

# Capacity-distortion trade-off in channels with state

Chiranjib Choudhuri, Young-Han Kim and Urbashi Mitra

**Abstract**—A problem of state information transmission over a state-dependent discrete memoryless channel (DMC) with independent and identically distributed (i.i.d.) states, known strictly causally at the transmitter is investigated. It is shown that block-Markov encoding coupled with channel state estimation conditioned on treating the decoded message and received channel output as side information at the decoder yields the minimum state estimation error. This same channel can also be used to send additional independent information at the expense of a higher channel state estimation error. The optimal tradeoff between the rate of the independent information that can be reliably transmitted and the state estimation error is characterized via the capacity-distortion function. It is shown that any optimal tradeoff pair can be achieved via a simple rate-splitting technique, whereby the transmitter appropriately allocates its rate between pure information transmission and state estimation.

## I. INTRODUCTION

In many communication scenarios, the communicating parties have some knowledge about the environment or the channel over which the communication takes place. For instance, the transmitter and the receiver may be able to monitor the interference level in the channel and only carry out communication when the interference level is low. In particular, we are interested in the study of data transmission over state-dependent channels. The case where state information is available at the transmitter has received considerable attention with prior work by Shannon [1], Kusnetsov and Tsybakov [2], Gel'fand and Pinsker [3], and Heegard and El Gamal [4]. Applications of this model include multimedia information hiding [5], digital watermarking [6], data storage over memory with defects [2], [4], secret communication systems [7], dynamic spectrum access systems [8], underwater acoustic/sonar applications [9] etc.

Most of the existing literature has focused on determining the channel capacity or devising practical capacity-achieving coding techniques [3], [4]. In certain communication scenarios, however, rather than communicating pure information across the channel, the transmitter may instead wish to help reveal the channel state to the receiver. An example of the above communication scenario is an analog-digital hybrid radio system

Chiranjib Choudhuri (cchoudhu@usc.edu) and Urbashi Mitra (ubli@usc.edu) are with the Ming Hsieh Department of Electrical Engineering, University of Southern California, University Park, Los Angeles, CA 90089, USA.

Young-Han Kim (yhk@ucsd.edu) is with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093, USA.

This research has been funded in part by the following grants and organizations: ONR N00014-09-1-0700, NSF CNS-0832186, NSF CNS-0821750 (MRI), and NSF CCF-0917343. The authors would also like to thank Sung Hoon Lim for enlightening discussions.

[10]. Here, digital refinement information is overlaid on the existing legacy analog transmission in order to help improve the detection and reconstruction of the original analog signal, which must be kept intact due to backward compatibility requirements. In this example, the existing analog transmission can be viewed as the channel state that the transmitter has access to and wishes to help reveal to the receiver. A key observation here is that the presence of the analog signal affects the channel over which the digital information is transmitted. At the same time, the digital transmission may itself interfere with the existing analog transmission, thereby degrading the quality of the original analog signal - the very thing that the digital information is designed to help improve.

In this paper, we study this problem of state information transmission over a state-dependent discrete memoryless channel. In this setup, the transmitter has access to the channel state in a strictly causal manner and wishes to help reveal it to the receiver with some fidelity criteria. We show that block-Markov encoding coupled with channel state estimation conditioned on treating the decoded message and received channel output as side information at the decoder is optimal for state information transmission.

This same channel can also be used to send additional independent information. This is, however, accomplished at the expense of a higher channel state estimation error. We characterize the tradeoff between the amount of independent information that can be reliably transmitted and the accuracy at which the receiver can estimate the channel state. There is a natural tension between sending pure information and revealing the channel state. Pure information transmission usually corrupts (or may even obliterate) the channel state, making it more difficult for the receiver to ascertain the channel state. Similarly, state information transmission takes away resources that may be used in transmitting pure information. We quantitatively characterize such a fundamental tension in this paper via the *capacity-distortion function* (first introduced in [11]). There is a fundamental difference between the capacity-distortion function and the rate-distortion function in lossy source coding [15]. The capacity-distortion function is defined with respect to a state-dependent channel, seeking to characterize the fundamental tradeoff between the rate of information transmission and the distortion of state estimation. In contrast, the rate-distortion function is defined with respect to a source distribution, seeking to characterize the fundamental tradeoff between the rate of its lossy description and the achievable distortion due to the description. We show that any optimal tradeoff pair can be achieved via a simple rate-splitting technique, whereby the transmitter is appropriately allocated its rate between pure information transmission and state estimation.

The problem formulation in [11] bears similarity to that

we consider: the destination is interested in both information transmission and channel state estimation. However, a critical distinction that differentiates our work from [11], is the fact that in [11], the transmitter and the receiver are assumed to be oblivious of the channel state realization. In our formulation, the source has a strictly causal knowledge of the channel state. In fact, we show that the results of [11] are a special case of our results. Similarly [12], [13] consider the rate-distortion trade-off for the state-dependent additive Gaussian channel, where the channel state is assumed to be *non-causally* known at the source. Along the same lines, [14] considered the problem of transmitting data over a state-dependent channel with state information available both non-causally and causally at the sender and at the same time conveying the information about the channel state itself to the receiver. The optimal tradeoff is characterized between the information transmission rate and the state uncertainty reduction rate, in which the decoder forms a list of possible state-sequences. There is a fundamental difference between the state uncertainty reduction rate and distortion, as under some distortion measure, it may so happen that a state sequence not in the list of decoder may yield a lower distortion.

The rest of this paper is organized as follows. Section II defines the notation used in the paper. Section III describes the basic channel model with discrete alphabets, and formulates the problem of characterizing the minimum achievable distortion at zero information rate. Section IV determines the minimum distortion, and Section V establishes its achievability. Section VI proves the converse part of the theorem. Section VII extends the results to the information rate-distortion trade-off setting, wherein we define and evaluate the capacity-distortion function. Section VIII illustrates the application of the capacity-distortion function through the examples of an additive state dependent Gaussian channel and an additive state dependent binary channel. Finally, Section IX concludes the paper.

## II. NOTATION

Before formulating the problem, we define notation that will be used throughout the paper. Capital letters are used to denote random variables and small letters are reserved for a particular realization of the random variable.  $x_a^b$  denotes a sequence  $(x_a, x_{a+1}, \dots, x_b)$ , whose elements are drawn from the same distribution. When  $a = 1$ , we usually omit the subscript. For  $X \sim P(x)$  and  $\epsilon \in (0, 1)$ , we define the set  $\mathcal{T}_\epsilon^{(n)}(X)$  of typical sequences  $x^n$  as

$$\mathcal{T}_\epsilon^{(n)}(X) := \left\{ x^n : \left| \frac{|\{j : x_j = x\}|}{n} - P(x) \right| \leq \epsilon P(x), \right. \\ \left. \text{for all } x \in \mathcal{X} \right\}. \quad (1)$$

Jointly typical sequences can be defined similarly. Consider the random variables  $A, B, C$ . If  $A$  and  $C$  are conditionally independent given  $B$ , we say they form a Markov chain. We denote this statistical relationship by  $A - B - C$ .

