbf{X},\mathbf{Y}) = \{(A^n,S^n,U^n,X^n,Y^n)\}_{n=1}^{\infty}$ such that (A^n,S^n,U^n,X^n,Y^n) obeys a probability distribution $P_{A^n}\cdot P_{S^n|A^n}\cdot P_{U^n|S^nA^n}\cdot P_{X^n|U^nS^n}\cdot P_{Y^n|X^nS^n}$ for some distributions $P_{A^n} \in \mathscr{P}(\mathscr{A}^n)$, $P_{U^n|S^nA^n} \in \mathscr{W}(\mathscr{U}^n|\mathscr{S}^n \times \mathscr{A}^n)$, and $P_{X^n|U^nS^n} \in \mathscr{W}(\mathscr{X}^n|\mathscr{U}^n \times \mathscr{S}^n)$.

Remark 2. The second equality of (2) always holds. This can be shown as follows: for any $(\mathbf{A},\mathbf{S},\mathbf{U},\mathbf{X},\mathbf{Y}) \in \mathscr{D}$, we have

$$\begin{aligned} \underline{I}(\mathbf{A},\mathbf{U};\mathbf{Y}) - \overline{I}(\mathbf{U};\mathbf{S}|\mathbf{A}) =& \underline{I}(\mathbf{A},\mathbf{U};\mathbf{Y}) - \overline{I}(\mathbf{A},\mathbf{U};\mathbf{S}|\mathbf{A}) \\ =& \underline{I}(\hat{\mathbf{U}};\mathbf{Y}) - \overline{I}(\hat{\mathbf{U}};\mathbf{S}|\mathbf{A}), \end{aligned}$$

where $\hat{\mathbf{U}} = \{(A^n,U^n)\}_{n=1}^{\infty}$. Since, for any $(\mathbf{A},\mathbf{U},\mathbf{S},\mathbf{X},\mathbf{Y}) \in \mathscr{D}$,

$(\mathbf{A},\hat{\mathbf{U}},\mathbf{S},\mathbf{X},\mathbf{Y})$ is also in the set $\mathscr{D}$, we have

$$\sup_{(\mathbf{A},\mathbf{U},\mathbf{S},\mathbf{X},\mathbf{Y})\in\mathscr{D}}[\underline{I}(\mathbf{A},\mathbf{U};\mathbf{Y})-\overline{I}(\mathbf{U};\mathbf{S}|\mathbf{A})] \le \sup_{(\mathbf{A},\hat{\mathbf{U}},\mathbf{S},\mathbf{X},\mathbf{Y})\in\mathscr{D}}[\underline{I}(\hat{\mathbf{U}};\mathbf{Y})-\overline{I}(\hat{\mathbf{U}};\mathbf{S}|\mathbf{A})].$$

The inequality in the opposite direction is trivial because, for any $(\mathbf{A},\mathbf{U},\mathbf{S},\mathbf{X},\mathbf{Y})\in\mathscr{D}$,

$$\underline{I}(\mathbf{A},\mathbf{U};\mathbf{Y})-\overline{I}(\mathbf{U};\mathbf{S}|\mathbf{A}) \ge \underline{I}(\mathbf{U};\mathbf{Y})-\overline{I}(\mathbf{U};\mathbf{S}|\mathbf{A}).$$

Thus, we have the second equality of (2).

We show the proof of this theorem in Section IV. In this theorem, by assuming that $|\mathscr{A}|=1$, we have the next corollary which shows a general formula of the capacity for the Gel'fand-Pinsker coding problem.

