A Formula for the Capacity of the General Gel'fand-Pinsker Channel

Vincent Y. F. Tan

Institute for Infocomm Research (I^{2}R), A*STAR, Email: tanyfv@i2r.a-star.edu.sg
ECE Dept., National University of Singapore (NUS), Email: vtan@nus.edu.sg

***Abstract*—We consider the Gel'fand-Pinsker problem in which the channel and state are *general*, i.e., possibly non-stationary, non-memoryless and non-ergodic. Using Verdú-Han's information spectrum method and a non-trivial modification of Wyner's piggyback coding lemma, we prove that the capacity can be expressed as an optimization over the difference of a spectral inf- and a spectral sup-mutual information rate. We consider various specializations including the case where the channel and state are memoryless but non-stationary. We then extend our result to obtain the capacity region of the general Gel'fand-Pinsker problem with rate-limited state information at the decoder.**

***Index Terms*—Gel'fand-Pinsker, General channel, General sources, Information spectrum method**

I. Introduction

In this paper, we consider the classical problem of channel coding with noncausal state information at the encoder, also know as the *Gel'fand-Pinsker* problem [1]. In this problem, we would like to send a uniformly distributed message over a state-dependent channel $W^n : \mathcal{X}^n \times \mathcal{S}^n \to \mathcal{Y}^n$, where $\mathcal{S}, \mathcal{X}$ and $\mathcal{Y}$ are the state, input and output alphabets respectively. The random state sequence $S^n \sim P_{S^n}$ is available noncausally at the encoder but not at the decoder. The Gel'fand-Pinsker problem consists in finding the maximum rate for which there exists a reliable code. Assuming that the channel and state sequence are stationary and memoryless, Gel'fand and Pinsker [1] showed that this maximum message rate or *capacity* $C = C(W, P_S)$ is given by

$$C = \max_{\substack{P_{U|S},\, g:\mathcal{U}\times\mathcal{S}\to\mathcal{X} \\ |\mathcal{U}|\le|\mathcal{X}||\mathcal{S}|+1}} I(U;Y) - I(U;S). \tag{1}$$

The coding scheme involves a covering step at the encoder to reduce the uncertainty due to the random state sequence and a packing step to decode the message [2, Chapter 7]. Thus, we observe the covering rate $I(U;S)$ and the packing rate $I(U;Y)$ in (1). A weak converse can be proved by using the Csiszár-sum-identity [2, Chapter 7]. A strong converse was proved by Tyagi and Narayan [3].

In this paper, we revisit the Gel'fand-Pinsker problem and instead of assuming stationarity and memorylessness on the channel and state sequence, we let the channel W^n be a *general* one in the sense of Verdú and Han [4], [5]. That is, $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$ is an arbitrary sequence of stochastic mappings from $\mathcal{X}^n \times \mathcal{S}^n$ to $\mathcal{Y}^n$. We also model the state as a *general* one $\mathbf{S} \sim \{P_{S^n} \in \mathcal{P}(\mathcal{S}^n)\}_{n=1}^{\infty}$. We prove an analogue of the Gel'fand-Pinsker capacity in (1) by using information spectrum analysis [5]. Our result is expressed in terms of the limit superior and limit inferior in probability operations [5]. For the direct part, we leverage on a technique used by Iwata and Muramatsu [6] for the general Wyner-Ziv problem. Our proof technique involves a non-trivial modification of Wyner's piggyback coding lemma (PBL) [7, Lemma 4.3]. We also find the capacity region for the case where *rate-limited state information* is available at the decoder. The stationary, memoryless case was solved in by Steinberg in [8].

A. Main Contributions

There are two contributions in this paper. First, by deriving a non-asymptotic upper bound on the average error probability for any Gel'fand-Pinsker problem, we prove that the capacity can be expressed as a supremum over conditional probability laws $\{P_{X^n,U^n|S^n}\}_{n=1}^{\infty}$ of the difference between $\underline{I}(\mathbf{U};\mathbf{Y})$ and $\overline{I}(\mathbf{U};\mathbf{S})$. These are respectively the the spectral inf- and sup-mutual information rates [5]. We specialize the result to the following scenarios: (i) common state information is available at the encoder and the decoder and (ii) channel and state are memoryless (but non-stationary). We note that L. Wang also derived the capacity of the general Gel'fand-Pinsker channel [9, Appendix A.1] but the expression [9, Eq. (A.6)] does not specialize to (1) whereas our expression in (11) does.

Second, we extend the above result to the case where coded (rate-limited) state information is available at the decoder [8], [10]. In this case, we combine our coding scheme with that of Iwata and Muramatsu for the general Wyner-Ziv problem [6] to obtain the tradeoff between R_{d}, the rate of the compressed state information that is available at the decoder, and R be the message rate. We show that the tradeoff (or capacity region) is the set of rate pairs (R, R_{d}) satisfying $R_{\mathrm{d}} \ge \overline{I}(\mathbf{V};\mathbf{S}) - \underline{I}(\mathbf{V};\mathbf{Y})$ and $R \le \underline{I}(\mathbf{U};\mathbf{Y}|\mathbf{V}) - \overline{I}(\mathbf{U};\mathbf{S}|\mathbf{V})$ for some $(\mathbf{U},\mathbf{V}) - (\mathbf{X},\mathbf{S}) - \mathbf{Y}$. This general result can also be specialized the stationary, memoryless setting [8].