## III. BASIC PROBLEM FORMULATION

In this section, we formulate the channel state estimation problem, where the receiver only wants to estimate the channel

state with minimum distortion, with the channel state is available strictly causally at the transmitter.

*Channel input:* A symbol  $x$  taken from a finite input alphabet  $\mathcal{X} = \{a^{(1)}, a^{(2)}, \dots, a^{(|\mathcal{X}|)}\}$ .

*Channel output:* A symbol  $y$  taken from a finite output alphabet  $\mathcal{Y} = \{b^{(1)}, b^{(2)}, \dots, b^{(|\mathcal{Y}|)}\}$ .

*Channel state:* A symbol  $s$  taken from a finite state alphabet  $\mathcal{S} = \{c^{(1)}, c^{(2)}, \dots, c^{(|\mathcal{S}|)}\}$ . For each channel use, the state is a random variable  $S$  which has a probability mass function (PMF)  $P_S(s)$ . Over any  $n$  consecutive channel uses, the channel state sequence  $S^n$  is memoryless,  $P(S^n) = \prod_{j=1}^n P_S(s_j)$ .

*Channel:* A collection of probability transition matrices each of which specifies the conditional probability distribution under a fixed channel state; that is,  $P(b^{(j)}|a^{(i)}, c^{(k)})$  represents the probability of output  $y = b^{(j)} \in \mathcal{Y}$  occurring given input  $x = a^{(i)} \in \mathcal{X}$  and state  $s = c^{(k)} \in \mathcal{S}$ , for any  $1 \leq i \leq |\mathcal{X}|, 1 \leq j \leq |\mathcal{Y}|$  and  $1 \leq k \leq |\mathcal{S}|$ . With  $n$  consecutive channel uses, the channel transitions are mutually independent, characterized by

$$P(y^n|x^n, s^n) = \prod_{j=1}^n P(y_j|x_j, s_j). \quad (2)$$

*Distortion:* For any two channel states, the distortion is a deterministic function,  $d : \mathcal{S} \times \hat{\mathcal{S}} \mapsto \mathbb{R}^+ \cup \{0\}$ . It is further assumed that  $d(\cdot, \cdot)$  is bounded, i.e.,  $d(c^{(i)}, c^{(j)}) \leq d_{max} < \infty$  for any  $1 \leq i, j \leq |\mathcal{S}|$ . For any two length- $n$  state sequences  $(s_1, \dots, s_n) \in \mathcal{S}^n, (\hat{s}_1, \dots, \hat{s}_n) \in \hat{\mathcal{S}}^n$ , the distortion is defined to be the average of the pairwise distortions,  $\frac{1}{n} \sum_{j=1}^n d(s_j, \hat{s}_j)$ .

*Coding:* A  $(f_j, h_j), 1 \leq j \leq n$  code for the channel is defined as:

- *Encoder:* A deterministic function  $f_j : \mathcal{S}^{j-1} \mapsto \mathcal{X}$  for each  $1 \leq j \leq n$ . Note that the state sequence is available at the encoder in a strictly causal manner.
- *State Estimator:* A deterministic function,  $h_j : \mathcal{Y}^n \mapsto \hat{\mathcal{S}}_j$ . We denote  $\hat{S}_j = h_j(Y^n)$  as the estimated channel states.

*Distortion for channel state estimation:* We consider the average distortion, which is defined as

$$\bar{D}^{(n)} = E \left[ \frac{1}{n} \sum_{j=1}^n d(S_j, \hat{S}_j) \right], \quad (3)$$

where the expectation is over the conditional joint distribution of  $(S^n, Y^n)$ , noting that  $\hat{S}^n$  is determined by  $Y^n$ .

In this paper, we wish to characterize  $D_{min}$  defined as

$$D_{min} = \liminf_{n \rightarrow \infty} \min_{f_j, h_j, 1 \leq j \leq n} E \left[ \frac{1}{n} \sum_{j=1}^n d(S_j, \hat{S}_j(Y^n)) \right], \quad (4)$$

which is the minimum distortion achievable for the channel model.

## IV. MAIN RESULT

To characterize the minimum distortion, we will need the following definition.

*Definition 1:* For a joint distribution  $P_{SUXY}$ , define the minimum possible estimation error of  $S$  given  $(U, X, Y)$  by

$$\xi(S|U, X, Y) = \min_{g: \mathcal{U} \times \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{S}} E[d(S, g(U, X, Y))], \quad (5)$$

where  $d(\cdot, \cdot)$  is a distortion measure.

Since by definition,  $X_j$  is independent of  $S_j$ , for all  $1 \leq j \leq n$ , one possible coding strategy (see [11]) would be to send a deterministic  $X_j$  that facilitates state estimation at the decoder and use a function of  $(X_j, Y_j)$  to estimate  $S_j$ . In that case, the minimum distortion  $D_{min}$  is given by,

$$D_{min} = \min_{x \in \mathcal{X}, g(\cdot)} E[d(S, g(x, Y))]. \quad (6)$$

This strategy, although optimal for the case of channel-state oblivious receiver (RX) and transmitter (TX) as shown in [11], is found to be suboptimal when we have some channel state information (CSI) at the TX as we can perform better by applying a block-Markov coding scheme, where at block  $i$ , the transmitter can use its knowledge of the state sequence of block  $i-1$  to select a code that can be employed to estimate the state sequence of block  $i-1$  at the decoder after receiving the channel output of block  $i$ . We could implement this by compressing the channel state of block  $i-1$  and then sending the compression index through the channel. This strategy can further be improved as the compression index can be sent at a much lower rate by realizing that the receiver has a side information of  $(X^n(i-1), Y^n(i-1))$ . We implement this coding scheme to achieve the minimum distortion, which is given by the following theorem.

*Theorem 1:* The minimum achievable distortion for the problem considered in last section is

$$d_{min} = \min_{\mathcal{P}} \xi(S|U, X, Y), \quad (7)$$

where

$$\mathcal{P} = \{P_X, P_{U|X, S} : I(U, X; Y) - I(U, X; S) \geq 0\}. \quad (8)$$

and  $U$  is an auxiliary random variable of finite alphabet size.