Corollary 1 ([8, Theorem 1])**.** Consider $\mathbf{P_S}=\{P_{S^n}\}_{n=1}^{\infty}$ and a channel $\mathbf{P_{Y|XS}}=\{P_{Y^n|X^nS^n}\}_{n=1}^{\infty}$ which are sequence of distributions $P_{S^n}\in\mathscr{P}(\mathscr{S}^n)$ and $P_{Y^n|X^nS^n}\in\mathscr{W}(\mathscr{Y}^n|\mathscr{X}^n\times\mathscr{S}^n)$, respectively. Then, we have

$$C=\sup_{(\mathbf{S},\mathbf{U},\mathbf{X},\mathbf{Y})\in\mathscr{D}_{\mathrm{GP}}}[\underline{I}(\mathbf{U};\mathbf{Y})-\overline{I}(\mathbf{U};\mathbf{S})],$$

where $\mathscr{D}_{\mathrm{GP}}$ is the set of all sequences of RVs $(\mathbf{S},\mathbf{U},\mathbf{X},\mathbf{Y})=\{(S^n,U^n,X^n,Y^n)\}_{n=1}^{\infty}$ such that (S^n,U^n,X^n,Y^n) obeys a probability distribution $P_{S^n}\cdot P_{U^n|S^n}\cdot P_{X^n|U^nS^n}\cdot P_{Y^n|X^nS^n}$ for some distributions $P_{U^n|S^n}\in\mathscr{W}(\mathscr{U}^n|\mathscr{S}^n)$ and $P_{X^n|U^nS^n}\in\mathscr{W}(\mathscr{X}^n|\mathscr{U}^n\times\mathscr{S}^n)$.

According to Theorem 2, we also have the next corollary which shows the capacity for nonstationary memoryless channels.

Corollary 2. For nonstationary memoryless channels $\mathbf{P_{S|A}}$ and $\mathbf{P_{Y|XS}}$ which are characterized by sequences of channels $\{P_{S_i|A_i}\}_{i=1}^{\infty}$ and $\{P_{Y_i|X_iS_i}\}_{i=1}^{\infty}$ as $P_{S^n|A^n}=\prod_{i=1}^n P_{S_i|A_i}$ and $P_{Y^n|X^nS^n}=\prod_{i=1}^n P_{Y_i|X_iS_i}$, let $(A_i^*,S_i^*,U_i^*,X_i^*,Y_i^*)$ be an optimal RV which maximizes

$$\max_{(A_i,S_i,U_i,X_i,Y_i)\in\mathscr{D}_{\mathrm{M}}(P_{S_i|A_i},P_{Y_i|X_iS_i})}[I(U_i;Y_i)-I(U_i;S_i|A_i)],$$

and $(A_i^{**},S_i^{**},U_i^{**},X_i^{**},Y_i^{**})$ be an optimal RV which maximizes

$$\max_{(A_i,S_i,U_i,X_i,Y_i)\in\mathscr{D}_{\mathrm{M}}(P_{S_i|A_i},P_{Y_i|X_iS_i})}[I(A_i,U_i;Y_i)-I(U_i;S_i|A_i)].$$

Assume that any one of the next four limits exist

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n I(U_i^*;Y_i^*),\ \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n I(U_i^*;S_i^*|A_i^*),$$
$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n I(A_i^{**},U_i^{**};Y_i^{**}),\ \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n I(U_i^{**};S_i^{**}|A_i^{**}).$$

Then, it holds that

$$\begin{aligned}C&=\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n[I(U_i^*;Y_i^*)-I(U_i^*;S_i^*|A_i^*)]\\&=\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n[I(A_i^{**},U_i^{**};Y_i^{**})-I(U_i^{**};S_i^{**}|A_i^{**})].\end{aligned}$$

Theorem 1 immediately obtained from this corollary.