B. Related Work

The study of general channels started with the seminal work by Verdú and Han [4] in which the authors characterized the capacity in terms of the limit inferior in probability of a sequence of information densities. This line of analysis provides deep insights into the fundamental limits of the transmission of information over general channels and the compressibility of general sources that may not be stationary,

memoryless or ergodic. Information spectrum analysis has been used for rate-distortion [11], the Wyner-Ziv problem [6], the Wyner-Ahlswede-Körner (WAK) problem [12] and the wiretap channel [13], [14]. The Wyner-Ziv and wiretap problems are the most closely related to the problem we solve in this paper. In particular, they involve differences of mutual informations akin to the Gel'fand-Pinsker problem.

II. System Model and Main Definitions

In this section, we state our notation and the definitions of the two problems that we consider in this paper.

A. Notation

We follow the notational conventions in Han [5].

Definition 1. *Let* $\mathbf{U} := \{U_n\}_{n=1}^{\infty}$ *be a sequence of real-valued random variables. The* limsup in probability *of* $\mathbf{U}$ *is an extended real-number defined as*

$$\mathfrak{p}\text{-}\limsup_{n\to\infty} U_n := \inf\left\{\alpha : \lim_{n\to\infty} \mathbb{P}(U_n > \alpha) = 0\right\}. \tag{2}$$

The liminf in probability *of* $\mathbf{U}$ *is defined as*

$$\mathfrak{p}\text{-}\liminf_{n\to\infty} U_n := -\,\mathfrak{p}\text{-}\limsup_{n\to\infty} (-U_n). \tag{3}$$

Definition 2. *Given a pair of stochastic processes* $(\mathbf{X}, \mathbf{Y}) = \{X^n, Y^n\}_{n=1}^{\infty}$ *with joint distributions* $\{P_{X^n,Y^n}\}_{n=1}^{\infty}$, *the* spectral sup-mutual information rate *is defined as*

$$\overline{I}(\mathbf{X};\mathbf{Y}) := \mathfrak{p}\text{-}\limsup_{n\to\infty} \frac{1}{n} \log \frac{P_{Y^n|X^n}(Y^n|X^n)}{P_{Y^n}(Y^n)}. \tag{4}$$

The spectral inf-mutual information rate $\underline{I}(\mathbf{X};\mathbf{Y})$ *is defined as in* (4) *with* $\mathfrak{p}\text{-}\liminf$ *in place of* $\mathfrak{p}\text{-}\limsup$. *The* spectral sup- and inf-conditional mutual information rates *are defined similarly.*

B. The Gel'fand-Pinsker Problem

We recall the definition of the Gel'fand-Pinsker problem [1].

Definition 3. *An* (n, M_n, ϵ) code for the Gel'fand-Pinsker problem *with channel* $W^n : \mathcal{X}^n \times \mathcal{S}^n \to \mathcal{Y}^n$ *and state distribution* $P_{S^n} \in \mathcal{P}(\mathcal{S}^n)$ *consists of (i) an encoder* $f_n : [1 : M_n] \times \mathcal{S}^n \to \mathcal{X}^n$ *and (ii) a decoder* $\varphi_n : \mathcal{Y}^n \to [1 : M_n]$ *such that the average error probability in decoding the message does not exceed* ϵ, *i.e.,*

$$\frac{1}{M_n} \sum_{s^n \in \mathcal{S}^n} P_{S^n}(s^n) \sum_{m=1}^{M_n} W^n(\mathcal{B}_m^c | f_n(m, s^n), s^n) \le \epsilon, \tag{5}$$

where $\mathcal{B}_m := \{y^n \in \mathcal{Y}^n : \varphi_n(y^n) = m\}$ *and* $\mathcal{B}_m^c := \mathcal{Y}^n \setminus \mathcal{B}_m$.

We assume that the message is uniformly distributed in $[1 : M_n]$ and that it is independent of the state $S^n \sim P_{S^n}$.

Definition 4. *The number* R *is an* achievable rate *if there exists a sequence of* (n, M_n, ϵ_n) *codes for which*

$$\liminf_{n\to\infty} \frac{1}{n} \log M_n \ge R, \qquad \lim_{n\to\infty} \epsilon_n = 0. \tag{6}$$

The capacity C *is the supremum of all achievable rates.*

C. The Gel'fand-Pinsker Problem With Coded State Information at the Decoder

In fact the information spectrum method allow us to solve a related problem which was first considered by Heegard and El Gamal [10] and subsequently solved by Steinberg [8].