*Remark:* These results hold for all possible finite delays, the key is the need for strict causality. In fact, our results hold as long as the delay is sub-linear in the codeword length  $n$ .

## V. PROOF OF ACHIEVABILITY

In this section, we will prove achievability for Theorem 1.

We fix the distributions  $P_X, P_{U|X, S}$  and  $\hat{s}(u, x, y)$  that achieve a distortion of  $D_{min}/(1 + \epsilon)$ .

*Codebook generation:*

- Choose  $2^{n\hat{R}}$  i.i.d  $x^n$  each with probability  $P(x^n) = \prod_{j=1}^n P(x_j)$ . Label these as  $x^n(w)$ ,  $w \in [1 : 2^{n\hat{R}}]$ .
- Choose, for each  $x^n(w)$ ,  $2^{nR'}$  i.i.d  $u^n$  each with probability  $P(u^n|x^n(w)) = \prod_{j=1}^n P(u_j|x_j(w))$ , where for  $x \in \mathcal{X}, u \in \mathcal{U}$ , we define

$$P(u|x) = \sum_{s \in \mathcal{S}} P(s)p(u|x, s).$$

Label these as  $u^n(l|w), l \in [1 : 2^{nR'}], w \in [1 : 2^{n\hat{R}}]$ .

- Partition the set of indices  $l \in [1 : 2^{nR'}]$  into equal-size subsets  $\mathcal{B}(w) := [ (m-1)2^{n(R'-\hat{R})} + 1 : m2^{n(R'-\hat{R})} ], w \in [1 : 2^{n\hat{R}}]$ .
- The codebook is revealed to the both encoder and decoder.

*Encoding:* Let,  $x^n(w_{i-1}) \in \mathcal{T}_\epsilon^{(n)}(X)$  be the codeword sent in block  $i-1$ .

- Knowing  $s^n(i-1)$  at the beginning of block  $i$ , the encoder looks for an index  $l_i \in [1 : 2^{nR'}]$  such that  $(u^n(l_i|w_{i-1}), s^n(i-1), x^n(w_{i-1})) \in \mathcal{T}_\epsilon^{(n)}(S, X, U)$ . If there is more than one such  $l_i$ , the smallest index is selected. If there is no such  $l_i$ , select  $l_i = 1$ .
- Determine the  $w_i$  such that  $l_i \in \mathcal{B}(w_i)$ . Codeword  $x^n(w_i)$  is transmitted in block  $i$ .

*Analysis of probability of error for encoding:* We define the following error events:

$$\begin{aligned} \mathcal{E}_{11i} &:= \{S^n(i-1) \notin \mathcal{T}_{\epsilon_1}^{(n)}(S)\}, \\ \mathcal{E}_{12i} &:= \{(U^n(l|w_{i-1}), S^n(i-1), X^n(w_{i-1})) \notin \mathcal{T}_{\epsilon_2}^{(n)}(S, X, U) \text{ for all } l \in [1 : 2^{nR'}]\}. \end{aligned}$$

The total probability of error for the encoding step is then upper bounded as

$$P(\mathcal{E}_{1i}) \leq P(\mathcal{E}_{11i}) + P(\mathcal{E}_{11i}^c \cap \mathcal{E}_{12i}). \quad (9)$$

We now bound each term:

- $P(\mathcal{E}_{11i})$  goes to 0 as  $n \rightarrow \infty$  by the Law of Large Numbers (LLN).
- As given  $X^n(w_{i-1}), U^n(l|w_{i-1}), l \in [1 : 2^{nR'}]$  is generated independently of  $S^n(i-1)$ , by the Covering Lemma (see [16]),  $P(\mathcal{E}_{11i}^c \cap \mathcal{E}_{12i}) \rightarrow 0$  if  $\epsilon_2$  is small,  $n$  is large and

$$R' > I(U; S|X) \stackrel{(a)}{=} I(U, X; S), \quad (10)$$

where (a) follows from the independence of  $X$  and  $S$ .

- Combining the results, we can conclude that  $P(\mathcal{E}_{1i}) \rightarrow 0$  as  $n \rightarrow \infty$  if  $R' > I(U, X; S)$ .

*Decoding and analysis of probability of error:* At the end of the block  $i$  the decoder does the following:

- The receiver declares  $\hat{w}_i$  was sent by looking for the uniquely typical  $x^n(w_i)$  with  $y^n(i)$ . Without loss of generality, let  $W_i = 1$  be the chosen index in block  $i$ . The decoder makes an error iff

$$\begin{aligned} \mathcal{E}_{21i} &:= \{X^n(1) \notin \mathcal{T}_{\epsilon_3}^{(n)}(X)\}, \\ \mathcal{E}_{22i} &:= \{S^n(i) \notin \mathcal{T}_{\epsilon_4}^{(n)}(S)\}, \\ \mathcal{E}_{23i} &:= \{(X^n(1), Y^n(i)) \notin \mathcal{T}_{\epsilon_5}^{(n)}(X, Y)\}, \\ \mathcal{E}_{24i} &:= \{(X^n(w), Y^n(i)) \in \mathcal{T}_{\epsilon_5}^{(n)}(X, Y) \text{ for some } w \neq 1\}. \end{aligned}$$

Thus, by the union of events bound, we have

$$\begin{aligned} P(\mathcal{E}_{2i}) &= P(\mathcal{E}_{2i}|W_i = 1) \leq P(\mathcal{E}_{21i}) + P(\mathcal{E}_{22i}) \\ &\quad + P(\mathcal{E}_{21i}^c \cap \mathcal{E}_{22i}^c \cap \mathcal{E}_{23i}) + P(\mathcal{E}_{21i}^c \cap \mathcal{E}_{22i}^c \cap \mathcal{E}_{24i}). \end{aligned} \quad (11)$$

We next bound each term:

