# Communication Over Discrete Channels Subject to State Obfuscation

Ligong Wang, *Senior Member, IEEE*, and Gregory W. Wornell, *Fellow, IEEE*

***Abstract*— We consider communication over a state-dependent discrete memoryless channel subject to a constraint that the output sequence must be nearly independent of the state sequence. We consider both cases where the transmitter knows (causally or noncausally) and where it does not know the states. When it does not know the states, we show that capacity can increase when the encoder uses some source of randomness that is not shared with the decoder. We consider three different cases for the state sequence: where it is independent and identically distributed across channel uses, where it is quasi-static, and where it has memory but is not quasi-static. We present single-letter capacity formulas for most combinations of the above scenarios, and also provide some illustrative examples.**

***Index Terms*— Channel capacity, channels with memory, quasi-static, state-dependent channel, state masking.**

## I. Introduction

STATE-DEPENDENT channels have been extensively studied in information theory [1], [2], [3]. The present work considers communication over a discrete state-dependent channel, with an additional requirement that the channel state should remain unknown to the receiver. A representative application for such a model is a scenario where the transmitter wishes to conceal its physical location: its location may affect the statistics of the channel to the receiver, hence can be modeled as a channel state.

The problem we study is closely related to "state masking" and, to a lesser extent, "state amplification" [4], [5], [6], [7], [8]. Consider a state-dependent discrete memoryless channel (DMC) where, given input $X = x$ and state $S = s$, the probability for the output $Y$ to equal $y$ is given by $W(y|x, s)$. Assume that the state is independent and identically distributed (IID) across channel uses according to a known distribution. The state-masking constraint considered in [4] is

$$\lim_{n\to\infty} \frac{1}{n} I(S^n; Y^n) \le E \tag{1}$$

for some parameter $E$, where $n$ denotes the number of times the channel is used. When channel-state information (CSI) is available noncausally to the transmitter (i.e., the transmitter knows the realization of $S^n$ before sending any input to the channel), a communication rate $R$ is achievable under the above constraint if, and only if [4, Theorem 2]

$$R \le I(U;Y) - I(U;S) \tag{2}$$

for some auxiliary random variable $U$ such that $U \multimap\!\!-\, (X,S) \multimap\!\!-\, Y$ forms a Markov chain, and that

$$I(S;U,Y) \le E. \tag{3}$$

Note that (2) is the Gel'fand-Pinsker rate expression [2], while the condition (3) concerns $I(S;U,Y)$ and not $I(S;Y)$.

In the present paper we are interested in problems where the states must be almost completely concealed from the receiver, namely, where the limit in (1) must equal zero. Our capacity formula in the case where CSI is available to the transmitter then follows almost immediately from [4]. We also consider situations where CSI is not available and derive similar capacity formulas. As we shall see, capacity differs between the cases where the transmitter must use a deterministic encoder and where it may use a stochastic encoder; we provide single-letter capacity formulas for both cases. In all three cases, we show that capacity is not affected by the state distribution: the state-obfuscation communication capacity is determined by the channel law $W(\cdot|\cdot,\cdot)$ alone.

We go beyond IID states to study two other scenarios: where the state is quasi-static, i.e., it is randomly generated and remains constant during the entire transmission, and where the state has memory (while not being quasi-static). Our motivation for studying these scenarios is two-fold. From a practical perspective, we recall that the state can model attributes such as the location of the sender, which should not change fast compared to the communication timescale, therefore IID states are not an appropriate model for such applications. On the mathematical side, since we know that the state distribution does not affect capacity as long as it is IID across channel uses, it is worth understanding whether its dependence between channel uses will affect capacity or not.

In the quasi-static scenario, we impose the state obfuscation constraint that $I(S;Y^n)$ approach zero; note that we do not divide it by $n$ as in (1), the latter trivially approaches zero as $n \to \infty$. When CSI is available to the transmitter, or when CSI is not available and the transmitter must use a deterministic encoder, the capacity turns out to be the same as in the IID-state scenario. Interestingly, when CSI is not available and

Manuscript received 28 February 2024; accepted 5 July 2024. Date of publication 23 July 2024; date of current version 22 November 2024. This work was supported in part by the Office of Naval Research (ONR) under Grant N000014-23-1-2803. An earlier version of this paper was presented in part at the 2020 International Zurich Seminar on Information and Communication [DOI: 10.3929/ethz-b-000402954]. *(Corresponding author: Ligong Wang.)*

Ligong Wang is with the Department of Information Technology and Electrical Engineering, ETH Zurich, 8092 Zurich, Switzerland (e-mail: ligwang@isi.ee.ethz.ch).

Gregory W. Wornell is with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: gww@mit.edu).

Communicated by Jun Chen, Associate Editor for Shannon Theory and Information Measures.

Digital Object Identifier 10.1109/TIT.2024.3432573

the transmitter may use a stochastic encoder, capacity can exceed that of the IID-state scenario, as we demonstrate via an example.

For states with memory, we restrict our attention to those for which the conditional probability of any state realization given the past is bounded away from zero; the state-obfuscation constraint is (1) with $E = 0$, as in the IID-state case. An example of such a state process is a time-invariant Markov process whose transition matrix does not contain zeros.[1] We show that a DMC with such a state sequence has the same state-obfuscation communication capacity as the same DMC with IID states when the transmitter is deterministic without CSI, when it is stochastic without CSI, and when it is stochastic with *causal* CSI. The direct parts of these capacity results are carried over from the corresponding results on IID states. The converse parts, however, call for some nonstandard proof techniques.

We consider IID states in Section II, quasi-static states in Section III, and states with memory in Section IV. We then conclude the paper with some remarks in Section V.

## II. IID States

Consider a DMC with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ that is affected by a random state $S$, which takes values in the set $\mathcal{S}$. The sets $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{S}$ are all assumed to be finite. The channel law is, given input $x \in \mathcal{X}$ and state $s \in \mathcal{S}$, the probability of the output being $y \in \mathcal{Y}$ is $W(y|x, s)$.

In this section, we assume that the states are drawn IID across channel uses following a probability mass function $P_S$. Without loss of generality, we assume throughout that

$$\operatorname{supp}(P_S) = \mathcal{S}. \tag{4}$$

The message $M$ to be communicated is drawn from the set $\{1, \ldots, \lfloor 2^{nR} \rfloor\}$, where $n$ denotes the total number of channel uses, and $R$ the rate of communication in bits per channel use. The message is fed to an encoder, which in turn produces the channel input sequence $x^n$. We consider both cases where the state realizations are known and unknown to the transmitter, respectively. When the states are unknown to the transmitter, we further distinguish between deterministic and stochastic encoders; details are provided below. In all cases, the receiver tries to guess the message based on the channel outputs $y^n$ (the receiver has no CSI).

The state-obfuscation constraint we impose in the IID-state case is

$$\lim_{n\to\infty} \frac{1}{n} I(S^n; Y^n) = 0, \tag{5}$$

where the mutual information is computed for the joint distribution induced by the encoder and a uniformly drawn message. In fact, all results in this section will continue to hold when we replace (5) by the following stronger constraint: irrespectively of the distribution of the message, we require

$$I(S^n; Y^n) = 0 \quad \text{for all } n. \tag{6}$$

In the following, we shall prove the stronger versions of both converse and direct parts of our results, namely, we shall prove converse results under the constraint (5) for a uniform message, and direct results under the constraint (6) for any message distribution. But, for simplicity, the theorems will be presented only under the former constraint.

### A. With CSI

We consider both noncausal and causal CSI. A deterministic encoder with noncausal CSI is a mapping

$$f^{\text{NC-CSI}} \colon \{1, \ldots, \lfloor 2^{nR} \rfloor\} \times \mathcal{S}^n \to \mathcal{X}^n,\ (m, s^n) \mapsto x^n. \tag{7}$$

The transmitter can use a *stochastic* encoder that is chosen randomly according to some distribution over all mappings of the form (7). The distribution is known to the decoder, but the actual choice is not.

A deterministic encoder with causal CSI is a sequence of mappings

$$f_i^{\text{C-CSI}} \colon \{1, \ldots, \lfloor 2^{nR} \rfloor\} \times \mathcal{S}^i \to \mathcal{X},\ (m, s^i) \mapsto x_i, \tag{8}$$

$i \in \{1, \ldots, n\}$. The transmitter can use a stochastic encoder that is chosen randomly according to some distribution over all sequences of mappings of the form (8). Again, the distribution but not the choice is known to the decoder.

