

Causal state amplification

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 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. 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

The effect of channel estimation errors on channel capacity has been long-studied in the context of wireless communications [1]. In these works, the focus is on maximizing channel capacity through the design of channel estimation strategies – the quality of the channel estimate is irrelevant. In contrast, herein, we examine the problem of jointly estimating an unknown channel and communicating over this channel. In particular, we examine the case where the transmitter has causal knowledge of the channel state and exploits this knowledge in codebook design. This scenario was first considered in the seminal work of [2] followed by that of [3], [4], [5]. There are a host of modern applications which can benefit from this analysis, including: multimedia information hiding [6], digital watermarking [7], data storage over memory with defects [3], [5], secret communication systems [8], dynamic spectrum access systems [9], underwater acoustic/sonar applications [10] etc.

In contrast to [4], [5], we study this problem of state information transmission over a state-dependent discrete memoryless channel; the transmitter has causal channel state information (CSI) 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. We also examine the case where additional

independent information is transmitted over the channel at the expense of a higher channel state estimation error. There is a natural tension between sending pure information and revealing the channel state. 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 via the *capacity-distortion function* (first introduced in [11]), which is fundamentally different from the rate-distortion function in lossy source coding [14]. 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.

At first glance, the work in [11] which examines the case where the transmitter and the receiver are oblivious of the channel state realization and [12] where the *rate-distortion* trade-off for the state-dependent additive Gaussian channel is considered for the case of the channel state being known *non-causally* at the transmitter appear to be unrelated; however, we show herein that these prior results can be recovered as special cases of our results. Our current work builds upon our prior efforts in [15] which treated the case where the transmitter has *strictly causal* state information. We observe that [13] considered the problem of transmitting data and state over a state-dependent channel with state information available both non-causally and causally at the sender. 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.

Our current results reveal two key insights: (1) the optimal strategy consists of non-coherently decoding the transmitted data and employing the decoded data as training to estimate the channel and (2) the additional degrees of freedom afforded by *causal* state information versus *strictly causal* state information at the transmitter provides a gain that is comparable to that achieved when going from no channel state information to *strictly causal* state information at the transmitter. We observe that (1) is in sharp contrast to how modern wireless systems are implemented, *i.e.* typically employing pilot sequences to aid in channel state information coupled with coherent decoding under the assumption that the estimated channel is close to the true channel.

This paper is organized as follows. Section II provides the channel model with discrete alphabets and formulates the problem of characterizing the minimum achievable distortion

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), CCF-0747111 and NSF CCF-0917343.

at zero information rate. Section III determines the minimum distortion, and Section IV presents a sketch of achievability and the proof of the converse for the key theorem of the work. Section V extends the results to the information rate-distortion trade-off setting, wherein we define and evaluate the capacity-distortion function. Section VI illustrates the application of the capacity-distortion function through the example of an additive state dependent binary channel. Finally, Section VII concludes the paper.

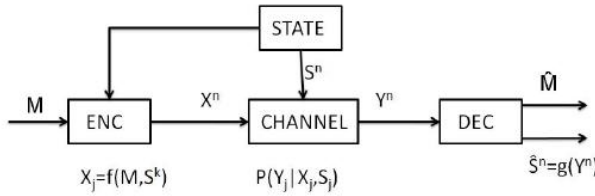


Fig. 1. Channel model for joint communication and estimation with causal ($k = i$) state information at transmitter.