Definition 5. *An* $(n, M_n, M_{\mathrm{d},n}, \epsilon)$ code for the Gel'fand-Pinsker problem *with channel* $W^n : \mathcal{X}^n \times \mathcal{S}^n \to \mathcal{Y}^n$ *and state distribution* $P_{S^n} \in \mathcal{P}(\mathcal{S}^n)$ *and with* coded state information *at the decoder consists of (i) A state encoder:* $f_{\mathrm{d},n} : \mathcal{S}^n \to [1 : M_{\mathrm{d},n}]$, *(ii) an encoder:* $f_n : [1 : M_n] \times \mathcal{S}^n \to \mathcal{X}^n$ *and (iii) a decoder:* $\varphi_n : \mathcal{Y}^n \times [1 : M_{\mathrm{d},n}] \to [1 : M_n]$ *such that the average error probability in decoding the message is no larger than* ϵ, *i.e.,*

$$\frac{1}{M_n} \sum_{s^n \in \mathcal{S}^n} P_{S^n}(s^n) \sum_{m=1}^{M_n} W^n(\mathcal{B}_{m,s^n}^c | f_n(m, s^n), s^n) \le \epsilon \tag{7}$$

where $\mathcal{B}_{m,s^n} := \{y^n \in \mathcal{Y}^n : \varphi_n(y^n, f_{\mathrm{d},n}(s^n)) = m\}$.

Definition 6. *The pair of numbers* (R, R_{d}) *is* an achievable rate pair *if there exists a sequence of* $(n, M_n, M_{\mathrm{d},n}, \epsilon_n)$ *codes such that in addition to the conditions in* (6),

$$\limsup_{n\to\infty} \frac{1}{n} \log M_{\mathrm{d},n} \le R_{\mathrm{d}}. \tag{8}$$

The capacity region $\mathscr{C}$ *is the set of all achievable rate pairs.*

It was shown by Steinberg [8] for the discrete memoryless channel and discrete memoryless state that the capacity region $\mathscr{C}$ is the set of rate pairs (R, R_{d}) such that

$$R_{\mathrm{d}} \ge I(V;S) - I(V;Y) \tag{9}$$

$$R \le I(U;Y|V) - I(U;S|V) \tag{10}$$

for some Markov chain $(U, V) - (X, S) - Y$. The first constraint is obtained using Wyner-Ziv coding with "source" S and "side-information" Y. The second constraint is analogous to the Gel'fand-Pinsker capacity where V is common to both encoder and decoder. A weak converse was proven using repeated applications of the Csiszár-sum-identity.

III. Information Spectrum Characterization of the General Gel'fand-Pinsker problem

In this Section, we first present the main result concerning the capacity of the general Gel'fand-Pinsker problem in Section III-A. These results are derived using the information spectrum method. We then derive the capacity for various special cases of the Gel'fand-Pinsker problem in Section III-B (two-sided common state information) and Section III-C (memoryless channels and state). The main ideas in the proof are discussed in Section III-D. Finally, in Section III-E, we extend our result to the general Gel'fand-Pinsker problem with coded state information at the decoder.

A. Main Result and Remarks

We now state the capacity of the Gel'fand-Pinsker problem (Definition 4) followed by some remarks. Proofs are in [15].

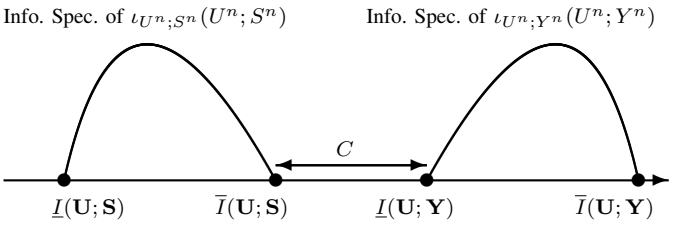

Fig. 1. Illustration of Theorem 1 where the information density $\iota_{U^n;S^n}(U^n;S^n) := n^{-1}\log[P_{U^n|S^n}(U^n|S^n)/P_{U^n}(U^n)]$ and similarly for $\iota_{U^n;Y^n}(U^n;Y^n)$. The capacity is the difference between $\underline{I}(\mathbf{U};\mathbf{Y})$ and $\overline{I}(\mathbf{U};\mathbf{S})$ evaluated at the optimal processes. The stationary, memoryless case (Corollary 4) corresponds to the situation in which $\underline{I}(\mathbf{U};\mathbf{S}) = \overline{I}(\mathbf{U};\mathbf{S}) = I(U;S)$ and $\underline{I}(\mathbf{U};\mathbf{Y}) = \overline{I}(\mathbf{U};\mathbf{Y}) = I(U;Y)$ so the information spectra are point masses at the mutual informations.

Theorem 1 (General Gel'fand-Pinsker Capacity)**.** *The capacity of the general Gel'fand-Pinsker channel with general states* $(\mathbf{W},\mathbf{S})$ *(see Definition 4) is*

$$C = \sup_{\mathbf{U}-(\mathbf{X},\mathbf{S})-\mathbf{Y}} \underline{I}(\mathbf{U};\mathbf{Y}) - \overline{I}(\mathbf{U};\mathbf{S}) \tag{11}$$

where the maximization is over all sequences of random variables $(\mathbf{U},\mathbf{X},\mathbf{S},\mathbf{Y}) = \{U^n,X^n,S^n,Y^n\}_{n=1}^{\infty}$ *forming the requisite Markov chain,*[1] *having the state distribution coinciding with* $\mathbf{S}$ *and having conditional distribution of* $\mathbf{Y}$ *given* $(\mathbf{X},\mathbf{S})$ *equal to the general channel* $\mathbf{W}$.