Proof of Corollary 2: Note that, according to Remark 1, $I(U_i^*;Y_i^*)-I(U_i^*;S_i^*|A_i^*)=I(A_i^{**},U_i^{**};Y_i^{**})-I(U_i^{**};S_i^{**}|A_i^{**})$. Hence, we have

$$\begin{aligned}&\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n[I(U_i^*;Y_i^*)-I(U_i^*;S_i^*|A_i^*)]\\&=\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n[I(A_i^{**},U_i^{**};Y_i^{**})-I(U_i^{**};S_i^{**}|A_i^{**})].\end{aligned}\tag{3}$$

First, we show that

$$\begin{aligned}C&\ge\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n[I(U_i^*;Y_i^*)-I(U_i^*;S_i^*|A_i^*)]\\&=\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n[I(A_i^{**},U_i^{**};Y_i^{**})-I(U_i^{**};S_i^{**}|A_i^{**})].\end{aligned}\tag{4}$$

Let $((A^n)^*,(S^n)^*,(U^n)^*,(X^n)^*,(Y^n)^*)$ be the n-length sequence of RVs such that the i-th element is $(A_i^*,S_i^*,U_i^*,X_i^*,Y_i^*)$, and all elements $\{(A_i^*,S_i^*,U_i^*,X_i^*,Y_i^*)\}_{i=1}^n$ are independent of each other. Then, for the sequence of RVs $(\mathbf{A}^*,\mathbf{S}^*,\mathbf{U}^*,\mathbf{X}^*,\mathbf{Y}^*)=\{((A^n)^*,(S^n)^*,(U^n)^*,(X^n)^*,(Y^n)^*)\}_{n=1}^{\infty}$, we have

$$\begin{aligned}C&\ge\underline{I}(\mathbf{U}^*;\mathbf{Y}^*)-\overline{I}(\mathbf{U}^*;\mathbf{S}^*|\mathbf{A}^*)\\&=\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n I(U_i^*;Y_i^*)-\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^n I(U_i^*;S_i^*|A_i^*),\end{aligned}\tag{5}$$

where the equality comes from Chebyshev's inequality. Similarly, we have

$$C\ge\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n I(A_i^{**},U_i^{**};Y_i^{**})-\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^n I(U_i^{**};S_i^{**}|A_i^{**}).\tag{6}$$

If there exists a limit for at least one term of the right-hand side of (5), we have

$$\begin{aligned}C&\ge\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n I(U_i^*;Y_i^*)-\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^n I(U_i^*;S_i^*|A_i^*)\\&=\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n[I(U_i^*;Y_i^*)-I(U_i^*;S_i^*|A_i^*)]\\&=\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n[I(A_i^{**},U_i^{**};Y_i^{**})-I(U_i^{**};S_i^{**}|A_i^{**})],\end{aligned}$$

where the first equality follows from the fact that $\liminf_{n\to\infty}a_n+\lim_{n\to\infty}b_n=\liminf_{n\to\infty}(a_n+b_n)$ if b_n converges, and the second equality comes from (3). Similarly, if there exists a limit for at least one term of the right-hand side of (6), we have

$$\begin{aligned}C&\ge\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n I(A_i^{**},U_i^{**};Y_i^{**})-\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^n I(U_i^{**};S_i^{**}|A_i^{**})\\&=\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n[I(A_i^{**},U_i^{**};Y_i^{**})-I(U_i^{**};S_i^{**}|A_i^{**})]\\&=\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n[I(U_i^*;Y_i^*)-I(U_i^*;S_i^*|A_i^*)].\end{aligned}$$

Thus, we have (4) under the assumption of the corollary.

Next, we show the inequality in the opposite direction. By following the proof of [10, Theorem 3.5.2], we can show that

$$\underline{I}(\mathbf{U};\mathbf{Y}) \le \liminf_{n\to\infty} \frac{1}{n} I(U^n;Y^n),$$
$$\bar{I}(\mathbf{U};\mathbf{S}|\mathbf{A}) \ge \limsup_{n\to\infty} \frac{1}{n} I(U^n;S^n|A^n).$$