- $P(\mathcal{E}_{21i})$  and  $P(\mathcal{E}_{22i})$  go to 0 as  $n \rightarrow \infty$  by LLN.
- Since  $\mathcal{E}_{21i}^c := \left\{ X^n(1) \in \mathcal{T}_{\epsilon_3}^{(n)}(X) \right\}$  and  $\mathcal{E}_{22i}^c := \left\{ S^n(i) \in \mathcal{T}_{\epsilon_4}^{(n)}(S) \right\}$ , it implies that  $(X^n(1), S^n(i)) \in \mathcal{T}_{\epsilon_5}^{(n)}(S, X)$  (as they are independent of each other) and thus by the Conditional Typicality Lemma, (see [16])  $P(\mathcal{E}_{21i}^c \cap \mathcal{E}_{22i}^c \cap \mathcal{E}_{23i}) \rightarrow 0$  due to the fact that  $(X^n(1), S^n(i), Y^n(i)) \in \mathcal{T}_{\epsilon_5}^{(n)}(S, X, Y)$ .
- By the Packing Lemma (see [16])  $P(\mathcal{E}_{21i}^c \cap \mathcal{E}_{22i}^c \cap \mathcal{E}_{24i}) \rightarrow 0$  if  $\epsilon_5$  is small,  $n$  is large and

$$\hat{R} < I(X; Y). \quad (12)$$

- Thus by (11),  $P(\mathcal{E}_{2i})$  vanishes as  $n \rightarrow \infty$  if  $\hat{R} < I(X; Y)$ .
- The receiver then declares that  $\hat{l}_i$  is sent if it is the unique message such that  $(u^n(\hat{l}_i|\hat{w}_{i-1}), y^n(i-1), x^n(\hat{w}_{i-1}))$  are jointly  $\epsilon$ -typical and  $\hat{l}_i \in \mathcal{B}(\hat{w}_i)$ ; otherwise it declares an error. Assume that  $(W_{i-1}, W_i, L_i)$  is sent in block  $i$  and let  $(\hat{W}_{i-1}, \hat{W}_i)$  be the receiver's estimate of  $(W_{i-1}, W_i)$ . Consider the following error events for the receiver,

$$\mathcal{E}_{31i} := \left\{ (U^n(L_i|\hat{W}_{i-1}), Y^n(i-1), X^n(\hat{W}_{i-1})) \notin \mathcal{T}_{\epsilon_6}^{(n)}(U, Y, X) \right\},$$

$$\mathcal{E}_{32i} := \left\{ (U^n(l|\hat{W}_{i-1}), Y^n(i-1), X^n(\hat{W}_{i-1})) \in \mathcal{T}_{\epsilon_6}^{(n)}(U, Y, X) \text{ for some } l \neq L_i, l \in \mathcal{B}(\hat{W}_i) \right\}.$$

The probability of decoding error at this step is upper bounded as

$$\begin{aligned} P(\mathcal{E}_{3i}) &\leq P\left(\mathcal{E}_{31i} \cup \mathcal{E}_{32i} \cup \left\{ \hat{W}_{i-1} \neq W_{i-1} \right\} \cup \mathcal{E}_{1i}\right) \\ &\leq P(\hat{W}_{i-1} \neq W_{i-1}) + P(\mathcal{E}_{1i}) + P(\mathcal{E}_{32i}) \\ &\quad + P\left(\mathcal{E}_{31i} \cap \left\{ \hat{W}_{i-1} = W_{i-1} \right\} \cap \mathcal{E}_{1i}^c\right). \quad (13) \end{aligned}$$

- The first two terms  $\rightarrow 0$  as  $n \rightarrow \infty$  if  $\hat{R} < I(X; Y)$  and  $R' > I(U, X; S)$ .
- Since  $\left\{ \hat{W}_{i-1} = W_{i-1} \right\} \cap \mathcal{E}_{1i}^c := \left\{ (U^n(L_i|\hat{W}_{i-1}), S^n(i-1), X^n(\hat{W}_{i-1})) \in \mathcal{T}_{\epsilon_2}^{(n)}(S, X, U) \right\}$  and since  $U - [S, X] - Y$ , by the Markov Lemma (see [17])  $P\left(\mathcal{E}_{31i} \cap \left\{ \hat{W}_{i-1} = W_{i-1} \right\} \cap \mathcal{E}_{1i}^c\right)$  goes to 0 as  $n \rightarrow \infty$ .
- To bound  $P(\mathcal{E}_{32i})$ , we first bound it above by

$$\begin{aligned} P(\mathcal{E}_{32i}) &= P\left\{ (U^n(l|\hat{W}_{i-1}), Y^n(i-1), X^n(\hat{W}_{i-1})) \in \mathcal{T}_{\epsilon_6}^{(n)}(U, Y, X) \text{ for some } l \neq L_i, l \in \mathcal{B}(\hat{W}_i) \right\} \\ &\leq P\left\{ (U^n(l|\hat{W}_{i-1}), Y^n(i-1), X^n(\hat{W}_{i-1})) \in \mathcal{T}_{\epsilon_6}^{(n)}(U, Y, X) \text{ for some } l \in \mathcal{B}(1) \right\}. \end{aligned}$$

The proof of this inequality is provided in [16]. Now, by the independence of the codebooks and by the

packing lemma  $P(\mathcal{E}_{32i})$  goes to 0 as  $n \rightarrow \infty$ , if

$$\hat{R} > R' - I(U; Y|X), \quad (14)$$

as there are  $2^{n(R' - \hat{R})}$  codewords  $U^n(\hat{L}_i|\hat{W}_{i-1}), \hat{L}_i \in \mathcal{B}(\hat{W}_i)$ .

- The total probability of error is then upper bounded by adding (9),(11) and (13) and is given by

$$P(\mathcal{E}_i) \leq P(\mathcal{E}_{1i}) + P(\mathcal{E}_{2i}) + P(\mathcal{E}_{3i}). \quad (15)$$

Combining the bounds in (10),(12) and (14) and performing Fourier-Motzkin elimination (see [16]) to remove  $\hat{R}$  and  $R'$ , we have shown that  $P(\mathcal{E}_i)$  goes to 0 as  $n \rightarrow \infty$  if

$$I(U, X; Y) > I(U, X; S) \quad (16)$$

- The reconstructed state sequence of block  $i - 1$  is then given by,

$$\begin{aligned} \hat{s}_j(i-1) &= f(u_j(l_i|w_{i-1}), x_j(w_{i-1}), y_j(i-1)), \\ &\quad \forall 1 \leq j \leq n. \quad (17) \end{aligned}$$

*Analysis of the expected distortion:* When there is no error,  $(S^n(i-1), X^n(W_{i-1}), U^n(L_i|W_{i-1}), Y^n(i-1)) \in \mathcal{T}_{\epsilon}^{(n)}(S, X, U, Y)$ . Thus the asymptotic distortion averaged over the random code and over  $(S^n, U^n, X^n, Y^n)$  is bounded as,

$$\begin{aligned} D &\leq \limsup_{n \rightarrow \infty} E \left[ d(S^n(i-1), \hat{S}^n(i-1)) \right] \\ &\stackrel{(a)}{=} \limsup_{n \rightarrow \infty} \sum_{j=1}^n E \left[ d(S_j(i-1), \hat{S}_j(i-1)) \right] \\ &\stackrel{(b)}{=} \limsup_{n \rightarrow \infty} \sum_{j=1}^n E \left[ d(S_j(i-1), f(U_j, X_j, Y_j)) \right] \\ &\stackrel{(c)}{\leq} \limsup_{n \rightarrow \infty} P(\mathcal{E}_i) D_{max} + (1 + \epsilon) P(\mathcal{E}_i^c) E \left[ d(S, f(U, X, Y)) \right] \\ &= D_{min} + \epsilon, \quad (18) \end{aligned}$$

where (a) follows from the definition of distortion and linearity of the expectation operator, (b) follows from the coding strategy and (c) follows from the Law of Total Expectation and Typical Average Lemma (see [16]).