See Fig. 1 for an illustration of Theorem 1. This theorem is proved by using a non-asymptotic bound on the error probability given in Lemma 6 in the appendix.

Remark 1. The general formula in (11) is the dual of that in Wyner-Ziv [6]. However, the proofs, and in particular, the constructions of the codebooks, the notions of typicality and the application of Wyner's PBL, are subtly different from [6], [12]. We discuss these issues in Section III-D. Another problem which involves difference of mutual information quantities is the wiretap channel [2, Chapter 22]. General formulas for the secrecy capacity are provided in [13] and [14]. They also involve the difference between spectral inf-mutual information rate (of the input and the legitimate receiver) and sup-mutual information rate (of the input and the eavesdropper).

Remark 2. The general formula in (11) can be slightly generalized to the Cover-Chiang (CC) setting [16] in which (i) the channel $W^n : \mathcal{X}^n \times \mathcal{S}_{\mathrm{e}}^n \times \mathcal{S}_{\mathrm{d}}^n \to \mathcal{Y}^n$ depends on two state sequences $(S_{\mathrm{e}}^n, S_{\mathrm{d}}^n) \sim P_{S_{\mathrm{e}}^n,S_{\mathrm{d}}^n}$ (in addition to X^n), (ii) partial channel state information S_{e}^n is available noncausally at the encoder and (iii) partial channel state information S_{d}^n is available at the decoder. In this case, replacing $\mathbf{Y}$ with $(\mathbf{Y},\mathbf{S}_{\mathrm{d}})$ and $\mathbf{S}$ with $\mathbf{S}_{\mathrm{e}}$ in (11) yields

$$C_{\mathrm{CC}} = \sup_{\mathbf{U}-(\mathbf{X},\mathbf{S}_{\mathrm{e}},\mathbf{S}_{\mathrm{d}})-(\mathbf{Y},\mathbf{S}_{\mathrm{d}})} \underline{I}(\mathbf{U};\mathbf{Y},\mathbf{S}_{\mathrm{d}}) - \overline{I}(\mathbf{U};\mathbf{S}_{\mathrm{e}}), \tag{12}$$

where the maximum is over all sequences of random variables $(\mathbf{U},\mathbf{X},\mathbf{S}_{\mathrm{e}},\mathbf{S}_{\mathrm{d}},\mathbf{Y})$ where the distribution of $(\mathbf{S}_{\mathrm{e}},\mathbf{S}_{\mathrm{d}})$ coincides with $\{P_{S_{\mathrm{e}}^n,S_{\mathrm{d}}^n}\}_{n=1}^{\infty}$ and $\mathbf{Y}$ given $(\mathbf{X},\mathbf{S}_{\mathrm{e}},\mathbf{S}_{\mathrm{d}})$ coincides with the sequence of channels $\{W^n\}_{n=1}^{\infty}$. Hence the optimization in (12) is over the conditionals $\{P_{X^n,U^n|S_{\mathrm{e}}^n}\}_{n=1}^{\infty}$.

[1]For three processes $(\mathbf{X},\mathbf{Y},\mathbf{Z}) = \{X^n,Y^n,Z^n\}_{n=1}^{\infty}$, we say that $\mathbf{X}-\mathbf{Y}-\mathbf{Z}$ *forms a Markov chain* if $X^n - Y^n - Z^n$ for all $n \in \mathbb{N}$.

B. Two-Sided Common State Information

Specializing (12) to the case where $\mathbf{S}_{\mathrm{e}} = \mathbf{S}_{\mathrm{d}} = \mathbf{S}$, i.e., the *same* side information is available to both encoder and decoder (ED), is not straightforward without further assumptions. Recall that in the stationary, memoryless scenario [16, Case 4, Corollary 1], we use the identification $U = X$ in (1) and chain rule for mutual information to assert that $I(X;Y,S) - I(X;S) = I(X;Y|S)$ evaluated at $P_{X|S}^*$ is the capacity. However, the chain rule does not hold for the spectral sup-mutual information rate. In fact, $\mathfrak{p}\text{-}\liminf$ is superadditive [5]. Nevertheless, under the assumption that a sequence of information densities converges in probability, we can derive the capacity of the general channel with general common state available at both terminals using Theorem 1.

Corollary 2 (State at ED)**.** *Consider the problem*

$$C_{\mathrm{ED}} = \sup_{\mathbf{X}} \underline{I}(\mathbf{X};\mathbf{Y}|\mathbf{S}), \tag{13}$$

where the supremum is over all $(\mathbf{X},\mathbf{S},\mathbf{Y})$ *such that* $\mathbf{S}$ *coincides with the given states* $\{P_{S^n}\}_{n=1}^{\infty}$ *and* $\mathbf{Y}$ *given* $(\mathbf{X},\mathbf{S})$ *coincides with the given channels* $\{W^n\}_{n=1}^{\infty}$. *Assume that the maximizer of* (13) *exists and denote it by* $\{P_{X^n|S^n}^*\}_{n=1}^{\infty}$. *Let the distribution of* $\mathbf{X}^*$ *given* $\mathbf{S}$ *be* $\{P_{X^n|S^n}^*\}_{n=1}^{\infty}$. *If*

$$\underline{I}(\mathbf{X}^*;\mathbf{S}) = \overline{I}(\mathbf{X}^*;\mathbf{S}), \tag{14}$$

then the capacity of the state-dependent channel with state $\mathbf{S}$ *available at both encoder and decoder is* C_{ED} *in* (13).