In both the noncausal and the causal cases, the decoder is a deterministic mapping[2]

$$g \colon \mathcal{Y}^n \to \{1, \ldots, \lfloor 2^{nR} \rfloor\},\ y^n \mapsto \hat{m}. \tag{9}$$

In both cases, a rate $R$ is said to be achievable if there exists a sequence of pairs of stochastic encoders and (deterministic) decoders as above such that, for a message $M$ uniformly drawn from $\{1, \ldots, \lfloor 2^{nR} \rfloor\}$, the probability that $\hat{M} \neq M$ tends to zero as $n$ grows to infinity, and, at the same time, (5) is satisfied. As usual, capacity is defined as the supremum over all achievable rates.

*Theorem 1:* The capacity of the channel when the transmitter has either noncausal or causal CSI is the same, and is given by

$$C_{\text{CSI}}^{\text{IID}} = \sup I(U; Y), \tag{10}$$

where $U$ takes values in some finite set $\mathcal{U}$, and the supremum is taken over joint probability distributions of the form

$$P_S(s) P_U(u) P_{X|US}(x|u, s) W(y|x, s) \tag{11}$$

subject to

$$I(S; U, Y) = 0. \tag{12}$$

*Proof:* It suffices to prove the converse in the noncausal case and the direct part in the causal case. The former follows directly from [4, Theorem 2] by noting that (12) requires that $U$ be independent of $S$.

To prove the direct part in the causal case, fix any joint distribution of the form (11), generate a random codebook $\{u^n(1), \ldots, u^n(\lfloor 2^{nR} \rfloor)\}$ IID according to $P_U$, and reveal it to

[1] A simple example where the condition is not satisfied is the aforementioned quasi-static states. Recall that (1) with $E = 0$ is trivially satisfied by quasi-static states.

[2] It is straightforward to show that there is no advantage in using a stochastic decoder.

the decoder. To send $m$, the encoder picks its input at time $i$, $i \in \{1, \ldots, n\}$, to be $x_i$ with probability $P_{X|US}(x_i|u_i(m), s_i)$ independently (conditional on $u^n(m)$ and $s^n$) of the inputs at other times.

We next show that, with probability one, the code generated as above satisfies (6). Note that (12) implies

$$P_{Y|US}(y|u,s) = P_{Y|U}(y|u) \tag{13}$$

for all $s, u, y$ such that $P_{SUY}(s,u,y) > 0$. When the code is used to transmit any message $M = m$, the probability of $Y^n = y^n$ and $S^n = s^n$, for any $y^n$ and $s^n$, can be written as

$$\begin{aligned}\Pr(S^n = s^n, Y^n = y^n | M = m)& \\ = \prod_{i=1}^{n} P_S(s_i) P_{Y|US}(y|u_i(m), s_i)& \qquad (14)\\ = \prod_{i=1}^{n} P_S(s_i) \cdot \prod_{i=1}^{n} P_{Y|U}(y_i|u_i(m)).& \qquad (15)\end{aligned}$$

Clearly, for any $m$ and any $n$,

$$I(S^n; Y^n | M = m) = 0. \tag{16}$$

It then follows that, irrespectively of the distribution of $M$,

$$\begin{aligned}I(S^n; Y^n) &\le I(S^n; Y^n, M) && (17)\\ &= I(S^n; M) + I(S^n; Y^n | M) && (18)\\ &= 0. && (19)\end{aligned}$$

It remains to analyze the probability of a decoding error, which is standard. Consider a channel whose input alphabet is $\mathcal{U}$, whose output alphabet is $\mathcal{Y}$, and whose transition law is given by $P_{Y|U}$ (so $X$ becomes part of the channel). Applying the standard proof as in, e.g., [9], to this new channel, we conclude that, for all $R < I(U;Y)$, there exists a sequence of codes (with nonzero probability of being generated) whose maximum error probability tends to zero as $n$ tends to infinity. It then follows that the average error probability must also tend to zero for any message distribution. ■

### B. No CSI, Deterministic Encoder

We next consider the case where no CSI is available to the transmitter, and where the encoder must be deterministic. Thus, instead of (7) or (8), the encoder is a deterministic mapping

$$f^{\text{det}} \colon \{1, \ldots, \lfloor 2^{nR} \rfloor\} \to \mathcal{X}^n, \;\; m \mapsto x^n. \tag{20}$$

The decoder remains to be of the form (9). Capacity is defined as in Section II-A, where the encoder with CSI is replaced by the deterministic encoder without CSI (20).

*Theorem 2:* When the transmitter has no CSI and must use a deterministic encoder, the capacity is given by

$$C_{\text{det}}^{\text{IID}} = \sup I(X;Y), \tag{21}$$

where the supremum is taken over joint distributions of the form

$$P_S(s) P_X(x) W(y|x,s) \tag{22}$$

subject to

$$I(S; X, Y) = 0. \tag{23}$$

*Proof:* For the direct part, we generate a codebook $\{x^n(1), \ldots, x^n(\lfloor 2^{nR} \rfloor)\}$ by generating each codeword IID according to $P_X$. The analysis is a straightforward modification of that in the proof of Theorem 1 and hence omitted.

For converse, take any sequence of codes with vanishing error probability as $n \to \infty$ and satisfying (5). By the fact that $X^n$ is a deterministic function of the message $M$, and by Fano's inequality, we have

$$H(X^n|Y^n) \le H(M|Y^n) \le n\epsilon_n \tag{24}$$

for some $\epsilon_n \downarrow 0$ as $n \to \infty$. We thus have

$$\begin{aligned}I(S^n; X^n, Y^n) &= I(S^n; X^n | Y^n) + I(S^n; Y^n) && (25)\\ &\le H(X^n|Y^n) + I(S^n; Y^n) && (26)\\ &\le n(\epsilon_n + \epsilon_n'), && (27)\end{aligned}$$

where $\epsilon_n' \downarrow 0$ as $n \to \infty$, and the last step follows by (24) and the constraint (5). We also have

$$\begin{aligned}I(S^n; X^n, Y^n) &= H(S^n) - H(S^n | X^n, Y^n) && (28)\\ &= \sum_{i=1}^{n} H(S_i) - H(S_i | X^n, Y^n, S^{i-1}) && (29)\\ &\ge \sum_{i=1}^{n} I(S_i; X_i, Y_i) && (30)\\ &\ge nI(S; \bar{X}, \bar{Y}), && (31)\end{aligned}$$

where $\bar{X}$ denotes a random variable whose distribution is the average of the marginal distributions for every $X_i$, $i = 1, \ldots, n$, and $\bar{Y}$ is the output corresponding to $\bar{X}$. Here, the last step follows because the distributions for $S_1, \ldots, S_n$ are identical, and by the convexity of mutual information in the conditional distribution of $(X,Y)$ given $S$. Combining (27) and (31) we obtain

$$I(S; \bar{X}, \bar{Y}) \le \epsilon_n + \epsilon_n'. \tag{32}$$

But by the standard converse proof procedure (see, e.g., [9]),

$$R \le I(\bar{X}; \bar{Y}) + \epsilon_n'', \tag{33}$$

where $\epsilon_n'' \downarrow 0$ as $n \to \infty$. So combining (32) and (33), noting that both $I(S; \bar{X}, \bar{Y})$ and $I(\bar{X}; \bar{Y})$ are continuous in $P_{\bar{X}}$ for fixed $P_S$ and $W(\cdot|\cdot,\cdot)$, and letting $n \to \infty$, we obtain that $C_{\text{det}}^{\text{IID}}$ is upper-bounded by the right-hand side of (21). This concludes the converse part of the proof. ■

### C. No CSI, Stochastic Encoder

Next we consider the case where the transmitter has no CSI, but is allowed to use a stochastic encoder, which is chosen randomly according to some distribution over all mappings of the form (20). The receiver knows the distribution according to which the mapping is chosen, but not the actual choice by the transmitter. The decoder is, as before, a mapping of the form (9). Capacity is defined as in Section II-A with the above-described encoder.

*Theorem 3:* When the transmitter has no CSI but can use a stochastic encoder, the capacity is given by

$$C^{\text{IID}}_{\text{sto}} = \sup I(U;Y), \tag{34}$$

where $U$ takes values in some finite set $\mathcal{U}$, and the supremum is taken over joint distributions of the form

$$P_S(s)P_U(u)P_{X|U}(x|u)W(y|x,s) \tag{35}$$

subject to

$$I(S;U,Y) = 0. \tag{36}$$

*Proof:* Achievability is proven by generating a codebook $\{u^n(1), \ldots, u^n(\lfloor 2^{nR} \rfloor)\}$ IID according to $P_U$, and picking each input symbol $x_i$ with probability $P_{X|U}(x_i|u_i(m))$, where $m$ is the message to be communicated. The analysis is similar to the previous cases and hence omitted.