Thus, we have

$$C = \sup_{(\mathbf{A},\mathbf{S},\mathbf{U},\mathbf{X},\mathbf{Y})\in\mathscr{D}} [\underline{I}(\mathbf{U};\mathbf{Y}) - \bar{I}(\mathbf{U};\mathbf{S}|\mathbf{A})]$$
$$\le \sup_{(\mathbf{A},\mathbf{S},\mathbf{U},\mathbf{X},\mathbf{Y})\in\mathscr{D}} \liminf_{n\to\infty} \frac{1}{n}[I(U^n;Y^n) - I(U^n;S^n|A^n)]. \quad (7)$$

On the other hand, by using similar inequalities [2, (8) – (15)] and using U^n in place of M in these inequalities, for any $(\mathbf{A},\mathbf{S},\mathbf{U},\mathbf{X},\mathbf{Y}) \in \mathscr{D}$, we have

$$I(U^n;Y^n) - I(U^n;S^n|A^n) \le \sum_{i=1}^{n} [I(U_i^*;Y_i^*) - I(U_i^*;S_i^*|A_i^*)]. \quad (8)$$

Hence, by combining (3), (7) and (8), we have

$$C \le \liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} [I(U_i^*;Y_i^*) - I(U_i^*;S_i^*|A_i^*)]$$
$$= \liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} [I(A_i^{**},U_i^{**};Y_i^{**}) - I(U_i^{**};S_i^{**}|A_i^{**})].$$

This completes the proof of Corollary 2. ■

IV. Proof of Theorem 2

The proof of Theorem 2 is based on the proof of [8, Theorem 1]. We can prove the converse part of the theorem in a similar way to the proof of [8, Theorem 1]. Hence, we only show the direct part of the theorem. To this end, we introduce a necessary lemma.

For an arbitrarily sequence of RVs $(\mathbf{A},\mathbf{S},\mathbf{U},\mathbf{X},\mathbf{Y}) \in \mathscr{D}$, any $\gamma > 0$, and any integer $M_n > 0$, we define the next two sets

$$\mathscr{T}_n^{(1)} \triangleq \Big\{ (a^n,u^n,y^n) \in \mathscr{A}^n \times \mathscr{U}^n \times \mathscr{Y}^n : \frac{1}{n}\log \frac{P_{Y^n|U^nA^n}(y^n|u^n,a^n)}{P_{Y^n}(y^n)} \ge R_n + \bar{I}(\mathbf{U};\mathbf{S}|\mathbf{A}) + 3\gamma \Big\},$$
$$\mathscr{T}_n^{(2)} \triangleq \Big\{ (a^n,s^n,u^n) \in \mathscr{A}^n \times \mathscr{S}^n \times \mathscr{U}^n : \frac{1}{n}\log \frac{P_{U^n|S^nA^n}(u^n|s^n,a^n)}{P_{U^n|A^n}(u^n|a^n)} \le \bar{I}(\mathbf{U};\mathbf{S}|\mathbf{A}) + \gamma \Big\},$$

where $R_n = \frac{1}{n}\log M_n$. We also define

$$\pi_{1,n} \triangleq \Pr\{(A^n,U^n,Y^n) \notin \mathscr{T}_n^{(1)}\},$$
$$\pi_{2,n} \triangleq \Pr\{(A^n,S^n,U^n) \notin \mathscr{T}_n^{(2)}\}.$$

Now we show the lemma.

Lemma 1. For any RVs $(\mathbf{A},\mathbf{S},\mathbf{U},\mathbf{X},\mathbf{Y}) \in \mathscr{D}$, any $n > 0$, $\gamma > 0$, and any integer $M_n > 0$, there exists a code $(f_{\mathrm{A}}^n, f_{\mathrm{C}}^n, \varphi_n)$ such that

$$\varepsilon_n(f_{\mathrm{A}}^n, f_{\mathrm{C}}^n, \varphi_n) \le 2\pi_{1,n}^{1/2} - \pi_{1,n} + \pi_{2,n} + \exp\{-\exp\{n\gamma\}\} + 2\exp\{-n\gamma\}. \quad (9)$$

Remark 3. The bound (9) is tighter than that in [8, Lemma 9]. In [8, Lemma 9], the second term of the right-hand side of (9) does not appear. As we will see in the proof of this lemma, this term comes from the complement of an error event.