If $(\mathbf{X}^*,\mathbf{S})$ satisfies (14) (and $\mathcal{X}$ and $\mathcal{S}$ are finite), it is called *information stable* [17]. We remark that a different achievability proof (that does not use Theorem 1) would allow us to dispense of the information stability assumption. We can simply develop a conditional version of Feinstein's lemma [5, Lemma 3.4.1] to prove the direct part of (13). The converse of Corollary 2 does not require (14). See [18].

C. Memoryless Channels and Memoryless States

To see how we can use Theorem 1 in concretely, we specialize it to the memoryless (but not necessarily stationary) setting. In the memoryless setting, the sequence of channels $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$ and the sequence of state distributions $P_{\mathbf{S}} = \{P_{S^n}\}_{n=1}^{\infty}$ are such that for every $(x^n,y^n,s^n) \in \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{S}^n$, we have $W^n(y^n|x^n,s^n) = \prod_{i=1}^n W_i(y_i|x_i,s_i)$, and $P_{S^n}(s^n) = \prod_{i=1}^n P_{S_i}(s_i)$ for some $\{W_i : \mathcal{X}\times\mathcal{S} \to \mathcal{Y}\}_{i=1}^{\infty}$ and some $\{P_{S_i} \in \mathcal{P}(\mathcal{S})\}_{i=1}^{\infty}$.

Corollary 3 (Memoryless Gel'fand-Pinsker Channel Capacity)**.** *Assume that* $\mathcal{X},\mathcal{Y}$ *and* $\mathcal{S}$ *are finite sets and the Gel'fand-Pinsker channel and state are memoryless. Define* $\phi(P_{X,U|S};W,P_S) := I(U;Y) - I(U;S)$. *Let the maximizers to the optimization problems indexed by* $i \in \mathbb{N}$

$$C(W_i,P_{S_i}) := \max_{\substack{P_{X,U|S} \\ |\mathcal{U}|\le|\mathcal{X}||\mathcal{S}|+1}} \phi(P_{X,U|S};W_i,P_{S_i}) \tag{15}$$

be denoted as $P^*_{X_i,U_i|S_i} : \mathcal{S} \to \mathcal{X} \times \mathcal{U}$. *Assume that either*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I(P_{S_i}, P^*_{U_i|S_i}), \quad or \quad \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I(P^*_{U_i}, P^*_{Y_i|U_i}) \tag{16}$$

exist, where $I(P, V)$ *denotes the mutual information calculated according to the joint distribution* $P \times V$. *Then the capacity of the memoryless Gel'fand-Pinsker channel is*

$$C_{\text{M'less}} = \liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} C(W_i, P_{S_i}). \tag{17}$$

The condition in (16) is only required for achievability. We illustrate (16) with Example 1 in the sequel. The proof of the direct part of Corollary 3 follows by taking the optimization in the general result (11) to be over memoryless conditional distributions. The converse follows by repeated applications of the Csiszár-sum-identity [2, Chapter 2]. If in addition to being memoryless, the channels and states are stationary, both limits in (16) exist since $P^*_{X_i,U_i|S_i}$ is the same for each $i \in \mathbb{N}$.

Corollary 4 (Stationary, Memoryless Gel'fand-Pinsker Channel Capacity)**.** *Assume that* $\mathcal{S}$ *is a finite set. In the stationary, memoryless case, the capacity of the Gel'fand-Pinsker channels given by* $C(W, P_S)$ *in* (1).

Only the converse of Corollary 4 requires the assumption that $|\mathcal{S}| < \infty$. The achievability of Corollary 4 follows easily from Khintchine's law of large numbers [5, Lemma 1.3.2].

Example 1. Let $\mathcal{J} := \{i \in \mathbb{N} : 2^{2k-1} \le i < 2^{2k}, k \in \mathbb{N}\}$ and let the set of even and odd positive integers be $\mathcal{E}$ and $\mathcal{O}$ respectively. Let $\mathcal{S}, \mathcal{X}, \mathcal{Y} = \{0, 1\}$. Consider a binary, nonstationary, memoryless channel $\mathbf{W}$ satisfying

$$W_i = \begin{cases} \tilde{W}_{\text{a}} & i \in \mathcal{O} \cap \mathcal{J} \\ \tilde{W}_{\text{b}} & i \in \mathcal{O} \cap \mathcal{J}^c \\ \tilde{W}_{\text{c}} & i \in \mathcal{E} \end{cases}, \tag{18}$$

where $\tilde{W}_{\text{a}}, \tilde{W}_{\text{b}}, \tilde{W}_{\text{c}} : \mathcal{X} \times \mathcal{S} \to \mathcal{Y}$. Also consider a binary, nonstationary, memoryless state $\mathbf{S} = \{S_i\}_{i=1}^{\infty}$ satisfying