To prove the converse part, we first use Fano's inequality to obtain, for some $\epsilon_n \downarrow 0$ as $n \to \infty$,

$$\begin{aligned} n(R-\epsilon_n) &\le I(M;Y^n) && (37)\\ &\le \sum_{i=1}^{n} I(M, Y^{i-1}; Y_i). && (38) \end{aligned}$$

We also have

$$\begin{aligned} I(S^n; M, Y^n) &= I(S^n; M|Y^n) + I(S^n;Y^n) && (39)\\ &\le H(M|Y^n) + I(S^n;Y^n) && (40)\\ &\le n\epsilon'_n, && (41) \end{aligned}$$

where $\epsilon'_n \downarrow 0$ as $n \to \infty$. The last step follows by Fano's inequality and the constraint (5). On the other hand,

$$\begin{aligned} I(S^n; M, Y^n) &= \sum_{i=1}^{n} I(S_i; M, Y^n, S^{i-1}) && (42)\\ &\ge \sum_{i=1}^{n} I(S_i; M, Y^{i-1}, Y_i). && (43) \end{aligned}$$

Let $U_i \triangleq (M, Y^{i-1})$, $i = 1, \ldots, n$. We have shown

$$\begin{aligned} \sum_{i=1}^{n} I(U_i;Y_i) &\ge n(R-\epsilon_n) && (44)\\ \sum_{i=1}^{n} I(S_i;U_i,Y_i) &\le n\epsilon'_n. && (45) \end{aligned}$$

Note that $U_i$ is independent of $S_i$ because $S^n$ is IID. Let $T$ be a time-sharing random variable that is uniformly distributed over $\{1, \ldots, n\}$, and denote $S \triangleq S_T$, $Y \triangleq Y_T$, and $U \triangleq (U_T, T)$, then

$$\begin{aligned} I(U;Y) &\ge R - \epsilon_n && (46)\\ I(S;U,Y) &\le \epsilon'_n. && (47) \end{aligned}$$

The proof is completed by letting $n \to \infty$ and exploiting continuity properties of the mutual information. ■

## D. Discussion and Examples

Bounds on the cardinality of the auxiliary alphabet $\mathcal{U}$ in Theorems 1 and 3 can be obtained using standard methods; see, e.g., [3]. For example, for Theorem 1, one can show that it suffices to have

$$|\mathcal{U}| \le \min\{|\mathcal{X}| \cdot |\mathcal{S}| + 1,\ |\mathcal{Y}| + |\mathcal{S}|\}, \tag{48}$$

while for Theorem 3 we only need

$$|\mathcal{U}| \le \min\{|\mathcal{X}|,\ |\mathcal{Y}|\} + 1. \tag{49}$$

Furthermore, in all of Theorems 1–3, the mutual information of interest is concave in the distributions of our choice, and the set of admissible distributions is convex, so the single-letter capacity expressions are all computable.

Next note that the three capacities that we analyzed clearly satisfy

$$C^{\text{IID}}_{\text{CSI}} \ge C^{\text{IID}}_{\text{sto}} \ge C^{\text{IID}}_{\text{det}}. \tag{50}$$

Their formulas can be interpreted more intuitively. We start with the smallest: the no-CSI deterministic-encoder capacity. The condition (23) in Theorem 2 is equivalent to the following: the input distribution must be restricted to those symbols that are not affected by the state $S$, i.e., input symbol $x \in \mathcal{X}$ can be used (with nonzero probability) only if

$$W(\cdot|x, s_1) = W(\cdot|x, s_2) \quad \text{for all } s_1, s_2 \in \mathcal{S}. \tag{51}$$

Conversely, any input distribution that only uses symbols satisfying (51) is permissible, in the sense that it satisfies condition (23).

Now consider the (middle) no-CSI stochastic-encoder capacity. Condition (36) is equivalent to requiring the codebook be restricted to symbols $u \in \mathcal{U}$ satisfying

$$W(\cdot|u, s_1) = W(\cdot|u, s_2) \quad \text{for all } s_1, s_2 \in \mathcal{S}. \tag{52}$$

Here, each $u$ corresponds to a distribution on $\mathcal{X}$. Thus, a stochastic encoder can use not only (deterministic) elements of $\mathcal{X}$, but also their mixtures.

For some channels, some input symbols do not satisfy (51), so they cannot be used by a deterministic encoder under the state-obfuscation constraint. But mixing these input symbols can result in a "super symbol" $u$ that satisfies (52), which can be used by a stochastic encoder. This is illustrated by the following example.

*Example 4:* Consider the channel depicted in Fig. 1, where $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$ and $\mathcal{S} = \{0, 1\}$. The channel law is, when $S = 0$, $Y = X$ with probability one; when $S = 1$, $Y = 0$ if $X = 0$, but the other two symbols are reversed: $Y = 2$ if $X = 1$ and $Y = 1$ if $X = 2$, all with probability one. The distribution of $S$ is arbitrary. A deterministic encoder can only use the input symbol 0, hence it cannot send any information:

$$C^{\text{IID}}_{\text{det}} = 0. \tag{53}$$

A stochastic encoder can choose $U \in \{0, 1\}$ uniformly, $X = 0$ with probability one if $U = 0$, and $X = 1$ or 2 equally likely if $U = 1$. This achieves one bit per channel use. It is straightforward to verify that this input strategy is optimal, i.e.,

$$C^{\text{IID}}_{\text{sto}} = 1 \text{ bit}. \tag{54}$$

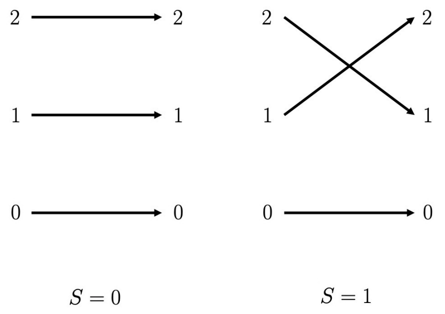

Fig. 1. The channel in Example 4. All transitions are with probability one.

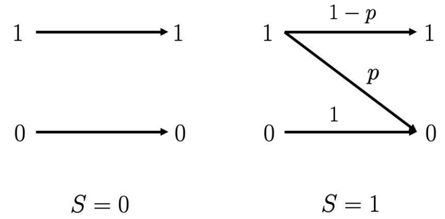

Fig. 2. The channel in Example 6.

For the case with CSI, where the capacity is still larger, the encoder again can only use $u$ if it satisfies (52), but now each $u$ is a mixture not of elements of $\mathcal{X}$, but of mappings from $\mathcal{S}$ to $\mathcal{X}$. We shall present an example where $C_{\text{CSI}}^{\text{IID}} > C_{\text{sto}}^{\text{IID}}$. Before doing so, we first make the following useful observation: the capacity in all three cases is upper-bounded by the worst-case capacity over $s \in \mathcal{S}$.

*Corollary 5:* In all settings above, capacity is upper-bounded by

$$\min_s \sup_{P_X} I(X;Y|S=s). \tag{55}$$

*Proof:* By (50), it suffices to prove the claim for the case with CSI. The condition (12) implies that

$$I(U;Y) = I(U;Y|S=s) \tag{56}$$

for every $s \in \mathcal{S}$. Hence,

$$\begin{aligned} C_{\text{CSI}}^{\text{IID}} &\le \sup_{P_U, P_{X|US}} \min_s I(U;Y|S=s) &(57)\\ &\le \min_s \sup_{P_U, P_{X|U}} I(U;Y|S=s) &(58)\\ &\le \min_s \sup_{P_X} I(X;Y|S=s), &(59)\end{aligned}$$

where the last step follows because $U - \circ - (X,S) - \circ - Y$ forms a Markov chain. ∎

*Example 6:* Consider a channel depicted in Fig. 2, where $\mathcal{X} = \mathcal{Y} = \mathcal{S} = \{0,1\}$. When $S=0$, the channel is a perfect bit pipe: $Y = X$ with probability one; when $S=1$, it is a Z-channel: $W(0|0,1)=1$ while $W(0|1,1) = p \in (0,1)$. (Again, the distribution of $S$ does not matter.) Corollary 5 implies that $C_{\text{CSI}}^{\text{IID}}$ cannot exceed the capacity of the Z-channel. We show that they are equal. Let $U$ be a binary random variable with the capacity-achieving input distribution of the Z-channel. Let $P_{X|US}$ be such that

$$P_{X|US}(1|0,s) = 0, \qquad s = 0,1 \tag{60a}$$

$$P_{X|US}(1|1,0) = 1-p \tag{60b}$$

$$P_{X|US}(1|1,1) = 1, \tag{60c}$$

namely, when $S=1$, we choose $X=U$ with probability one; when $S=0$, $X$ is produced by passing $U$ through the above Z-channel. By this choice, we have the same Z-channel from $U$ to $Y$ irrespectively of the value of $S$, hence $I(S;U,Y)=0$, whereas $I(U;Y)$ equals the capacity of the Z-channel.