II. 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 available causally at the transmitter.

Consider a discrete-memoryless channel $(\mathcal{X} \times \mathcal{S}, p(y|x, s), \mathcal{Y})$, which consists of a finite input alphabet $x \in \mathcal{X}$, a finite output alphabet $y \in \mathcal{Y}$, a finite state alphabet $s \in \mathcal{S}$, and a collection of conditional pmfs $p(y|x, s)$ on \mathcal{Y} for each value of $x \in \mathcal{X}$ and $s \in \mathcal{S}$. The state sequence $\{S_i\}$ is an i.i.d. process with $S_i \sim p(s_i)$ for each channel use. The channel is memoryless in a sense that, without feedback, $p(y^n|x^n, s^n) = \prod_{i=1}^n p(y_i|x_i, s_i)$, for $x_i \in \mathcal{X}$, $s_i \in \mathcal{S}$ and $y_i \in \mathcal{Y}$. 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., $d(s_i, \hat{s}_j) \leq D_{max} < \infty$ for any $1 \leq i, j \leq |\mathcal{S}|$. For any two length- n state sequences, the distortion is defined to be the average of the pairwise distortions, $\frac{1}{n} \sum_{j=1}^n d(s_j, \hat{s}_j)$.

A $(f_i, h_i), 1 \leq i \leq n$ code for the channel is defined as:

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

Distortion for channel state estimation: We consider the average distortion

$$\bar{D}^{(n)} = E \left[\frac{1}{n} \sum_{i=1}^n d(S_i, \hat{S}_i) \right],$$

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

$$D_{min} = \liminf_{n \rightarrow \infty} \min_{f_i, h_i, 1 \leq i \leq n} E \left[\frac{1}{n} \sum_{i=1}^n d(S_i, \hat{S}_i(Y^n)) \right],$$

which is the minimum distortion achievable for the channel model.

III. ZERO INFORMATION RATE CASE

It is not obvious how to use the casual knowledge of the channel state at the transmitter (TX). Motivated from our result in [15], one strategy would be to perform a block-Markov coding scheme, where at block j , the transmitter can use its knowledge of the state sequence of block $j-1$ to select a code that can be employed to estimate the state sequence of block $j-1$ at the decoder after receiving the channel output of block j . We could implement this by compressing the channel state of block $j-1$ and then sending the compression index through the channel at the capacity of the channel. This strategy can be improved as the compression index can be sent at a much lower rate by observing that the receiver has a side information of $(X^n(j-1), Y^n(j-1))$. Additionally, causal knowledge of the state at the transmitter can be used to further increase the rate compared with strictly causal state information by correlating the input codeword with current state (see [2]).

To illustrate this strategy, we consider a simple state-dependent Gaussian channel.

$$Y_i = X_i(S^k) + S_i + Z_i, \quad 1 \leq i \leq n,$$

where $S_i \sim \mathcal{N}(0, Q)$, $Z_i \sim \mathcal{N}(0, N)$, the transmitted signal has a power constraint of P . The receiver wants to estimate the channel state in minimum possible distortion. We consider the mean-squared Error (MSE) distortion measure. We wish to characterize the minimum achievable distortion D_{min} .

When transmitter is *oblivious* of the channel state $k = 0$, the minimum distortion is achieved by incoherent decoding, i.e., by decoding X first, and then use it as training symbol to estimate S . The minimum distortion is given by $D_{min} = Q \frac{N}{Q+N}$ (proved in [11]).

With *strictly causal* knowledge of state $k = i-1$, the minimum distortion is achieved by quantizing the channel state of the previous block and sending the quantization index across the channel using X . The D_{min} achieved with this strategy is given by $D_{min} = Q \frac{N}{P+Q+N}$ (see [15]).

The minimum distortion achievable for the state-dependent Gaussian channel with causal state information at the transmitter is given by

$$D_{min} = Q \frac{N}{(\sqrt{P} + \sqrt{Q})^2 + N}.$$

Achievability can be shown by choosing $X = \sqrt{\frac{P}{Q}} S$ and upon receiving the channel output, the receiver forms an minimum mean-square estimate of the channel state given the output to achieve the distortion D_{min} . The converse follows from the fact that D_{min} is also the minimum distortion

achieved when the channel state is known non-causally at the transmitter (see [12]).