$$P_{S_i} = \begin{cases} Q_{\text{a}} & i \in \mathcal{O} \\ Q_{\text{b}} & i \in \mathcal{E} \end{cases}, \tag{19}$$

where $Q_{\text{a}}, Q_{\text{b}} \in \mathcal{P}(\mathcal{S})$. In addition, assume that $\tilde{W}_m(\cdot|\cdot, s)$ for $(m, s) \in \{\text{a}, \text{b}\} \times \mathcal{S}$ are binary symmetric channels with arbitrary crossover probabilities $q_{ms} \in (0, 1)$. Let $V^*_{m,l} : \mathcal{S} \to \mathcal{U}$ be the $\mathcal{U}$-marginal of the maximizer in (15) when the channel is $\tilde{W}_m, m \in \{\text{a}, \text{b}\}$ and the state distribution is $Q_l, l \in \{\text{a}, \text{b}\}$. For $m \in \{\text{a}, \text{b}\}$ (odd blocklengths), due to the symmetry of the channels, the optimal $V^*_{m,\text{a}}(u|s)$ is Bernoulli($\frac{1}{2}$) and independent of s [2, Problem 7.12(c)]. Thus, for all odd blocklengths, the mutual informations in the first limit in (16) are equal to zero. Clearly, the first limit in (16) exists, equalling $\frac{1}{2} I(Q_{\text{b}}, V^*_{\text{c,b}})$ (contributed by the even blocklengths). Therefore, Corollary 3 applies and we can show that the Gel'fand-Pinsker capacity in (17) simplifies to

$$C_{\text{M'less}} = \frac{1}{2} \left[G(Q_{\text{a}}) + C(\tilde{W}_{\text{c}}, Q_{\text{b}}) \right] \tag{20}$$

where $C(W, P_S)$ in (15) is given explicitly in [2, Problem 7.12(c)] and $G : \mathcal{P}(\mathcal{S}) \to \mathbb{R}$ is defined as

$$G(Q) := \frac{2c_{\min}}{3} + \frac{c_{\max}}{3} \tag{21}$$

where $c_{\min} := \min\{C(\tilde{W}_{\text{a}}, Q), C(\tilde{W}_{\text{b}}, Q)\}$ and $c_{\max} := \max\{C(\tilde{W}_{\text{a}}, Q), C(\tilde{W}_{\text{b}}, Q)\}$. Equation (20) implies that the capacity consists of two parts: $C(\tilde{W}_{\text{c}}, Q_{\text{b}})$ represents the performance of the system $(\tilde{W}_{\text{c}}, Q_{\text{b}})$ at even n, while $G(Q_{\text{a}})$ represents the non-ergodic behavior of the channel at odd n with state distribution Q_{a}.

D. Proof Idea of Theorem 1

1) Direct part: The proof of the direct part is similar to that for usual Gel'fand-Pinsker coding [1] which involves covering to reduce the uncertainty due to the state and packing to decode the codeword. However, to use the information spectrum method for the general channel and general state, the definitions of "typicality" have to be restated in terms of information densities. We need to show that the probability that the transmitted codeword U^n is not "typical" with the channel output Y^n vanishes. In regular Gel'fand-Pinsker coding, one appeals to the conditional typicality lemma [1, Lemma 2] [2, Chapter 2] (which holds for "strongly typical sets") to assert that this error probability is small. But the "typical sets" used in information spectrum analysis do not allow us to apply the conditional typicality lemma in a straightforward manner. For example, our decoder is a threshold test involving the information density $n^{-1} \log(P_{Y^n|U^n}/P_{Y^n})$. It is not clear when there is no covering error that the transmitted U^n codeword passes the threshold test (i.e., $n^{-1} \log(P_{Y^n|U^n}/P_{Y^n})$ exceeds a certain threshold) with high probability.

To get around this problem, we modify Wyner's PBL [7, Lemma 4.3] accordingly. Wyner derived an analog of the Markov lemma [2, Chapter 12] by introducing a new "typical set" defined in terms of conditional probabilities. This definition is particularly useful for problems involving covering and packing and having some Markov structure. Our direct part proceeds roughly as follows: For a sequence $\{P_{X^n,U^n|S^n}\}_{n=1}^{\infty}$, define the following subsets of $\mathcal{U}^n \times \mathcal{Y}^n$ and $\mathcal{U}^n \times \mathcal{S}^n$:

$$\mathcal{T}_1 := \left\{ (u^n, y^n) : \frac{P_{Y^n|U^n}(y^n|u^n)}{P_{Y^n}(y^n)} \ge 2^{n(\underline{I}(\mathbf{U};\mathbf{Y}) - \gamma_1)} \right\}, \tag{22}$$

$$\mathcal{T}_2 := \left\{ (u^n, s^n) : \frac{P_{U^n|S^n}(u^n|s^n)}{P_{U^n}(u^n)} \le 2^{n(\overline{I}(\mathbf{U};\mathbf{S}) + \gamma_2)} \right\}, \tag{23}$$

where $\gamma_j > 0, j = 1, 2$, are arbitrarily small constants. Define

$$\eta(u^n, s^n) := \sum_{\substack{x^n \in \mathcal{X}^n, \\ (u^n, y^n) \in \mathcal{T}_1^c}} W^n(y^n|x^n, s^n) P_{X^n|U^n,S^n}(x^n|u^n, s^n). \tag{24}$$