On the other hand, one can show that $C_{\text{sto}}^{\text{IID}} = 0$ (which in turn implies $C_{\text{det}}^{\text{IID}} = 0$). To see this, observe that the auxiliary variable $u$ can be any mixture of $X=0$ and $X=1$, but the only such mixture that is not affected by $S$ is $X=0$ with probability one.

We shall return to this example at the end of the next Section.

Our last observation in this section is that, in all three studied cases, the state distribution does not affect capacity.

*Corollary 7:* Suppose $P_S$ and $Q_S$ are two probability mass functions on $\mathcal{S}$, both with support $\mathcal{S}$. When the state distribution of the channel is changed from $P_S$ to $Q_S$, all three capacities $C_{\text{det}}^{\text{IID}}$, $C_{\text{sto}}^{\text{IID}}$, and $C_{\text{CSI}}^{\text{IID}}$ remain unchanged.

*Proof:* Consider $C_{\text{det}}^{\text{IID}}$. First notice that the set of permissible input distributions remains unchanged when $P_S$ is replaced by $Q_S$. Indeed, an input distribution is permissible if and only if it only uses symbols that satisfy (51). Furthermore, for such an input distribution, $I(X;Y)$ is not affected by the state distribution. This is because $I(X;Y|S=s)$ is the same for all $s \in \mathcal{S}$. Recalling Theorem 2, we can then conclude that $C_{\text{det}}^{\text{IID}}$ indeed remains unchanged.

The proofs for the other two capacities are similar and therefore omitted. ∎

## III. Quasi-Static States

Consider again a state-dependent DMC with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, state alphabet $\mathcal{S}$, and channel law $W(y|x,s)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, $s \in \mathcal{S}$. In this section, we assume the state to be quasi-static instead of IID. This means the state is generated randomly according to a distribution $P_S$ before communication starts, and remains the same throughout the $n$ channel uses when a transmission takes place. We again assume (4) to hold.

As in the previous section, the message to be communicated is drawn from the set $\{1,\ldots,\lfloor 2^{nR}\rfloor\}$, and we consider three different settings for the encoder: with CSI, deterministic without CSI, and stochastic without CSI. The decoder is, as in Section II, a mapping from $y^n$ to a guess of the message.

For state obfuscation, we now require, for a uniformly drawn message,

$$\lim_{n\to\infty} I(S;Y^n) = 0. \tag{61}$$

(Note that the mutual information is *not* divided by $n$.) All our claims in this section will continue to hold under the stronger condition that, irrespectively of the message distribution,

$$I(S;Y^n) = 0 \quad \text{for all } n. \tag{62}$$

Indeed, we shall prove converse results under (61) and direct results under (62).

In all three settings, we define capacity as the supremum over all rates for which one can find a sequence of encoder-decoder pairs such that (61) is satisfied while the decoding error probability will approach zero when $n$ grows large. Note that this definition, together with (4), requires that the error probability be small conditional on $S = s$ for every $s \in \mathcal{S}$.

*Remark 8:* Since our channel model is not information stable [10], there is in general a tradeoff between outage probability and supportable rate, as in, e.g., [11]. Indeed, allowing a nonvanishing error probability can in some cases increase the maximum possible communication rate, as we see in the following simple example: When $S = 0$, the channel is a perfect bit pipe; when $S = 1$, $Y$ is a uniform binary random variable and is independent of $X$. By sending uncoded bits, we can send one bit per use of this channel, such that $I(S;Y^n) = 0$, and the probability of correct decoding is approximately $P_S(0)$. However, when we require the error probability to vanish as $n \to \infty$ (as in our analysis), capacity is clearly zero (even without a state obfuscation constraint).

### A. With CSI

When CSI is available to the transmitter,[3] the encoder is a possibly random mapping from $(s, m)$ to $x^n$, where $m$ denotes the message and $x^n$ the input sequence. The capacity in this case is the same for quasi-static and for IID states:

*Theorem 9:* For any DMC with transition law $W(\cdot|\cdot,\cdot)$ and state distribution $P_S$, the capacity when $S$ is quasi-static and when CSI is available to the transmitter is

$$C_{\text{CSI}}^{\text{static}} = C_{\text{CSI}}^{\text{IID}}, \tag{63}$$

where $C_{\text{CSI}}^{\text{IID}}$ is given by Theorem 1. Furthermore, $C_{\text{CSI}}^{\text{static}}$ does not depend on $P_S$ (as long as $\operatorname{supp}(P_S) = \mathcal{S}$).

*Proof:* The direct part of the proof is very similar to that for Theorem 1. Fix a joint distribution of the form (11). The codebook $\{u^n(1), \ldots, u^n(\lfloor 2^{nR} \rfloor)\}$ is generated in the same way as in the proof of Theorem 1. To send message $m$, the encoder picks input at time $i$ to be $x_i$ with probability $P_{X|US}(x_i|u_i(m), s)$, where $s$ is the state. To see that (62) is satisfied, we write, for every $m \in \{1, \ldots, \lfloor 2^{nR} \rfloor\}$, $s \in \mathcal{S}$, and $y^n \in \mathcal{Y}^n$,

$$\begin{aligned} &\Pr(S = s, Y^n = y^n | M = m) \\ &\quad = P_S(s) \prod_{i=1}^{n} P_{Y|US}(y|u_i(m), s) && (64) \\ &\quad = P_S(s) \prod_{i=1}^{n} P_{Y|U}(y|u_i(m)). && (65) \end{aligned}$$

This shows that $I(S;Y^n|M = m) = 0$, which in turn implies $I(S;Y^n) = 0$.

We next recall that (12) implies that $P_{Y|US}$ does not depend on $S$. Therefore, even though $S$ is quasi-static and not ergodic, the channel from $U$ to $Y$ is memoryless under such a joint distribution. Thus we can apply without change the error probability analysis used in the proof of Theorem 1, which completes the direct part.

It remains to prove the converse. To this end, we define auxiliary random variables

$$U_i \triangleq (M, Y^{i-1}), \quad i = 1, \ldots, n. \tag{66}$$

Using Fano's inequality and the chain rule, we have, for some $\epsilon_n \downarrow 0$ as $n \to \infty$,

$$\begin{aligned} n(R - \epsilon_n) &\leq I(M;Y^n) && (67) \\ &\leq \sum_{i=1}^{n} I(M, Y^{i-1}; Y_i) && (68) \\ &= \sum_{i=1}^{n} I(U_i; Y_i). && (69) \end{aligned}$$

We next show that $I(S;U_i,Y_i)$ must tend to zero as $n \to \infty$ for every $i$. Clearly,

$$I(S;U_i,Y_i) = I(S;M,Y^i) \leq I(S;M,Y^n), \tag{70}$$

so it suffices to prove that $I(S;M,Y^n)$ must tend to zero as $n$ grows large. To this end, define a binary random variable $F$ that equals 0 when decoding is correct and equals 1 when decoding is incorrect. Then we have

$$\begin{aligned} &I(S;M,Y^n) \\ &\quad = I(S;Y^n) + I(S;M|Y^n) && (71) \\ &\quad = I(S;Y^n) + I(S;M,F|Y^n) && (72) \\ &\quad = I(S;Y^n) + I(S;F|Y^n) + I(S;M|Y^n,F) && (73) \\ &\quad \leq I(S;Y^n) + H(F) + I(S;M|Y^n,F). && (74) \end{aligned}$$

The first two terms on the right-hand side of (74) both tend to zero as $n \to \infty$, the first by (61), and the second because the probability of a decoding error must tend to zero. For the last term, we have

$$\begin{aligned} &I(S;M|Y^n,F) \\ &= P_F(0) \sum_{y^n} \Pr(Y^n = y^n|F = 0) I(S;M|Y^n = y^n, F = 0) \\ &\quad + P_F(1) \sum_{y^n} \Pr(Y^n = y^n|F = 1) I(S;M|Y^n = y^n, F = 1) \\ & && (75) \\ &\leq P_F(0) \sum_{y^n} \Pr(Y^n = y^n|F = 0) \cdot 0 \\ &\quad + P_F(1) \sum_{y^n} \Pr(Y^n = y^n|F = 1) \cdot \log |\mathcal{S}| && (76) \\ &= P_F(1) \log |\mathcal{S}|, && (77) \end{aligned}$$

which also must tend to zero as $n \to \infty$ because $P_F(1)$ must tend to zero. Hence, as $n \to \infty$, the right-hand side of (74) must tend to zero, and consequently $I(S;U_i,Y_i)$ must tend to zero for every $i$. This, together with (69) and a continuity argument, completes the converse. ■

[3] Since the state remains constant during transmission, there is no difference between causal and noncausal CSI.