Proof: For an arbitrarily given RVs $(\mathbf{A},\mathbf{S},\mathbf{U},\mathbf{X},\mathbf{Y}) \in \mathscr{D}$ and $\gamma > 0$, we define the mapping $\eta : \mathscr{A}^n \times \mathscr{S}^n \times \mathscr{U}^n \to \mathbb{R}_+$ as

$$\eta(a^n,s^n,u^n) \triangleq \sum_{x^n\in\mathscr{X}^n} \sum_{y^n\in\mathscr{Y}^n:(a^n,u^n,y^n)\notin\mathscr{T}_n^{(1)}} P_{X^n|U^nS^n}(x^n|u^n,s^n) \times P_{Y^n|X^nS^n}(y^n|x^n,s^n),$$

and define the set $\mathscr{B}$ as

$$\mathscr{B} \triangleq \Big\{ (a^n,s^n,u^n) \in \mathscr{A}^n \times \mathscr{S}^n \times \mathscr{U}^n : \eta(a^n,s^n,u^n) \le \pi_{1,n}^{1/2} \Big\}.$$

Then, we randomly generate a code as follows:

Random generation of $\mathscr{C}_{\mathrm{a}}$: For each $m \in \mathscr{M}_n$, randomly and independently generate sequences $a^n(m)$, each drawn according to P_{A^n}. We denote the set of sequences $\{a^n(1), a^n(2), \cdots, a^n(M_n)\}$ by $\mathscr{C}_{\mathrm{a}}$.

Random generation of $\mathscr{C}_{\mathrm{c}}(m)$: Let $\tilde{\mathscr{M}}_n$ be the set of integers $\{1,2,\cdots,\tilde{M}_n\}$, where $\tilde{M}_n = \lceil \exp\{n(\bar{I}(\mathbf{U};\mathbf{S}|\mathbf{A}) + 2\gamma)\} \rceil$. Then, for each $m \in \mathscr{M}_n$ and $l \in \tilde{\mathscr{M}}_n$, randomly and independently generate sequences $u^n(l)$, each drawn according to $P_{U^n|A^n}(\cdot|a^n(m))$ where $a^n(m)$ is the element of $\mathscr{C}_{\mathrm{a}}$ corresponding to m. We denote the set of sequences $\{u^n(1), \cdots, u^n(\tilde{M}_n)\}$ by $\mathscr{C}_{\mathrm{c}}(m)$.

Action encoder: For a given $m \in \mathscr{M}_n$, choose $a^n(m) \in \mathscr{C}_{\mathrm{a}}$ as the action.

Channel encoder: We use the following two-step encoding.

1) For a given $m \in \mathscr{M}_n$ and the state sequence $s^n \in \mathscr{S}^n$, finds the sequence $u^n(\hat{l}) \in \mathscr{C}_{\mathrm{c}}(m)$ with the smallest index $\hat{l}$ satisfying
$$(a^n(m), s^n, u^n(\hat{l})) \in \mathscr{B}.$$
If there is no such sequence, set $\hat{l} = 1$.
2) For the index $\hat{l}$ found in the step 1), randomly generate a codeword x^n drawn from $P_{X^n|U^nS^n}(x^n|u^n(\hat{l}),s^n)$, and input it to the channel $P_{Y^n|X^nS^n}$.

Decoder: For a given channel output $y^n \in \mathscr{Y}^n$, decoder determines the unique $\hat{m} \in \mathscr{M}_n$ as the message such that

$$(a^n(\hat{m}), u^n(l), y^n) \in \mathscr{T}_n^{(1)}$$

for some $u^n(l) \in \mathscr{C}_{\mathrm{c}}(\hat{m})$. If there is no such unique $\hat{m} \in \mathscr{M}_n$, an error is declared.