Given $\eta(u^n, s^n)$ and analogous to [7, Lemma 4.3], define

$$\mathcal{A} := \left\{ (u^n, s^n) : \eta(u^n, s^n)^2 \le \mathbb{P}\left[(u^n, y^n) \in \mathcal{T}_1^c \right] \right\}, \tag{25}$$

and also the rate

$$R := \underline{I}(\mathbf{U}; \mathbf{Y}) - 2\gamma_1 - (\overline{I}(\mathbf{U}; \mathbf{S}) + 2\gamma_2). \tag{26}$$

Construct $\exp(nR)$ subcodebooks each indexing a message in $[1 : M_n]$ and containing $\exp(n(\overline{I}(\mathbf{U};\mathbf{S}) + 2\gamma_2))$ sequences drawn independently from P_{U^n}. The covering step at the encoder involves finding an auxiliary sequence u^n such that $(u^n, s^n) \in \mathcal{A}$ and the packing step at the decoder involves finding a u^n such that $(u^n, y^n) \in \mathcal{T}_2$. The resulting error probability is provided in (33) in Lemma 6.

Note that unlike in [6], we construct a random codebook and use it in subsequent steps rather than to assert the existence of a single codebook via random selection and subsequently regard it as being deterministic. This is because unlike Wyner-Ziv, for Gel'fand-Pinsker, we need to construct exponentially many subcodebooks each indexing a message in $[1 : M_n]$. We also require each of these subcodebooks to be *different* and *identifiable* based on the channel output. Our analogue of Wyner's PBL set $\mathcal{A}$ in (25) is different from that in [6] because the function η in (24) is different. In particular, it includes both the channel W^n and the conditional distribution $P_{X^n|U^n,S^n}$.

2) Converse part: Consider a sequence of (n, M_n, ϵ_n) codes with $\epsilon_n \to 0$. Let $U^n \in \mathcal{U}^n$ denote an arbitrary random variable representing the uniform choice of a message in $[1 : M_n]$. Because the message is independent of the state, this induces the joint distribution $P_{S^n} \circ P_{U^n} \circ P_{X^n|U^n,S^n} \circ W^n$ where $P_{X^n|U^n,S^n}$ models possible stochastic encoding. Clearly by the independence, $\overline{I}(\mathbf{U};\mathbf{S}) = 0$. Let the set of processes $(\mathbf{S}, \mathbf{U}, \mathbf{X}, \mathbf{Y})$ in which each collection of random variables (S^n, U^n, X^n, Y^n) is distributed as $P_{S^n} \circ P_{U^n} \circ P_{X^n|U^n,S^n} \circ W^n$ (resp. $P_{S^n} \circ P_{U^n|S^n} \circ P_{X^n|U^n,S^n} \circ W^n$) be $\mathcal{I}_{\mathbf{W},\mathbf{S}}$ for "independent" (resp. $\mathcal{D}_{\mathbf{W},\mathbf{S}}$ for "dependent"). For every $\gamma > 0$,

$$
\begin{aligned}
R &\le \liminf_{n\to\infty} \frac{1}{n} \log M_n \le \underline{I}(\mathbf{U};\mathbf{Y}) + \gamma && (27)\\
&= \underline{I}(\mathbf{U};\mathbf{Y}) - \overline{I}(\mathbf{U};\mathbf{S}) + \gamma && (28)\\
&\le \sup_{(\mathbf{S},\mathbf{U},\mathbf{X},\mathbf{Y}) \in \mathcal{I}_{\mathbf{W},\mathbf{S}}} \left\{ \underline{I}(\mathbf{U};\mathbf{Y}) - \overline{I}(\mathbf{U};\mathbf{S}) \right\} + \gamma, && (29)\\
&\le \sup_{(\mathbf{S},\mathbf{U},\mathbf{X},\mathbf{Y}) \in \mathcal{D}_{\mathbf{W},\mathbf{S}}} \left\{ \underline{I}(\mathbf{U};\mathbf{Y}) - \overline{I}(\mathbf{U};\mathbf{S}) \right\} + \gamma, && (30)
\end{aligned}
$$

where the second inequality in (27) follows from an application of the Verdú-Han converse [5, Lemma 3.2.2] and (30) follows from $\mathcal{I}_{\mathbf{W},\mathbf{S}} \subset \mathcal{D}_{\mathbf{W},\mathbf{S}}$. Now let $\gamma \to 0$.

E. Coded State Information at Decoder

We now state the capacity region of the coded state information problem (Definition 5).