This completes the proof of achievability for Theorem 1.

## VI. PROOF OF THE CONVERSE

In this section, we prove that for every code, the achieved distortion  $D \geq D_{min}$ . Before proving the converse, we introduce one key Lemma.

*Lemma 1:* For any three arbitrary random variables  $Z \in \mathcal{Z}$ ,  $V \in \mathcal{V}$  and  $T \in \mathcal{T}$ , where  $Z - T - V$  form a Markov chain and for a distortion function  $d : \mathcal{Z} \times \mathcal{Z} \mapsto \mathbb{R}^+ \cup \{0\}$ , we have

$$E[d(Z, f(V))] \geq \min_{g: \mathcal{T} \mapsto \mathcal{Z}} E[d(Z, g(T))], \quad (19)$$

for some arbitrary function  $f : \mathcal{V} \mapsto \mathcal{Z}$ .

This Lemma can be interpreted as the data-processing inequality for estimation theory.

*Proof:*

Using the law of iterated expectation, we have

$$E[d(Z, f(V))] = E_T[E[d(Z, f(V))|T]]. \quad (20)$$

Now, for each  $t \in \mathcal{T}$ ,

$$\begin{aligned} E[d(Z, f(V))|T = t] &= \sum_{z \in \mathcal{Z}, v \in \mathcal{V}} P(z|t)P(v|t)d(z, f(v)) \\ &= \sum_{v \in \mathcal{V}} P(v|t) \sum_{z \in \mathcal{Z}} P(z|t)d(z, f(v)) \\ &\geq \min_{v \in \mathcal{V}} \sum_{z \in \mathcal{Z}} P(z|t)d(z, f(v)) \quad (21) \\ &= \sum_{z \in \mathcal{Z}} P(z|t)d(z, f(v^*(t))), \end{aligned}$$

where  $v^*(t)$  attains the minimum in (21) for the given  $t$ . Define  $g(t) = f(v^*(t))$ . Then (20) becomes

$$\begin{aligned} E[d(Z, f(V))] &= E_T[E[d(Z, f(V))|T]] \\ &\geq E_T \left[ \sum_{z \in \mathcal{Z}} P(z|t)d(z, g(t)) \right] \\ &= E[d(Z, g(T))] \\ &\geq \min_{\tilde{g}: \mathcal{T} \rightarrow \mathcal{Z}} E[d(Z, \tilde{g}(T))], \end{aligned}$$

which completes the proof.  $\blacksquare$

Now, consider a  $(f_j, h_j, n)$ ,  $1 \leq j \leq n$ -code with distortion  $D$ . We have to show that  $D \geq D_{min}$ . We define  $U_j := (S^{j-1}, Y_{j+1}^n)$ , with  $(S_0, Y_{n+1}) = (\emptyset, \emptyset)$ . Note that as desired,  $U_j - [X_j, S_j] - Y_j$ ,  $1 \leq j \leq n$  form a Markov chain. Thus,

$$\begin{aligned} \sum_{j=1}^n I(U_j, X_j; S_j) &= \sum_{j=1}^n I(S^{j-1}, Y_{j+1}^n, X_j; S_j) \\ &\stackrel{(a)}{=} \sum_{j=1}^n I(S^{j-1}, Y_{j+1}^n; S_j) \\ &= \sum_{j=1}^n I(S^{j-1}; S_j) + I(Y_{j+1}^n; S_j|S^{j-1}) \\ &\stackrel{(b)}{=} \sum_{j=1}^n I(Y_{j+1}^n; S_j|S^{j-1}) \\ &\stackrel{(c)}{=} \sum_{j=1}^n I(S^{j-1}; Y_j|Y_{j+1}^n) \\ &\leq \sum_{i=1}^n I(S^{j-1}, Y_{j+1}^n, X_j; Y_j) \\ &= \sum_{j=1}^n I(U_j, X_j; Y_j), \quad (22) \end{aligned}$$

where (a) is true because  $X_j$  is a function of  $S^{j-1}$ , (b) follows from the fact that the channel state sequence is memoryless, and (c) follows from the Csiszár sum identity (see [16]). Let  $Q$  be a uniform random variable with  $P_Q(q) = 1/n$ ,  $1 \leq q \leq n$

that is independent of  $(S^n, X^n, Y^n)$ . Now,

$$\begin{aligned} \sum_{j=1}^n I(U_j, X_j; S_j) &\stackrel{(a)}{=} nI(U_Q, X_Q; S_Q|Q) \\ &\stackrel{(b)}{=} nI(U_Q, Q, X_Q; S_Q) \\ &\stackrel{(c)}{\leq} \sum_{j=1}^n I(U_j, X_j; Y_j) \\ &= nI(U_Q, X_Q; Y_Q|Q) \\ &\stackrel{(d)}{\leq} nI(U_Q, Q, X_Q; Y_Q), \quad (23) \end{aligned}$$

where (a) follows from the definition of conditional mutual information, (b) follows from the fact that  $Q$  is independent of  $S_Q$ , (c) follows from (22) and (d) follows from the chain rule. Now the distortion  $D$  can be bounded below as,

$$\begin{aligned} D &= E[d(S^n, \hat{S}^n)] \\ &\stackrel{(a)}{=} \sum_{j=1}^n \frac{1}{n} E[d(S_j, \hat{S}_j(Y^n))] \\ &\stackrel{(b)}{\geq} \sum_{j=1}^n \frac{1}{n} \min_{g_j} E[d(S_j, g_j(U_j, X_j, Y_j))] \\ &\stackrel{(c)}{\geq} \min_g E[d(S_Q, g(U_Q, Q, X_Q, Y_Q))], \quad (24) \end{aligned}$$

where (a) follows from the definition of distortion and the linearity of expectation operator and for (b) we apply Lemma 1. We recognize  $S_i$  as  $Z$ ,  $Y^n$  as  $V$ , and  $(U_j, X_j(S^{j-1}), Y_j)$  as  $T$ , and it is easy to verify that with  $(U_j, X_j(S^{j-1}), Y_j)$  given,  $Y^n$  is independent of  $S_j$  due to the fact that input codeword is a strictly causal function of the state sequence and also because of the memoryless property of the channel. Therefore Lemma 1 yields (b). The inequality in (c) follows from the definition of conditional expectation.