We now show an upper bound of the error probability of the code generated as the above procedure. We assume that $m = 1$ because of the symmetry of the random generation, and define the following three events.

$$\mathscr{E}_1 \triangleq \{ {}^{\forall}u^n \in \mathscr{C}_{\mathrm{c}}(1), (A^n(1), S^n, u^n) \notin \mathscr{B} \},$$
$$\mathscr{E}_2 \triangleq \{ (A^n(1), U^n(L), Y^n) \notin \mathscr{T}_n^{(1)} \},$$
$$\mathscr{E}_3 \triangleq \{ {}^{\exists}\tilde{m} \ne 1, {}^{\exists}\tilde{u}^n \in \mathscr{C}_{\mathrm{c}}(\tilde{m}), (A^n(\tilde{m}), \tilde{u}^n, Y^n) \in \mathscr{T}_n^{(1)} \},$$

where L denotes the random index chosen in the channel encoder. Then, the error probability can be bounded as

$$E[\varepsilon_n(f_{\mathrm{A}}^n, f_{\mathrm{C}}^n, \varphi_n)] \le \Pr\{\mathscr{E}_1\} + \Pr\{\mathscr{E}_2 \cap \mathscr{E}_1^c\} + \Pr\{\mathscr{E}_3\}, \quad (10)$$

where $E[\cdot]$ denotes the expectation over the random choice of codes, and $\{\cdot\}^c$ denotes the complement of the set. We can upper bound the first and the third terms of (10) in a similar way to the proof of [8, Theorem 1] as

$$\Pr\{\mathscr{E}_1\} \leq \pi_{1,n}^{1/2} + \pi_{2,n} + \exp\{-\exp\{n\gamma\}\}, \quad (11)$$

and

$$\Pr\{\mathscr{E}_3\} \leq \exp\{-n\gamma\} + \exp\{-n(\overline{I}(\mathbf{U};\mathbf{S}|\mathbf{A}) + 3\gamma)\}. \quad (12)$$

We show an upper bound of $\Pr\{\mathscr{E}_2 \cap \mathscr{E}_1^c\}$. We denote $U^n(L)$ as $F_{A^n(1),\mathscr{C}_c(1)}(S^n) \in \mathscr{U}^n$ to make clear from which the randomness comes. We denote c_c as a specific realization of $\mathscr{C}_c(1)$. Then, we have

$$\begin{aligned}
&\Pr\{\mathscr{E}_2 \cap \mathscr{E}_1^c\} \\
&= \Pr\Big\{\{(A^n(1), F_{A^n(1),\mathscr{C}_c(1)}(S^n), Y^n) \notin \mathscr{T}_n^{(1)}\} \cap \\
&\quad \{{}^\exists u^n \in \mathscr{C}_c(1), (A^n(1), S^n, u^n) \in \mathscr{B}\}\Big\} \\
&= \sum_{\substack{a^n,s^n,c_c,x^n,y^n:(a^n,F_{a^n,c_c}(s^n),y^n)\notin\mathscr{T}_n^{(1)},\\ {}^\exists u^n\in c_c,(a^n,s^n,u^n)\in\mathscr{B}}} \\
&\quad \times \Pr\{A^n(1)=a^n, S^n=s^n, \mathscr{C}_c(1)=c_c, X^n=x^n, Y^n=y^n\},
\end{aligned}$$