### *B. No CSI, Deterministic Encoder*

Assume now that the encoder must be a deterministic mapping from the message $m$ to an input sequence $x^n$ as in (20). The capacity is again the same as in the IID-state case.

*Theorem 10:* For any $W(\cdot|\cdot,\cdot)$ and $P_S$, the capacity when $S$ is quasi-static, and when the transmitter has no CSI and must use a deterministic encoder, is

$$C_{\text{det}}^{\text{static}} = C_{\text{det}}^{\text{IID}}, \tag{78}$$

where $C_{\text{det}}^{\text{IID}}$ is given in Theorem 2.

*Proof:* The direct part is essentially the same as before and omitted. For converse, we have, for every $i \in \{1,\ldots,n\}$,

$$\begin{aligned} I(S;X_i,Y_i) &\le I(S;X_i,Y^n) && (79)\\ &= I(S;Y^n) + I(S;X_i|Y^n) && (80)\\ &\le I(S;Y^n) + H(X_i|Y^n). && (81) \end{aligned}$$

Since the encoder is deterministic, the decoder should be able to correctly guess every $X_i$ from $Y^n$ (by first guessing $M$) with high probability. By Fano's inequality, $H(X_i|Y^n)$ must vanish together with the error probability. Hence, for every $i$,

$$\lim_{n\to\infty} I(S;X_i,Y_i) = 0. \tag{82}$$

Next consider the communication rate $R$. By Fano's inequality, for some vanishing $\epsilon_n$,

$$\begin{aligned} n(R-\epsilon_n) &\le I(X^n;Y^n) && (83)\\ &\le I(X^n,S;Y^n) && (84)\\ &\le \sum_{i=1}^{n} I\left(X^n,S,Y^{i-1};Y_i\right) && (85)\\ &= \sum_{i=1}^{n} I(X_i,S;Y_i) && (86)\\ &\le \sum_{i=1}^{n} I(X_i;Y_i) + I(S;X_i,Y_i). && (87) \end{aligned}$$

Combining (82) and (87) and letting $n \to \infty$ complete the converse. ■

### *C. No CSI, Stochastic Encoder: Examples*

When the transmitter has no CSI, a stochastic encoder is a random mapping from $m$ to $x^n$, as in Section II-C. The decoder knows the distribution used by the stochastic encoder, but not which codebook is chosen. Denote the capacity in this case subject to (61) by $C_{\text{sto}}^{\text{static}}$. We have not been able to develop a single-letter expression for $C_{\text{sto}}^{\text{static}}$. It is straightforward to verify that the direct part of Theorem 3 is still valid. We thus have

$$C_{\text{sto}}^{\text{IID}} \le C_{\text{sto}}^{\text{static}} \le C_{\text{CSI}}^{\text{static}} (= C_{\text{CSI}}^{\text{IID}}), \tag{88}$$

where the second inequality follows because additional information cannot reduce capacity. Both inequalities in (88) can be strict, as we show via the next two examples. In particular, unlike the previous two cases, here capacity need not be the same for quasi-static and for IID states.

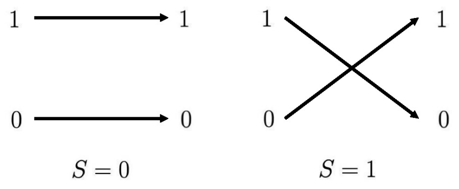

Fig. 3. The channel in Example 11.

*Example 11:* Let $\mathcal{X} = \mathcal{Y} = \mathcal{S} = \{0,1\}$. When $S = 0$ the channel is a noiseless bit pipe; when $S = 1$ the bit is flipped at the output with probability one; see Fig. 3. We have $C_{\text{sto}}^{\text{IID}} = 0$ because, without CSI and when the states are IID, it is impossible for the transmitter to send any information, even without the constraint (5). We have

$$C_{\text{sto}}^{\text{static}} = 1 \text{ bit}. \tag{89}$$

Indeed, consider the following simple scheme. The transmitter generates a random variable $A$ uniformly over $\{0,1\}$. To send $(n-1)$ information bits $(B_1,\ldots,B_{n-1})$ over $n$ channel uses, it sends

$$\begin{aligned} X_1 &= A && (90)\\ X_i &= B_{i-1} \oplus A, \quad i = 2,\ldots,n. && (91) \end{aligned}$$

The output string is then given by

$$\begin{aligned} Y_1 &= A \oplus S && (92)\\ Y_i &= B_{i-1} \oplus A \oplus S, \quad i = 2,\ldots,n. && (93) \end{aligned}$$

It is IID and uniform irrespectively of the value of $S$, so $I(S;Y^n) = 0$. The receiver can recover all information bits by computing

$$B_i = Y_{i+1} \oplus Y_1, \quad i = 1,\ldots,n-1. \tag{94}$$

*Example 12:* Consider the same channel as in Example 6 and Fig. 2, except that now the state remains the same for all $n$ channel uses. Recall that $C_{\text{CSI}}^{\text{IID}}$ equals the capacity of the Z-channel on the right-hand side of Fig. 2; by Theorem 9, so does $C_{\text{CSI}}^{\text{static}}$. We shall show that

$$C_{\text{sto}}^{\text{static}} = 0. \tag{95}$$

To this end, consider any sequence of encoder-decoder pairs, and define

$$\begin{aligned} A_n &\triangleq \sum_{i=1}^{n} X_i && (96)\\ B_n &\triangleq \sum_{i=1}^{n} Y_i. && (97) \end{aligned}$$

Further define

$$\alpha \triangleq P\text{-}\limsup_{n\to\infty} \frac{A_n}{n}, \tag{98}$$

where $P\text{-}\limsup$ denotes the limit-supremum in probability: $\alpha$ is the smallest real number for which the probability that $\frac{A_n}{n} > \alpha$ tends to zero as $n \to \infty$. Assume that $\alpha > 0$. Note that, when $S=0$, $B_n = A_n$ with probability one. Thus we have

$$\beta \triangleq \limsup_{n\to\infty} \Pr\left(\frac{B_n}{n} \ge \left(1-\frac{p}{2}\right)\alpha \,\middle|\, S=0\right) > 0. \tag{99}$$

When $S=1$, $B_n$ is conditionally a binomial distribution with parameters $A_n$ and $p$, therefore

$$P\text{-}\limsup_{n\to\infty} \left(\frac{B_n}{n}\,\middle|\, S=1\right) = (1-p)\alpha. \tag{100}$$

This implies

$$\lim_{n\to\infty} \Pr\left(\frac{B_n}{n} \ge \left(1-\frac{p}{2}\right)\alpha \,\middle|\, S=1\right) = 0. \tag{101}$$

Let $\delta(\cdot,\cdot)$ denote the total variation distance between two probability distributions. It follows from (99) and (101) that

$$\limsup_{n\to\infty} \delta\left(P_{B_n|S=0}, P_{B_n|S=1}\right) \ge \beta. \tag{102}$$

By Pinsker's inequality and the data processing inequality [9], this further implies that

$$\limsup_{n\to\infty} I(S;Y^n) \ge \limsup_{n\to\infty} I(S;B_n) \ge 2\beta^2. \tag{103}$$

Thus the assumption that $\alpha > 0$ is incompatible with the requirement (61); in other words, state obfuscation (61) requires that $\alpha = 0$. But, clearly, having $\alpha = 0$ does not permit communication at a positive rate. We have thus proven (95).

## IV. States With Memory

Consider the same state-dependent DMC as in the previous sections. We now turn to the situation where the state sequence $S^n$ is neither IID nor quasi-static. We make the following technical assumption: there exists some $a > 0$ such that, for all $i \in \mathbb{Z}^+$ and all $s^i \in \mathcal{S}^i$,

$$\Pr\left(S_i = s_i \middle| S^{i-1} = s^{i-1}\right) \ge a. \tag{104}$$

Aside from (104), we do not make any further assumptions on $S^n$, such as stationarity or ergodicity.