Now by defining  $(U_Q, Q) = U, S_Q = S, X_Q = X$  and  $Y_Q = Y$  we have the proof of the converse of Theorem 1.

## VII. CAPACITY-DISTORTION TRADE-OFF

In this section, we consider a scenario where, in addition to assisting the receiver in estimating the channel state, the transmitter also wishes to send additional pure information, independent of the state, over the discrete memoryless channel. Formally, based on the message index  $m \in [1, 2^{nR})$  and the channel state  $S^{j-1}$ , the transmitter chooses  $X_j(m, S^{j-1})$ ,  $1 \leq j \leq n$  and transmits it over the channel. After receiving  $Y^n$ , the receiver decodes  $\hat{m} \in [1, 2^{nR})$ , and forms an estimate  $\hat{S}^n(Y^n)$  of the channel state  $S^n$ .

The probability of a message decoding error and the state estimation error are given by

$$\lambda^{(n)} = \max_{m \in \mathcal{M}} \Pr[g_n(Y^n) \neq m | m \text{ is transmitted}] \quad (25)$$

$$\bar{D}^{(n)} = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} E \left[ \frac{1}{n} \sum_{i=1}^n d(S_i, \hat{S}_i) | m \text{ is transmitted} \right], \quad (26)$$

where the expectation is over the conditional joint distribution of  $(S^n, Y^n)$  conditioned on the message  $m \in \mathcal{M}$ .

A pair  $(R, D)$ , denoting a transmission rate and a state estimation distortion is said to be achievable if there exists a sequence of  $(\lceil 2^{nR} \rceil, n)$ -codes, indexed by  $n = 1, 2, \dots$ , such that  $\lim_{n \rightarrow \infty} \lambda^{(n)} = 0$ , and  $\lim_{n \rightarrow \infty} \bar{D}^{(n)} \leq D$ .

*Capacity-distortion function [11]:* For every  $D \geq 0$ , the capacity-distortion function  $C_{SC}(D)$  is the supremum of rates  $R$  such that  $(R, D)$  is an achievable transmission-state estimation tradeoff. The results of the last section can be utilised to characterize  $C_{SC}(D)$ .

*Theorem 2:* The capacity-distortion function for the problem considered is

$$C_{SC}(D) = \max_{\mathcal{P}_{\mathcal{D}}} I(U, X; Y) - I(U, X; S), \quad (27)$$

where

$$\mathcal{P}_{\mathcal{D}} = \{P_X, P_{U|X,S} : \xi(S|U, X, Y) \leq D\}. \quad (28)$$

where  $U$  is an auxiliary random variable of finite alphabet size.

Note that by choosing  $U = \emptyset$ , we can recover the capacity-distortion function results of [11], where it is assumed that both the transmitter and receiver have no knowledge of the channel state-sequence.

Theorem 2 can be shown by splitting the rate between pure information transmission and channel state estimation. The proof is omitted for brevity.

We summarize a few properties of  $C_{SC}(D)$  in Corollary 1 without proof.

*Corollary 1:* The capacity-distortion function  $C_{SC}(D)$  in Theorem 2 has the following properties:

- 1)  $C_{SC}(D)$  is a non-decreasing concave function of  $D$  for all  $D \geq D_{min}$ .
- 2)  $C_{SC}(D)$  is a continuous function of  $D$  for all  $D > D_{min}$ .
- 3)  $C_{SC}(D_{min}) = 0$  if  $D_{min} \neq 0$  and  $C_{SC}(D_{min}) \geq 0$  when  $D_{min} = 0$ .
- 4)  $C_{SC}(\infty)$  is the unconstrained channel capacity and is given by,

$$C_{SC}(\infty) = \max_{P_X} I(X; Y). \quad (29)$$

- 5) The condition for zero distortion at zero information rate is given by,  $\max_{P_X} I(X, S; Y) \geq H(S)$ .

*Remark:* The characterization in Theorem 2, albeit very compact, does not bring out the intrinsic tension between pure information transmission and state information transmission. Here, without proof, we provide a different, but equivalent, characterization of the region, which reveals more explicitly the capacity-distortion trade-off:

$$C_{SC}(D) = \max_{P_X} \left( I(X; Y) - E_X \left[ R_{WZ}^{(x)}(D) \right] \right), \quad (30)$$

where

$$R_{WZ}^{(x)}(D) = \min_{P_{U|x,S} : \xi(S|U,x,Y) \leq D} I(U; S|x, Y), \quad (31)$$

where  $x \in \mathcal{X}$ ,

is the Wyner-Ziv rate-distortion function for a given  $x \in \mathcal{X}$ . We observe the following:

- Eqn (30) states that the transmitter pays a rate penalty of  $R_{WZ}^{(x)}(D)$  to estimate the channel state at the receiver by transmitting  $x$ . Alternatively,  $R_{WZ}^{(x)}(D)$  can be viewed as the *estimation cost* due to signaling with  $x$ .
- Combining Property 3) of Corollary 1 and Equation (30), we can have an alternate characterization of the minimum achievable distortion, when  $D_{min} \neq 0$  and it is given by the solution of the following equation.

$$\max_{P_X} \left( I(X; Y) - E_X \left[ R_{WZ}^{(x)}(D_{min}) \right] \right) = 0, \quad (32)$$

where  $R_{WZ}^{(x)}(D_{min})$  is as defined in (31).

## VIII. ILLUSTRATIVE EXAMPLES

### A. State dependent Gaussian channel

Consider the state-dependent Gaussian channel:

$$Y_j = X_j(m, S^{j-1}) + S_j + Z_j, \quad 1 \leq j \leq n, \quad (33)$$

where  $S_j \sim \mathcal{N}(0, Q)$ ,  $Z_j \sim \mathcal{N}(0, N)$ , the transmitted signal has a power constraint of  $P$  and the channel input codeword is a strictly causal function of state  $S$  and the message  $m \in [1, 2^{nR})$ . The receiver wants to decode the message with vanishing probability of error for large  $n$  and also estimate the channel state in some distortion  $D$ . We consider the mean-squared Error (MSE) distortion measure. We wish to characterize the capacity-distortion  $(R, D)$  trade-off region.