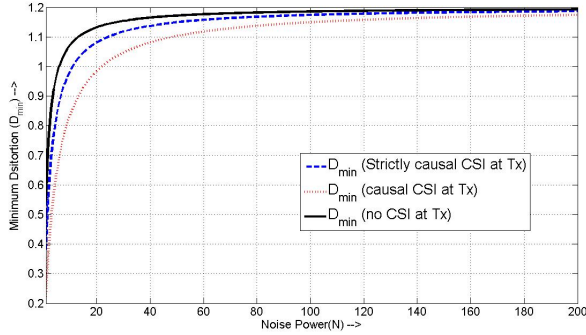


Fig. 2. Minimum achievable distortion with $P = 1, Q = 1.2$ and N varying from 1 to 200 for strictly causal, causal and no state information at the Tx

Figure 2 compares the D_{min} for the cases of strictly causal, causal and no CSI at the transmitter with varying N . It is evident that knowing channel state causally helps in achieving less distortion compared to strictly causal and no state information at the transmitter as then the channel codeword X can be made to be correlated with the current channel yielding a better estimate. For zero noise power N , D_{min} for all cases can be shown to converge towards 0, as then the decoded channel codeword can be effectively subtracted from the received signal to determine the channel state with zero distortion and hence a channel codeword correlated with the state cannot provide any further improvement.

To characterize the minimum distortion, we need the following:

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

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

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

With this definition, the minimum distortion with causal CSI at the transmitter is given by the following theorem.

Theorem 1: The minimum achievable distortion for the problem considered is

$$D_{min} = \min_{\mathcal{P}} \xi(S|U, V, Y),$$

where

$$\mathcal{P} = \{P_U, P_{V|U, S}, X = h(U, S) : I(U, V; Y) - I(U, V; S) \geq 0\}$$

and U and V are auxiliary random variables of finite alphabet size.

To minimize state estimation error, one might be tempted to use the channel in such a way that the pure information rate is maximized (i.e., is made equal to the channel capacity) and then use this pure information to describe the channel state. This technique can be shown to be suboptimal.

IV. PROOF OF THEOREM 1

In this section, we will give a sketch of the proofs of achievability and converse for Theorem 1.

A. Achievability

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

Codebook generation:

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

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

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

- 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 both the encoder and the decoder.

Encoding: Let, $u^n(w_{j-1})$ be the codeword sent in block $j - 1$. Knowing $s^n(j - 1)$ at the beginning of block j , the encoder looks for an index $l_j \in [1 : 2^{nR'}]$ such that $(v^n(l_j|w_{j-1}), s^n(j - 1), u^n(w_{j-1}))$ are jointly typical. If there is more than one such l_j , the smallest index is selected. If there is no such l_j , select $l_j = 1$. Determine w_j such that $l_j \in \mathcal{B}(w_j)$. Codeword $u^n(w_j)$ is selected for transmission in block j and it is transmitted to the receiver via $x^n(u^n(w_j), s^n)$ by symbol-by-symbol mapping $x_i(u^n(w_j), s^n) = h(u_i(w_j), s_i), \forall 1 \leq i \leq n$.

Decoding: At the end of the block j , the receiver declares \hat{w}_j was sent by looking for the uniquely typical $u^n(w_j)$ with $y^n(j)$. The receiver then declares that \hat{l}_j is sent if it is the unique message such that $(v^n(\hat{l}_j|\hat{w}_{j-1}), y^n(j - 1), u^n(\hat{w}_{j-1}))$ are jointly ϵ -typical and $\hat{l}_j \in \mathcal{B}(\hat{w}_j)$; otherwise it declares an error. The reconstructed state sequence of block $j - 1$ is then given by

$$\hat{s}_i(j - 1) = f(u_i(\hat{l}_j|\hat{w}_{j-1}), x_i(\hat{w}_{j-1}), y_i(j - 1)) \quad \forall 1 \leq i \leq n.$$