For state obfuscation, we impose the same requirement as in the IID-state case, i.e., (5). We consider the same three types of encoders as in the previous sections, namely, with CSI, deterministic without CSI, and stochastic without CSI. The decoder is also of the same form as in the previous sections. A rate is said to be achievable if there exists a sequence of encoder-decoder pairs such that (5) is satisfied while the decoding error probability approaches zero as $n \to \infty$.

By the same arguments as in Section III, we can easily see that the achievability schemes used in Section II continue to work when the states have memory, and that these schemes even satisfy the stronger requirement that $I(S^n;Y^n) = 0$ for all $n$. This means that, for every type of encoders, the capacity of a channel with states satisfying (104) is larger than or equal to the corresponding capacity when the states are IID. In the rest of this section we shall not repeat these achievability arguments.

In this section, we change the order to first study the cases where the transmitter has no CSI and uses deterministic and stochastic encoders, respectively. We then study the case where the transmitter has *causal* CSI. In all these three cases, we show that capacity for any channel and any state process satisfying (104) is the same as in the corresponding IID-state cases. A single-letter expression for the capacity when the transmitter has *noncausal* CSI remains an open problem.

### A. No CSI, Deterministic Encoder

In the no-CSI, deterministic-encoder case, capacity for any state process satisfying (104) is the same as $C_{\text{det}}^{\text{IID}}$. We note that (104) allows the marginal distributions for each $S_i$ to be different, therefore $I(S;X,Y)$ in (23) is not properly defined under (104). In the following theorem we replace (23) by its equivalent form, discussed in Section II-D.

*Theorem 13:* For any state sequence satisfying (104), the capacity when the transmitter has no CSI and must use a deterministic encoder, under the constraint (5), is equal to

$$C_{\text{det}}^{\text{mem}} = C_{\text{det}}^{\text{IID}} = \sup_{\text{supp}(P_X)\subseteq\tilde{\mathcal{X}}} I(X;Y), \tag{105}$$

where

$$\tilde{\mathcal{X}} \triangleq \{x \in \mathcal{X} \colon W(\cdot|x,s_1) = W(\cdot|x,s_2) \text{ for all } s_1, s_2 \in \mathcal{S}\}. \tag{106}$$

*Proof:* As discussed earlier, we shall not repeat the direct part. For the converse part, take any sequence of codes that has vanishing error probability and satisfies (5). Like (27), we have

$$I(S^n;X^n,Y^n) \le n(\epsilon_n + \epsilon_n'), \tag{107}$$

where both $\epsilon_n$ and $\epsilon_n'$ tend to zero as $n \to \infty$. Using the chain rule, we lower-bound the left-hand side of (107) as

$$\begin{aligned} I(S^n;X^n,Y^n) &= \sum_{i=1}^{n} I(S_i;X^n,Y^n|S^{i-1}) && (108)\\ &\ge \sum_{i=1}^{n} I(S_i;X_i,Y_i|S^{i-1}). && (109)\end{aligned}$$

To transform the above bound into one that involves $\sum_{i=1}^n I(S_i;X_i,Y_i)$ (without the conditioning), we observe the following. For any $i \in \{1,\ldots,n\}$ and $s^{i-1} \in \mathcal{S}^{i-1}$,

$$\begin{aligned} &I(S_i;X_i,Y_i)\\ &\quad= \sum_s P_{S_i}(s) D\left(P_{X_iY_i|S_i=s} \middle\| P_{X_iY_i}\right) && (110)\\ &\quad= \sum_s P_{S_i}(s) D\left(P_{X_iY_i|S_i=s} \middle\| P_{X_iY_i|S^{i-1}=s^{i-1}}\right)\\ &\qquad - D\left(P_{X_iY_i} \middle\| P_{X_iY_i|S^{i-1}=s^{i-1}}\right) && (111)\\ &\quad\le \sum_s P_{S_i}(s) D\left(P_{X_iY_i|S_i=s} \middle\| P_{X_iY_i|S^{i-1}=s^{i-1}}\right) && (112)\\ &\quad\le \sum_s \frac{1-a}{a} P_{S_i|S^{i-1}}(s|s^{i-1})\\ &\qquad \cdot D\left(P_{X_iY_i|S_i=s} \middle\| P_{X_iY_i|S^{i-1}=s^{i-1}}\right) && (113)\\ &\quad= \frac{1-a}{a} I(S_i;X_i,Y_i|S^{i-1}=s^{i-1}). && (114)\end{aligned}$$

Here, (113) follows by the assumption (104), and because (104) further implies $P_{S_i}(s) \leq 1-a$; and (114) follows because $S^{i-1} -\!\!\circ\!\!- S_i -\!\!\circ\!\!- (X_i, Y_i)$ forms a Markov chain. Averaging the above over $s^{i-1}$, summing it over $i$, and recalling (107) and (109), we obtain

$$\sum_{i=1}^{n} I(S_i; X_i, Y_i) \leq \sum_{i=1}^{n} \frac{1-a}{a} I(S_i; X_i, Y_i | S^{i-1}) \tag{115}$$

$$\leq \frac{1-a}{a} \cdot n(\epsilon_n + \epsilon_n'). \tag{116}$$

On the other hand, starting with a standard argument using Fano's inequality, we have, for some $\epsilon_n''$ that tends to zero as $n \to \infty$,

$$\begin{aligned} n(R - \epsilon_n'') &\leq I(X^n; Y^n) && (117)\\ &\leq I(X^n, S^n; Y^n) && (118)\\ &\leq \sum_{i=1}^{n} I(X^n, S^n, Y^{i-1}; Y_i) && (119)\\ &= \sum_{i=1}^{n} I(X_i, S_i; Y_i) && (120)\\ &\leq \sum_{i=1}^{n} I(X_i; Y_i) + \sum_{i=1}^{n} I(S_i; X_i, Y_i) && (121)\\ &\leq \sum_{i=1}^{n} I(X_i; Y_i) + \frac{1-a}{a} \cdot n(\epsilon_n + \epsilon_n'), && (122) \end{aligned}$$

where the last step follows from (116).

The proof is essentially completed by combining (116) and (122) and letting $n \to \infty$, but there are some additional technicalities. Since the marginals of $S_i$, $i = 1, \ldots, n$ are not specified and need not be the same, we cannot apply a convexity argument to the left-hand side of (116), as we did in the proof of Theorem 2. Instead, we make an argument via the total variation distance. We present this remaining part of the proof in the Appendix. ■

### B. No CSI, Stochastic Encoder

Like in the previous case, here capacity for any state process satisfying (104) is the same as in the IID-state case. Again, instead of (34), we express this capacity in an equivalent form, following the discussion in Section II-D.

*Theorem 14:* For any state sequence satisfying (104), the capacity when the transmitter has no CSI and may use a stochastic encoder, under constraint (5), is equal to

$$C_{\text{sto}}^{\text{mem}} = C_{\text{sto}}^{\text{IID}} = \sup I(U; Y), \tag{123}$$

where the supremum is taken over a choice of finite set $\mathcal{U}$, distribution $P_U$ on $\mathcal{U}$, and conditional distribution $P_{X|U}$, satisfying, for all $s_1, s_2 \in \mathcal{S}$, and for all $u \in \mathcal{U}$ such that $P_U(u) > 0$,

$$\sum_{x} P_{X|U}(x|u) W(\cdot|x, s_1) = \sum_{x} P_{X|U}(x|u) W(\cdot|x, s_2), \tag{124}$$

and the conditional distribution $P_{Y|U=u}$ is (124) computed for any $s \in \mathcal{S}$.