The capacity-distortion function  $C_{SC}(D)$  of the state-dependent Gaussian channel with strictly causal state information at the transmitter is given by:

$$C_{SC}(D) = \begin{cases} \frac{1}{2} \log \frac{(P+Q+N)D}{QN}, & \frac{QN}{P+Q+N} \leq D \leq \frac{QN}{Q+N} \\ 0, & 0 \leq D \leq \frac{QN}{P+Q+N} \\ \frac{1}{2} \log \left( 1 + \frac{P}{Q+N} \right), & D \geq \frac{QN}{Q+N}. \end{cases} \quad (34)$$

Equation (34) can be proved using the alternative characterization of the capacity-distortion function provided in the last Section. If  $R_{WZ}^{(x)}(D)$  is independent of  $x$  or same for all  $x$ , then it is easy to see that the alternative characterization of  $C_{SC}(D)$  reduces to,

$$C_{SC}(D) = C_{SC}(\infty) - R_{WZ}(D). \quad (35)$$

We will use this observation to characterize  $C_{SC}(D)$  in this and the subsequent example. It is a well-known result that for additive Gaussian channel,  $R_{WZ}^{(x)}(D)$  is same as the rate-distortion function when both transmitter and receiver has the side information (see [19]). Thus  $R_{WZ}^{(x)}(D)$  is independent of  $x$  as by knowing  $x$  at both the transmitter and receiver, it can form an equivalent channel which is independent of  $x$ . Thus  $C_{SC}(D)$  for the Gaussian example is given by (35), where  $C_{SC}(\infty)$  and  $R_{WZ}(D)$  can be evaluated using standard results (for details see [16]).

*Discussion:* (1) It is obvious from Equation (34) that  $D_{min} = \frac{QN}{P+Q+N}$  and unconstrained capacity of the channel is given by  $C(\infty) = \frac{1}{2} \log \left( 1 + \frac{P}{Q+N} \right)$ , which is achieved by treating the interfering channel state as noise.

(2)  $D_{min} < D^* = QN/(Q+N)$ , where  $D^*$  is the minimum distortion achievable when the transmitter has no knowledge of

the state and it is achieved by first decoding the transmitted message  $X^n$  in a "non-coherent" fashion, then utilizing the reconstructed channel inputs along with the channel outputs  $Y^n$  to estimate the channel states as seen in [11].

Fig. 1 plots the rate-distortion region of the state-dependent additive Gaussian channel as we vary different channel parameters. As we increase the variance  $N$  of the additive Gaussian noise or the variance  $Q$  of the additive channel state sequence, the rate-distortion region decreases since, to achieve a distortion level  $D$  for the channel state, we need to allocate more rates to describe the channel state in the presence of high noise. If we increase the source power  $P$ , by keeping all the other parameters constant, the rate-distortion region increases as source has more power to describe the channel state. Note that the saturated part of each of the rate-distortion plot quantifies the unconstrained capacity of the channel.

### B. State dependent additive binary Channel

We next consider the example of a state-dependent additive binary channel given by,

$$Y_j = X_j \oplus S_j \oplus Z_j, \quad 1 \leq j \leq n, \quad (36)$$

where  $\oplus$  is the binary addition,  $S_j \sim \text{Ber}(p)$ ,  $Z_j \sim \text{Ber}(q)$ ,  $(p, q) \in [0, 1/2]$  and  $X_j$  is a binary random variable, which is a function of the message  $m$  and the state-sequence  $S^{j-1}$ . We note the following regarding the computation of the capacity-distortion function:

- $p = 0$  is a trivial case as under this condition,  $S = 0$  with probability 1.
- When  $q = 0$ , we can achieve zero distortion ( $D_{min} = 0$ ) by decoding  $X_j$  at the decoder and then cancelling its effect from the received  $Y_j$ , and the capacity distortion function in this case is constant and is given by the unconstrained capacity  $C_{SC}(D) = C_{SC}(\infty) = 1 - H_2(p)$ .
- When either of  $p$  or  $q = 1/2$ , then the capacity-distortion function is zero as the unconstrained capacity is zero in this case. Under this condition, it is easy to see that  $D_{min} = \min\{p, q\}$ .

For all other values of  $(p, q)$ , the capacity-distortion function  $C_{SC}(D)$  for the state-dependent additive binary channel with  $(p, q) \in (0, 1/2)$  is given by,

$$C_{SC}(D) = 1 - H_2(p * q) - R_{WZ}(D), \quad (37)$$

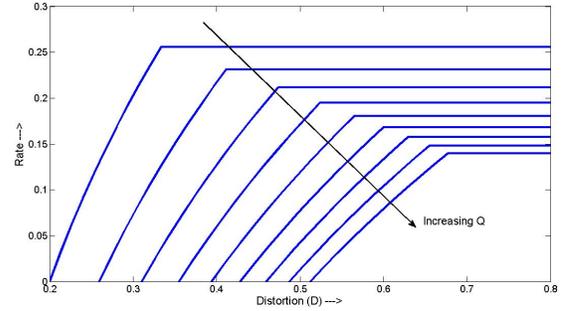
where

$$R_{WZ}(D) = \min_{\mu, \beta: \mu\beta + (1-\mu)\min\{p, q\} \leq D} \mu[H_2(p) - H_2(\beta) - H_2(p * q) + H_2(\beta * q)], \quad (38)$$

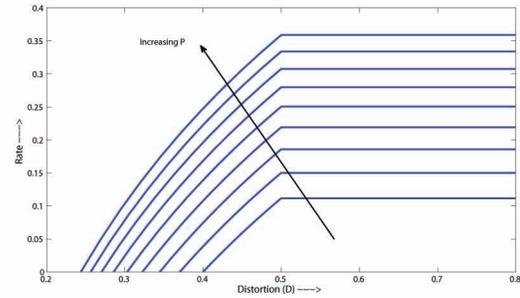
is the Wyner-Ziv rate-distortion function and  $H_2(\cdot)$  is a binary entropy function.