where

$$\begin{aligned}
&\Pr\{A^n(1)=a^n, S^n=s^n, \mathscr{C}_c(1)=c_c, X^n=x^n, Y^n=y^n\} \\
&= P_{A^n}(a^n) P_{S^n|A^n}(s^n|a^n) \left(\prod_{u^n \in c_c} P_{U^n|A^n}(u^n|a^n) \right) \\
&\quad \times P_{X^n|U^nS^n}(x^n|F_{a^n,c_c}(s^n), s^n) P_{Y^n|X^nS^n}(y^n|x^n,s^n).
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&\Pr\{\mathscr{E}_2 \cap \mathscr{E}_1^c\} \\
&= \sum_{a^n,s^n,c_c:{}^\exists u^n\in c_c,(a^n,s^n,u^n)\in\mathscr{B}} \Pr\{A^n(1)=a^n, S^n=s^n, \mathscr{C}_c(1)=c_c\} \\
&\quad \times \sum_{x^n} \sum_{y^n:(a^n,F_{a^n,c_c}(s^n),y^n)\notin\mathscr{T}_n^{(1)}} P_{X^n|U^nS^n}(x^n|F_{a^n,c_c}(s^n), s^n) \\
&\quad \times P_{Y^n|X^nS^n}(y^n|x^n,s^n) \\
&= \sum_{a^n,s^n,c_c:{}^\exists u^n\in c_c,(a^n,s^n,u^n)\in\mathscr{B}} \Pr\{A^n(1)=a^n, S^n=s^n, \mathscr{C}_c(1)=c_c\} \\
&\quad \times \eta(a^n, s^n, F_{a^n,c_c}(s^n)) \\
&\leq \Pr\{\mathscr{E}_1^c\} \pi_{1,n}^{1/2}, \qquad (13)
\end{aligned}$$

where the last inequality comes from the fact that, for the sequence (a^n, s^n, c_c) such that $(a^n, s^n, u^n) \in \mathscr{B}$ for some $u^n \in c_c$, the sequence $(a^n, s^n, F_{a^n,c_c}(s^n))$ must satisfy $(a^n, s^n, F_{a^n,c_c}(s^n)) \in \mathscr{B}$.

Now by combining (11) and (13), we have

$$\begin{aligned}
&\Pr\{\mathscr{E}_1\} + \Pr\{\mathscr{E}_2 \cap \mathscr{E}_1^c\} \\
&\leq \Pr\{\mathscr{E}_1\} + \Pr\{\mathscr{E}_1^c\} \pi_{1,n}^{1/2} \\
&= \pi_{1,n}^{1/2} + (1 - \pi_{1,n}^{1/2}) \Pr\{\mathscr{E}_1\} \\
&\leq 2\pi_{1,n}^{1/2} - \pi_{1,n} + (1 - \pi_{1,n}^{1/2})(\pi_{2,n} + \exp\{-\exp\{n\gamma\}\}). \quad (14)
\end{aligned}$$

Finally by combining (14) and (12), we have

$$\begin{aligned}
E[\varepsilon_n(f_A^n, f_C^n, \varphi_n)] \leq& 2\pi_{1,n}^{1/2} - \pi_{1,n} + (1 - \pi_{1,n}^{1/2})(\pi_{2,n} \\
&+ \exp\{-\exp\{n\gamma\}\}) + \exp\{-n\gamma\} \\
&+ \exp\{-n(\overline{I}(\mathbf{U};\mathbf{S}|\mathbf{A}) + 3\gamma)\},
\end{aligned}$$

or more simply

$$\begin{aligned}
E[\varepsilon_n(f_A^n, f_C^n, \varphi_n)] \leq& 2\pi_{1,n}^{1/2} - \pi_{1,n} + \pi_{2,n} + \exp\{-\exp\{n\gamma\}\} \\
&+ 2\exp\{-n\gamma\}.
\end{aligned}$$

Hence, there exists a code $(f_A^n, f_C^n, \varphi_n)$ satisfying (9). ■

Now, we prove the direct part of Theorem 2.