Theorem 5 (Coded State Information at Decoder)**.** *The capacity region of the Gel'fand-Pinsker problem with coded state information at the decoder $\mathscr{C}$ (see Definition 6) is given by the set of pairs (R, R_{d}) satisfying*

$$
\begin{aligned}
R &\le \underline{I}(\mathbf{U};\mathbf{Y}|\mathbf{V}) - \overline{I}(\mathbf{U};\mathbf{S}|\mathbf{V}) && (31)\\
R_{\mathrm{d}} &\ge \overline{I}(\mathbf{V};\mathbf{S}) - \underline{I}(\mathbf{V};\mathbf{Y}) && (32)
\end{aligned}
$$

for $(\mathbf{U}, \mathbf{V}, \mathbf{X}, \mathbf{S}, \mathbf{Y}) = \{U^n, V^n, X^n, S^n, Y^n\}_{n=1}^{\infty}$ satisfying $(\mathbf{U}, \mathbf{V}) - (\mathbf{X}, \mathbf{S}) - \mathbf{Y}$, having the state distribution coinciding with $\mathbf{S}$ and having conditional distribution of $\mathbf{Y}$ given $(\mathbf{X}, \mathbf{S})$ equal to the general channel $\mathbf{W}$.

For the direct part, we combine Wyner-Ziv [6] and Gel'fand-Pinsker coding (Theorem 1). To prove the converse, we use exploit the independence of the message and the state, the Verdú-Han lemma [5, Lemma 3.2.2] and the proof technique for the converse of the general rate-distortion problem [5, Section 5.4]. Because the proof of Theorem 5 is very similar to Theorem 1, we only provide a sketch in [15]. Note that (9) and (10) follows as a corollary of Theorem 5.

APPENDIX

Lemma 6 (Nonasymptotic upper bound on error probability for Gel'fand-Pinsker)**.** *Fix a sequence of conditional distributions $\{P_{X^n,U^n|S^n}\}_{n=1}^{\infty}$. This specifies $\underline{I}(\mathbf{U};\mathbf{Y})$ and $\overline{I}(\mathbf{U};\mathbf{S})$. For every $n \in \mathbb{N}$, there exists an $(n, \exp(nR), \rho_n)$ code (Definition 3) for the general Gel'fand-Pinsker channel where R is defined in (26), and ρ_n is defined as follows:*

$$
\rho_n := 2\sqrt{\mathbb{P}(\mathcal{T}_1^c)} + \mathbb{P}(\mathcal{T}_2^c) + \exp[-\exp(n\gamma_2)] + \exp(-n\gamma_1). \tag{33}
$$

REFERENCES

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

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

[3] H. Tyagi and P. Narayan, "The Gelfand-Pinsker channel: Strong converse and upper bound for the reliability function," in *Proc. of IEEE Intl. Symp. on Info. Theory*, Seoul, Korea, 2009.

[4] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. on Inf. Th.*, vol. 40, no. 4, pp. 1147–57, Apr 1994.

[5] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer Berlin Heidelberg, Feb 2003.

[6] K.-I. Iwata and J. Muramatsu, "An information-spectrum approach to rate-distortion function with side information," *IEICE Trans. on Fundamentals of Electronics, Communications and Computer*, vol. E85-A, no. 6, pp. 1387–95, 2002.

[7] A. D. Wyner, "The wire-tap channel," *The Bell Systems Technical Journal*, vol. 54, pp. 1355–1387, 1975.

[8] Y. Steinberg, "Coding for channels with rate-limited side information at the decoder, with applications," *IEEE Trans. on Inf. Th.*, vol. 54, no. 9, pp. 4283–95, Sep 2008.

[9] L. Wang, "Information-theoretic aspects of optical communications," Ph.D. dissertation, ETH Zurich, 2011.

[10] C. Heegard and A. El Gamal, "On the capacity of computer memory with defects," *IEEE Trans. on Inf. Th.*, vol. 29, no. 5, pp. 731–739, May 1983.

[11] Y. Steinberg and S. Verdú, "Simulation of random processes and rate-distortion theory," *IEEE Trans. on Inf. Th.*, vol. 42, no. 1, pp. 63–86, Jan 1996.

[12] S. Miyake and F. Kanaya, "Coding theorems on correlated general sources," *IEICE Trans. on Fundamentals of Electronics, Communications and Computer*, vol. E78-A, no. 9, pp. 1063–70, 1995.

[13] M. Hayashi, "General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel," *IEEE Trans. on Inf. Th.*, vol. 52, no. 4, pp. 1562–1575, April 2006.

[14] M. Bloch and J. N. Laneman, "On the secrecy capacity of arbitrary wiretap channels," in *Allerton Conference on Communication, Control, and Computing*, 2008, pp. 818–25.

[15] V. Y. F. Tan, "A formula for the capacity of the general Gel'fand-Pinsker channel," *arXiv:1210.1091*, Oct 2012.

[16] T. M. Cover and M. Chiang, "Duality between channel capacity and rate distortion with two-sided state information," *IEEE Trans. on Inf. Th.*, vol. 48, no. 6, pp. 1629–38, Jun 2002.

[17] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. on Inf. Th.*, vol. 39, no. 3, pp. 752–72, Mar 1993.

[18] M. Tomamichel and V. Y. F. Tan, "ε-Capacity and strong converse for channels with general state," 2013, submitted to IEEE Inf. Th. Workshop.