We will again use the alternate characterization to evaluate the capacity-distortion function of state-dependent binary channel. Using [19],  $R_{WZ}^{(x)}(D)$  for  $x = 0$  can be determined to be

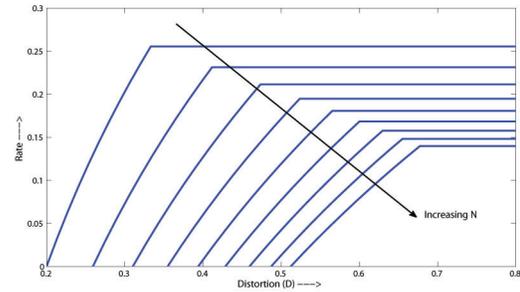
$$R_{WZ}^{(x=0)}(D) = \min_{\mu, \beta: \mu\beta + (1-\mu)\min\{p, q\} \leq D} \mu[H_2(p) - H_2(\beta) - H_2(p * q) + H_2(\beta * q)].$$



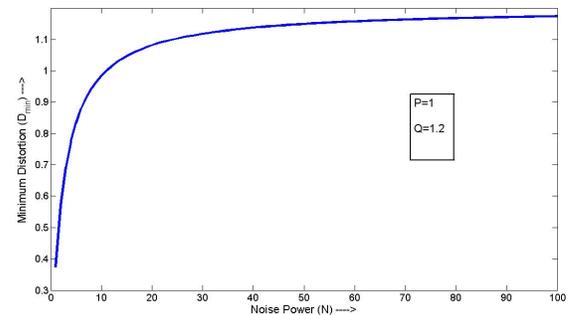
(a) Capacity-distortion regions with  $P = 1$ ,  $N = 1$  and  $Q$  varying from 0.5 to 2.



(b) Capacity-distortion regions with  $Q = 1$ ,  $N = 1$  and  $P$  varying from 0.5 to 2.



(c) Capacity-distortion regions with  $P = 1$ ,  $Q = 1$  and  $N$  varying from 0.5 to 2.



(d) Minimum achievable distortion with  $P = 1$ ,  $Q = 1.2$  and  $N$  varying from 1 to 100

Fig. 1. Capacity-distortion region for various channel parameter values

For  $x = 1$ , the Wyner-Ziv rate-distortion function can be similarly calculated and is given by

$$R_{WZ}^{(x=1)}(D) = \min_{\mu, \beta: \mu\beta + (1-\mu)\min\{p, q\} \leq D} \mu [H_2(p) - H_2(\beta) - H_2(p * (1 - q)) + H_2(\beta * (1 - q))].$$

It is easy to see that  $R_{WZ}^{(x)}(D)$  is same for both  $x = 0$  and  $x = 1$  because of fact that  $H_2(r) = H_2(1 - r)$ ,  $\forall r \in [0, 1]$ . Now if  $R_{WZ}^{(x)}(D)$  is same for all  $x$ , then by (35), the capacity distortion function for the binary channel can be achieved by choosing  $X \sim \text{Ber}(1/2)$ .

*Discussion:* (1) When  $(p, q) \in (0, 1/2)$ , it can be easily shown that if  $H_2(p) + H_2(q) < 1$ , then  $D_{min} = 0$ , otherwise  $D_{min}$  is given by the solution of the following equation:

$$1 - H_2(p * q) = R_{WZ}(D_{min}). \quad (39)$$

(2) When  $D \geq \min\{p, q\}$ , the capacity-distortion function is the unconstrained capacity  $1 - H_2(p * q)$ .

## IX. CONCLUDING REMARKS

The joint information transmission and channel state estimation problem for state-dependent channels was studied in [11], [13]. In [11], the case where the transmitter is oblivious of the channel state information was investigated and [13] studied state-dependent additive Gaussian channel with states available non-causally at the transmitter. In this paper, we bridge the gap between these two results by considering the joint communication and state estimation problem when the transmitter knows the channel state in a strictly causal manner. We showed that when the goal is to minimize the state estimation error at the receiver, the optimal transmission technique is to use block coding coupled with channel state estimation conditioned on treating the decoded message and received channel output as side information at the receiver. Pure information transmission obscures the receiver's view of the channel state, thereby increasing the state estimation error. For this intrinsic conflict, a simple rate-splitting technique achieves the optimal tradeoff. We also showed that the capacity-distortion function when transmitter is oblivious of the state information is a special case of our result.

## REFERENCES

- [1] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop.*, vol. 2, pp. 289–293, 1958.
- [2] A. V. Kusnetsov and B. S. Tsybakov, "Coding in a memory with defective cells," *Probl. Pered. Inform.*, vol. 10, no. 2, pp. 52–60, Apr./Jun. 1974. Translated from Russian.
- [3] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Contr. Inform. Theory*, vol. 9, no. 1, pp. 19–31, 1980.
- [4] C. Heegard and A. El Gamal, "On the capacity of computer memories with defects," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 5, pp. 731–739, Sep. 1983.
- [5] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 563–593, Mar. 2003.
- [6] B. Chen and G.W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.
- [7] W. Lee, D. Xiang, "Information-theoretic measures for anomaly detection," *IEEE Symposium on Security and Privacy*, 2001.
- [8] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE J. Select. Areas Comm.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.
- [9] M. Stojanovic, "Recent advances in high-speed underwater acoustic communications," in *IEEE Journal of Oceanic Engineering*, vol. 21, no. 2, pp. 125–137, Apr. 1996.
- [10] H. C. Papadopoulos and C.-E. W. Sundberg, "Simultaneous broadcasting of analog FM and digital audio signals by means of adaptive precanceling techniques," *IEEE Trans. Commun.*, vol. 46, no. 9, pp. 1233–1242, Sep. 1998.
- [11] W. Zhang, S. Vedantam, and U. Mitra, "A constrained channel coding approach to joint communication and channel estimation," in *Proc. IEEE Int. Symp. Inform. Theory*, Jul. 2008, Toronto, Canada.
- [12] A. Sutivong, T. M. Cover, M. Chiang, and Y.-H. Kim, "Rate vs. distortion trade-off for channels with state information," in *Proc. IEEE Int. Symp. Inform. Theory*, Jun./Jul. 2002, Lausanne, Switzerland.
- [13] A. Sutivong, M. Chiang, T. M. Cover, and Y.-H. Kim, "Channel capacity and state estimation for state-dependent Gaussian channels," *IEEE Transactions on Information Theory*, vol. 51(4), pp. 1486–1495, April 2005.
- [14] Y.-H. Kim, A. Sutivong and T. M. Cover, "State amplification," *IEEE Trans. on Information Theory*, vol. 54(5), pp. 1850–1859, May 2008.
- [15] T.M. Cover and J.A. Thomas, "Elements of Information Theory," New York: Wiley, 1991.
- [16] A. El Gamal, Y.-H. Kim, "Lecture Notes on Network Information Theory," arXiv:1001.3404.
- [17] G. Kramer, "Foundations and Trends in Communications and Information Theory," Vol. 4, Issue 4-5, ISSN 1567-2190.
- [18] T. M. Cover and A. El Gamal, "Capacity Theorems for the Relay Channels," *IEEE Trans. on Information Theory*, vol. 25(5), pp. 572–584, Sept. 1979.
- [19] A. D. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder," *IEEE Trans. on Information Theory*, vol. 22(1), pp. 1–10, Jan. 1976.