*Proof:* The direct part is omitted. To prove the converse part, let $T$ be a random variable uniformly distributed over $\{1, \ldots, n\}$, and define

$$\begin{aligned} U_i &\triangleq (M, S^{i-1}, Y^{i-1}), \quad i = 1, \ldots, n && (125)\\ U &\triangleq (U_T, T) && (126)\\ S &\triangleq S_T && (127)\\ Y &\triangleq Y_T. && (128) \end{aligned}$$

Note that $(M, Y^{i-1}) -\!\!\circ\!\!- S^{i-1} -\!\!\circ\!\!- S_i$ forms a Markov chain, hence (104) implies

$$P_{S|U}(s|u) \geq a \qquad \text{for all } s \in \mathcal{S}, u \in \mathcal{U}. \tag{129}$$

By Fano's inequality, we have, for some $\epsilon_n$ that tends to zero as $n \to \infty$,

$$\begin{aligned} &n(R - \epsilon_n)\\ &\quad \leq I(M; Y^n) && (130)\\ &\quad = \sum_{i=1}^{n} I(M; Y_i | Y^{i-1}) && (131)\\ &\quad \leq \sum_{i=1}^{n} I(M, S^i; Y_i | Y^{i-1}) && (132)\\ &\quad = \sum_{i=1}^{n} I(S^i; Y_i | Y^{i-1}) + \sum_{i=1}^{n} I(M; Y_i | Y^{i-1}, S^i) && (133)\\ &\quad \leq \sum_{i=1}^{n} I(S^i; Y_i | Y^{i-1}) + \sum_{i=1}^{n} I(U_i; Y_i | S_i). && (134) \end{aligned}$$

The first summation on the right-hand side of (134) can be bounded using the condition (5): for some $\epsilon_n'$ that tends to zero as $n \to \infty$,

$$\begin{aligned} n\epsilon_n' &\geq I(S^n; Y^n) && (135)\\ &= \sum_{i=1}^{n} I(S^n; Y_i | Y^{i-1}) && (136)\\ &\geq \sum_{i=1}^{n} I(S^i; Y_i | Y^{i-1}). && (137) \end{aligned}$$

Combining (134) and (137), we obtain

$$\begin{aligned} R - \epsilon_n - \epsilon_n' &\leq \frac{1}{n} \sum_{i=1}^{n} I(U_i; Y_i | S_i) && (138)\\ &= I(U_T; Y_T | S_T, T) && (139)\\ &\leq I(U_T, T; Y_T | S_T) && (140)\\ &= I(U; Y | S). && (141) \end{aligned}$$

On the other hand, starting with a bound like (41), we have, for some $\epsilon_n''$ that tends to zero as $n \to \infty$,

$$\begin{aligned} n\epsilon_n'' &\geq I(S^n; M, Y^n) && (142)\\ &= \sum_{i=1}^{n} I(S_i; M, Y^n | S^{i-1}) && (143)\\ &\geq \sum_{i=1}^{n} I(S_i; M, Y^i | S^{i-1}) && (144) \end{aligned}$$

$$
\begin{aligned}
&\geq \sum_{i=1}^{n} I(S_i; Y_i | U_i) &(145)\\
&= nI(S_T; Y_T | U_T, T) &(146)\\
&= nI(S; Y | U). &(147)
\end{aligned}
$$

Combining (141) and (147), recalling (129), and letting $n$ tend to infinity, we obtain that the state-obfuscation capacity is upper-bounded by

$$\sup I(U; Y | S) \tag{148}$$

over distributions of the form (note the Markov chain $S_i -\!\!\circ\!\!- (M, Y^{i-1}, S^{i-1}) -\!\!\circ\!\!- X_i$)

$$P_S(s) P_{U|S}(u|s) P_{X|U}(x|u) W(y|x,s) \tag{149}$$

subject to

$$
\begin{aligned}
&P_{S|U}(s|u) \geq a \qquad \text{for all } s, u &(150)\\
&I(S; Y | U) = 0. &(151)
\end{aligned}
$$

It remains to reduce (148)–(151) to the claimed expression. To this end, note that (150) and (151) together imply that, for every $u \in \mathcal{U}$ with $P_U(u) > 0$, and for every $s \in \mathcal{S}$,

$$P_{Y|US}(y|u,s) = P_{Y|U}(y|u). \tag{152}$$

Also note that, by (149),

$$P_{Y|US}(y|u,s) = \sum_x P_{X|U}(u) W(y|x,s). \tag{153}$$

Thus, (152) is a constraint on the choice of $\mathcal{U}$ and $P_{X|U}$; whether it is satisfied or not is not affected by the choice of $P_{U|S}$. We now have

$$
\begin{aligned}
&I(U; Y | S)\\
&\quad = \sum_s P_S(s) \sum_u P_{U|S}(u|s)\\
&\qquad \cdot D\left(P_{Y|US}(\cdot|u,s) \middle\| \sum_{u'} P_{U|S}(u'|s) P_{Y|US}(\cdot|u',s)\right) &(154)\\
&\quad = \sum_s P_S(s) \sum_u P_{U|S}(u|s)\\
&\qquad \cdot D\left(P_{Y|U}(\cdot|u) \middle\| \sum_{u'} P_{U|S}(u'|s) P_{Y|U}(\cdot|u')\right) &(155)\\
&\quad \leq \sum_s P_S(s) \sup_{P_{U|S}(\cdot|s)} I\big(P_{U|S}(\cdot|s), P_{Y|U}\big) &(156)\\
&\quad = \sup_{U \perp\!\!\!\perp S} I(P_U, P_{Y|U}), &(157)
\end{aligned}
$$

where in the last two lines we use $I(P, W)$ to denote the mutual information computed according to input distribution $P$ and transition law $W$. We thus obtain that (148) is upper-bounded by

$$\sup I(U; Y) \tag{158}$$

over distributions of the form

$$P_S(s) P_U(u) P_{X|U}(x|u) W(y|x,s) \tag{159}$$

with the condition

$$\sum_x P_{X|U}(u) W(y|x,s) = P_{Y|U}(y|u) \qquad \text{for all } u, y, s \text{ with } P_U(u) > 0, \tag{160}$$

which is equivalent to the claimed capacity expression. ■

*Remark 15:* The role played by the condition (104) and its consequence (129) in the proof is the following: they require that (152) must hold for *every* $s$. Without (129), there could be a pair $(u, s)$ with $P_{S|U}(s|u) = 0$, so it would be possible to have $P_{Y|US}(\cdot|u,s) \neq P_{Y|U}(\cdot|u)$, but $P_S(s) \neq 0$, then (155) need not hold.

### C. Causal CSI

When the transmitter has causal CSI, the encoder is a sequence of mappings: at time $i$, it maps the message $m$ and the states $s^i$ to the input symbol $x_i$. This mapping may be stochastic, i.e., the encoder may choose $x_i$ according to a certain distribution conditional on $m$ and $s^i$. Capacity is defined in the same way as before, and equals the capacity when the states are IID and with (causal or noncausal) CSI at the transmitter.

*Theorem 16:* For any state sequence satisfying (104), the capacity when the transmitter has causal CSI, under constraint (5), is equal to

$$C^{\mathrm{mem}}_{\text{C-CSI}} = C^{\mathrm{IID}}_{\mathrm{CSI}} = \sup I(U; Y), \tag{161}$$

where the supremum is taken over a choice of finite set $\mathcal{U}$, distribution $P_U$ on $\mathcal{U}$, and conditional distribution $P_{X|US}$, satisfying, for all $s_1, s_2 \in \mathcal{S}$, and for all $u \in \mathcal{U}$ such that $P_U(u) > 0$,

$$
\begin{aligned}
&\sum_x P_{X|US}(x|u,s_1) W(\cdot|x,s_1)\\
&\qquad = \sum_x P_{X|US}(x|u,s_2) W(\cdot|x,s_2), &(162)
\end{aligned}
$$

and the conditional distribution $P_{Y|U=u}$ in (161) is computed for any $s_1$ or $s_2$.

*Proof:* The proof is the same as the proof of Theorem 14, except that (149) is replaced by the joint distribution

$$P_S(s) P_{U|S}(u|s) P_{X|US}(x|u,s) W(y|x,s). \tag{163}$$

■

*Remark 17:* The reason why the above proof does not apply to noncausal CSI is that, with noncausal CSI, the Markov chain $(M, Y^{i-1}) -\!\!\circ\!\!- S^{i-1} -\!\!\circ\!\!- S_i$ no longer holds, so we can no longer establish (129).

## V. Concluding Remarks

We have presented information-theoretic capacity expressions for several instances of communication subject to state obfuscation. The case where the state is quasi-static and unknown to the transmitter, and where the transmitter can use a stochastic encoder, is yet unsolved. We have demonstrated via examples that the capacity in this case differs from both the IID-state no-CSI stochastic-encoder case and the quasi-static-state with-CSI case. Another unsolved case is where the states have memory, and where noncausal CSI is available at the encoder; in this case we do not know whether capacity differs from the corresponding IID-state case or not.

Our capacity results for IID states, specifically Theorems 2 and 3, can be extended to state masking, where the constraint (5) is replaced by (1); we do not elaborate within the present paper.