Following the analysis of [15] and [16], it can be easily shown that the scheme achieves the distortion given in Theorem 1 and the detailed proof is omitted. In comparing the strictly causal to the causal case, we see that knowledge of the current state provides an additional degree of freedom as evidenced by the additional auxiliary variable V . In fact the strictly causal results can be recovered from the causal results by substituting $U = X$ and $V = U$ in Theorem 1, yielding the result of Theorem 1 in [15]. Thus, the minimum distortion achievable with causal state information at the transmitter is upper bounded by D_{min} of the strictly causal case; the additional degree of freedom is exploited here by

performing the encoding operation over the set of all functions $\{x_u(s) : \mathcal{S} \mapsto \mathcal{X}\}$ indexed by u as the input alphabet. This technique of coding over functions onto \mathcal{X} instead of actual symbols in \mathcal{X} is referred to as the *Shannon strategy* and was introduced in [2].

B. Proof of the converse

To prove the converse, we must show that for every code, the achieved distortion $D \geq D_{min}$. Before sketching the converse, we review one key Lemma from [15], which can be interpreted as the data-processing inequality for estimation of a random variable in some distortion.

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

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

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

Now, consider a (f_i, h_i, n) , $1 \leq i \leq n$ -code with distortion D . We have to show that $D \geq D_{min}$. Define $U_i := S^{i-1}$, $V_i = Y_{i+1}^n$ with $(S_0, Y_{n+1}) = (\emptyset, \emptyset)$. Note that as desired, $U_i - [X_i, S_i] - Y_i$, $1 \leq i \leq n$ form a Markov chain. Now using standard information theoretic inequalities and Lemma 1 (similar to the converse proof of Theorem 1 in [15]), we can proof the converse of Theorem 1.

V. 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^i , the transmitter chooses $X_i(m, S^i)$, $1 \leq i \leq n$ 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

$$\lambda^{(n)} = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \Pr [g_n(Y^n) \neq m | m \text{ is transmitted}]$$

and

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

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 $\lim_{n \rightarrow \infty} \bar{D}^{(n)} \leq D$. We wish to characterize the capacity-distortion function $C_C(D)$ defined in [11].

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

$$C_C(D) = \max_{\mathcal{P}_D} I(U, V; Y) - I(U, V; S),$$

where

$$\mathcal{P}_D = \{P_U, P_{V|U, S}, X = h(U, S) : \xi(S|U, V, Y) \leq D\}.$$

where $U \in \mathcal{U}$ and $V \in \mathcal{V}$ are auxiliary random variables with cardinality $|\mathcal{U}| \leq \min\{(|\mathcal{X}| - 1)|\mathcal{S}| + 1, |\mathcal{Y}|\}$ and $|\mathcal{V}| \leq |\mathcal{S}| + 1$.

Note that by choosing $U = \emptyset$ and $V = X$, where X is independent of S , 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. And by choosing $U = X$, we recover our results on strictly causal state amplification (see [15]).

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 $C_C(D)$ in Corollary 1 (similar to the Corollary 1 in [15]) without proof.

Corollary 1: The capacity-distortion function $C_C(D)$ in Theorem 2 has the following properties:

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

$$C_C(\infty) = \max_{P_U, X=h(U, S)} I(U; Y),$$

which is the capacity of state-dependent channel with causal CSI at the transmitter only (see [2]).

(5) The condition for zero distortion at zero information rate is given by $\max_{P_{X|S}} I(X, S; Y) \geq H(S)$, which is also proved in [13].

Comparing with Corollary 1 of [15], we see that causal state knowledge gives an improvement in the unconstrained capacity of the channel as compared to the strictly causal case. Also note that the condition for zero distortion at zero information rate for the causal case is very similar to the strictly causal case, the only difference is in the maximizing distribution, as knowing the state sequence causally can be utilized to make the channel codeword X correlated with the current channel state S unlike the strictly causal case, where X is independent of S .