Proof of the direct part: For any fixed $(\mathbf{A}, \mathbf{S}, \mathbf{U}, \mathbf{X}, \mathbf{Y}) \in \mathscr{D}$ and $\gamma > 0$, we set

$$M_n = \lfloor \exp\{n(\underline{I}(\mathbf{A},\mathbf{U};\mathbf{Y}) - \overline{I}(\mathbf{U};\mathbf{S}|\mathbf{A}) - 4\gamma)\} \rfloor.$$

Then, according to Lemma 1, there exists a code such that

$$\liminf_{n\to\infty} R_n \geq \underline{I}(\mathbf{A},\mathbf{U};\mathbf{Y}) - \overline{I}(\mathbf{U};\mathbf{S}|\mathbf{A}) - 4\gamma, \quad (15)$$

and the error probability satisfies (9). Since

$$R_n \leq \underline{I}(\mathbf{A},\mathbf{U};\mathbf{Y}) - \overline{I}(\mathbf{U};\mathbf{S}|\mathbf{A}) - 4\gamma,$$

we have

$$\pi_{1,n} \leq \Pr\left\{ \frac{1}{n} \log \frac{P_{Y^n|U^nA^n}(Y^n|U^n,A^n)}{P_{Y^n}(Y^n)} < \underline{I}(\mathbf{A},\mathbf{U};\mathbf{Y}) - \gamma \right\},$$

and, according to the definition of $\pi_{2,n}$,

$$\pi_{2,n} = \Pr\left\{ \frac{1}{n} \log \frac{P_{U^n|S^nA^n}(U^n|S^n,A^n)}{P_{U^n|A^n}(U^n|A^n)} > \overline{I}(\mathbf{U};\mathbf{S}|\mathbf{A}) + \gamma \right\}.$$

Thus, according to the definition of the limit superior and inferior in probability, $\pi_{1,n}$ and $\pi_{2,n}$ goes to zero as the block length tends to infinity. Then, according to (9), the error probability also goes to zero as the block length tends to infinity. Hence, $\underline{I}(\mathbf{A},\mathbf{U};\mathbf{Y}) - \overline{I}(\mathbf{U};\mathbf{S}|\mathbf{A}) - 4\gamma$ is achievable. Since $(\mathbf{A}, \mathbf{S}, \mathbf{U}, \mathbf{X}, \mathbf{Y}) \in \mathscr{D}$ and γ are arbitrarily, we have

$$\begin{aligned}
C &\geq \sup_{(\mathbf{A},\mathbf{S},\mathbf{U},\mathbf{X},\mathbf{Y})\in\mathscr{D}} [\underline{I}(\mathbf{A},\mathbf{U};\mathbf{Y}) - \overline{I}(\mathbf{U};\mathbf{S}|\mathbf{A})] \\
&= \sup_{(\mathbf{A},\mathbf{S},\mathbf{U},\mathbf{X},\mathbf{Y})\in\mathscr{D}} [\underline{I}(\mathbf{U};\mathbf{Y}) - \overline{I}(\mathbf{U};\mathbf{S}|\mathbf{A})],
\end{aligned}$$

where the equality follows from Remark 2. This completes the proof of the direct part. ■

REFERENCES

[1] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2012.

[2] T. Weissman, "Capacity of channels with action-dependent states," *IEEE Trans. Inform. Theory*, vol. 56, no. 11, pp. 5396–5411, Nov. 2010.

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

[4] K. Kittichokechai, T. Oechtering, and M. Skoglund, "On the capacity of a channel with action-dependent state and reversible input," *Proc. IEEE Int. Symp. on Inform. Theory*, pp. 331–335, July–Aug. 2011.

[5] C. Choudhuri and U. Mitra, "Action dependent strictly causal state communication," *Proc. IEEE Int. Symp. on Inform. Theory*, pp. 3058–3062, Jul. 2012.

[6] B. Ahmadi and O. Simeone, "On channels with action-dependent states," *Arxiv preprint arXiv:1202.4438*, Jul. 2012.

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

[8] V. Y. F. Tan, "The capacity of the general Gel'fand-Pinsker channel and achievable second-order coding rates," *arXiv preprint arXiv:1210.1091*, 2012.

[9] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.

[10] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer, 2003.