Partially due to the stringent state-obfuscation constraint (5) or (61), the state distribution has limited influence on the communication capacity. Indeed, in most cases we have studied, capacity depends on neither the marginal distribution of the state for a specific channel use, nor the dependence of states across channel uses. A notable exception is, as mentioned above, the quasi-static-state stochastic-encoder case.

Compared to IID or quasi-static channel states (the "compound channel" is an example of the latter), states with memory are less widely considered in the literature. The converse proofs in Section IV use some nonstandard techniques. For example, in (129), dependence between the current and previous states is reflected as dependence between the state and an auxiliary random variable, which plays a crucial part in the proof.

To analyze scenarios of practical interest where the transmitter wishes to guarantee a low probability of geolocation by the receiver, one may consider channel models with continuous alphabets. Our proof techniques do in general apply to continuous-alphabet channels, although the cardinality bounds do not, so additional work may be needed to obtain computable capacity formulas.

Some practically relevant channel and state models may be the following. In line-of-sight multiple-antenna wireless communication, the state $S$ may correspond to the phase difference between observation at receive antennas. For free-space optical communication, $S$ may correspond to attenuation of the transmitted signal. Examples 6 and 12 may be considered a first step towards modeling the latter channel. Analysis of these and other such scenarios is the subject of ongoing work.

## Appendix

In this appendix we complete the proof of Theorem 13. We continue from (116) and (122). For any $\hat{x} \in \mathcal{X} \setminus \tilde{\mathcal{X}}$, let $s_1, s_2 \in \mathcal{S}$ be such that

$$W(\cdot|\hat{x}, s_1) \neq W(\cdot|\hat{x}, s_2). \tag{164}$$

We have the following bound:

$$\begin{aligned}
&I(S_i; X_i, Y_i)\\
&\quad \geq P_{S_i}(s_1) D\left(P_{X_iY_i|S_i=s_1} \middle\| P_{X_iY_i}\right)\\
&\qquad + P_{S_i}(s_2) D\left(P_{X_iY_i|S_i=s_2} \middle\| P_{X_iY_i}\right) &(165)\\
&\quad \geq a \cdot D\left(P_{X_iY_i|S_i=s_1} \middle\| P_{X_iY_i}\right)\\
&\qquad + a \cdot D\left(P_{X_iY_i|S_i=s_2} \middle\| P_{X_iY_i}\right) &(166)\\
&\quad \geq 2a \cdot \delta^2(P_{X_iY_i|S_i=s_1}, P_{X_iY_i})\\
&\qquad + 2a \cdot \delta^2(P_{X_iY_i|S_i=s_2}, P_{X_iY_i}) &(167)\\
&\quad \geq a \cdot \left(\delta(P_{X_iY_i|S_i=s_1}, P_{X_iY_i}) + \delta(P_{X_iY_i|S_i=s_2}, P_{X_iY_i})\right)^2 &(168)\\
&\quad \geq a \cdot \delta(P_{X_iY_i|S_i=s_1}, P_{X_iY_i|S_i=s_2})^2 &(169)\\
&\quad \geq a \cdot (P_{X_i}(\hat{x}))^2 \cdot \delta^2\big(W(\cdot|\hat{x}, s_1), W(\cdot|\hat{x}, s_2)\big). &(170)
\end{aligned}$$

Here, (165) follows by dropping all $s \in \mathcal{S}$ except $s_1, s_2$; (166) by (104); (167) by Pinsker's inequality; (168) by simple algebra; (169) by the triangle inequality for total variation distance; and (170) by dropping all $x \in \mathcal{X}$ except $\hat{x}$. Combining (116) with (170), and letting $n \to \infty$, we see that the following must hold:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (P_{X_i}(\hat{x}))^2 = 0. \tag{171}$$

Define

$$B_i \triangleq \mathbf{1}\left\{X_i \notin \tilde{\mathcal{X}}\right\}, \quad i = 1, \ldots, n, \tag{172}$$

with $\mathbf{1}\{\cdot\}$ denoting the indicator function. Noting

$$\left(\frac{1}{n}\sum_{i=1}^{n} P_{X_i}(\hat{x})\right)^2 \leq \frac{1}{n}\sum_{i=1}^{n} (P_{X_i}(\hat{x}))^2, \tag{173}$$

and applying (171) to all $\hat{x} \in \mathcal{X} \setminus \hat{\mathcal{X}}$, we obtain

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathsf{E} B_i = 0, \tag{174}$$

We now turn back to (122). For every $i \in \{1, \ldots, n\}$,

$$\begin{aligned}
I(X_i; Y_i) &= I(X_i, B_i; Y_i) &(175)\\
&= I(B_i; Y_i) + P_{B_i}(1) I(X_i; Y_i|B_i = 1)\\
&\quad + P_{B_i}(0) I(X_i; Y_i|B_i = 0) &(176)\\
&\leq H(B_i) + P_{B_i}(1) \log|\mathcal{X}| + C_{\text{det}}^{\text{IID}}. &(177)
\end{aligned}$$

Plugging (174) and (177) into (122) and letting $n \to \infty$ prove that

$$R \leq C_{\text{det}}^{\text{IID}}. \tag{178}$$

This completes the proof.

## References

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

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

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

[4] N. Merhav and S. Shamai, "Information rates subject to state masking," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2254–2261, Jun. 2007.

[5] Y.-H. Kim, A. Sutivong, and T. M. Cover, "State amplification," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1850–1859, May 2008.

[6] O. O. Koyluoglu, R. Soundararajan, and S. Vishwanath, "State amplification under masking constraints," in *Proc. 49th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Monticello, IL, USA, Sep. 2011, pp. 936–943.

[7] T. A. Courtade, "Information masking and amplification: The source coding setting," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Jul. 2012, pp. 189–193.

[8] M. Dikshtein and S. Shamai, "Broadcasting information subject to state masking," 2018, *arXiv:1810.11781*.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Hoboken, NJ, USA: Wiley, 2006.

[10] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. San Francisco, CA, USA: Holden-Day, 1964.

[11] İ. E. Telatar, “Capacity of multi-antenna Gaussian channels,” *Euro. Trans. Telecommun.*, vol. 10, pp. 585–595, Dec. 1999.

**Ligong Wang** (Senior Member, IEEE) received the B.E. degree in electronic engineering from Tsinghua University, Beijing, China, in 2004, and the M.Sc. and Dr.Sc. degrees in electrical engineering from ETH Zurich, Switzerland, in 2006 and 2011, respectively.

Between 2011 and 2014, he was a Post-Doctoral Associate with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA. He joined the CNRS, France, as a Researcher (Chargé de Recherche) in 2014; he has been on leave of absence from CNRS since 2023. He is currently a Senior Researcher with the Signal and Information Processing Laboratory, ETH Zurich, and also a Lecturer with Lucerne University of Applied Sciences and Arts (HSLU), Switzerland. From 2019 to 2023, he served as an Associate Editor for Shannon Theory for IEEE TRANSACTIONS ON INFORMATION THEORY.

**Gregory W. Wornell** (Fellow, IEEE) received the B.A.Sc. degree from the University of British Columbia, Canada, and the S.M. and Ph.D. degrees from the Massachusetts Institute of Technology, all in electrical engineering and computer science, in 1985, 1987 and 1991, respectively.

Since 1991 he has been on the faculty at MIT, where he is the Sumitomo Professor of Engineering in the department of Electrical Engineering and Computer Science (EECS) and the Schwarzman College of Computing, and area co-chair of the EECS doctoral program. At MIT he leads the Signals, Information, and Algorithms Laboratory, and is affiliated with the Research Laboratory of Electronics (RLE), the Computer Science and Artificial Intelligence Laboratory (CSAIL), and the Institute for Data, Systems and Society (IDSS). He has held visiting appointments at the Department of Electrical Engineering and Computer Science at the University of California, Berkeley, CA, in 1999-2000, at Hewlett-Packard Laboratories, Palo Alto, CA, in 1999, and at AT&T Bell Laboratories, Murray Hill, NJ, in 1992-1993.

His research interests and publications span the areas of signal processing, information theory, statistical inference, artificial intelligence, and information security, and include architectures for sensing, learning, computing, communication, and storage; systems for computational imaging, vision, and perception; aspects of computational biology and neuroscience; and the design of wireless networks. He has been involved in the Information Theory and Signal Processing societies of the IEEE in a variety of capacities, and maintains a number of close industrial relationships and activities. He has won a number of awards for both his research and teaching, including the 2019 IEEE Leon K. Kirchmayer Graduate Teaching Award.