VI. ILLUSTRATIVE EXAMPLE

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

$$Y_i = X_i \oplus S_i \oplus Z_i, \quad 1 \leq i \leq n,$$

where \oplus is the binary addition, $S_i \sim \text{Ber}(\frac{1}{2})$, $Z_i \sim \text{Ber}(q)$, $q \in [0, 1/2]$ and X_i is a binary random variable, which is a

function of the message m and the state-sequence S^i . The tradeoff between communication and channel estimation is straightforward to observe if we consider this channel with $q = 0$: for good estimation of S , we want deterministic X as often as possible ($X = 0$ or 1 w.p. 1), whereas this would reduce the achieved information rate.

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_i at the decoder and then cancelling its effect from the received Y_i .
- If $D \geq \frac{1}{2}$, then the capacity distortion function is given by the unconstrained capacity $C_C(D) = C_C(\infty) = 1 - H_2(q)$.
- When $q = 1/2$, 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 $D_{min} = p$.

For all other values of q , the capacity-distortion function $C_C(D)$ is given by

$$C_C(D) = \max_{\lambda, t, \beta_1, \beta_2} \lambda t(1 - H_2(p) - H_2(\beta_1 * q) + H_2(\beta_1)) + \lambda(1 - t)(1 - H_2(p)) + (1 - \lambda)(1 - H_2(q) - H_2(p) + H_2(\beta_2))$$

subject to,

$$\lambda t \beta_1 + \lambda(1 - t)q + (1 - \lambda)\beta_2 \leq D.$$

The achievability of this capacity-distortion function uses time sharing argument (see [17]) and the converse can be shown by extending the converse proof of Wyner-Ziv rate-distortion function of binary source (see [19] for details) and is omitted here for brevity.

Note that substituting $\lambda = 1$ retrieves the $C_{SC}(D)$ result of state-dependent binary additive channel with strictly causal state at encoder (see [15]), which could be easily explained from the achievability, where in the first strategy we have to choose $X = U$, independent of S and thus it is same as in the strictly causal case. The improvement in the causal case comes from the second strategy, where we leverage the benefits of knowing the current channel state by choosing $U = X \oplus S$.

VII. CONCLUDING REMARKS

In this work, we bridge the gap between two problems of joint information transmission and channel state estimation for state-dependent channels: the case where the transmitter is oblivious to the channel [11] and that where the channel state is available non-causally [12]. Our current work extends our approach in [15] where strictly causal channel state is available to the transmitter; herein we show that measurable improvement is achieved by the knowledge of the current channel state (the so-called causal case). In contrast to the traditional practical system approach where training information is sent to the receiver to estimate the channel which is then employed for coherent decoding, our work shows that

non-coherent decoding should be employed as this decoded data should be used to learn the channel at the receiver. In particular we show that for the zero-information rate case, 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. For the transmission of additional independent information, a higher channel state estimation is achieved; however 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.

REFERENCES

- [1] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49(4), pp. 951-963, April 2003.
- [2] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop.*, vol. 2, pp. 289-293, 1958.
- [3] A. V. Kusunetsov 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] W. Lee, D. Xiang, "Information-theoretic measures for anomaly detection," *IEEE Symposium on Security and Privacy*, 2001.
- [9] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE J. Select. Areas Comm.*, vol. 23, no. 2, pp. 201-220, Feb. 2005.
- [10] 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.
- [11] W. Zhang, S. Vedantam, and U. Mitra, "A Constrained Channel Coding Approach to Joint Transmission and State Estimation Problem," accepted for publication in *IEEE Trans. Inform. Theory*, March 2011.
- [12] 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.
- [13] Y.-H. Kim, A. Sutivong and T. M. Cover, "State amplification," *IEEE Trans. on Information Theory*, vol. 54(5), pp. 1850-1859, May 2008.
- [14] T.M. Cover and J.A. Thomas, "Elements of Information Theory," New York: Wiley, 1991.
- [15] C. Choudhuri, Y.-H. Kim and Urbashi Mitra, "Capacity-distortion tradeoff in channels with state", *Allerton Conference*, Monticellon IL, Oct. 2010.
- [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.