Simultaneous Communication of Data and State

Thomas M. Cover
Information Systems Laboratory
Stanford University
Stanford, CA 94305, USA
cover@stanford.edu

Young-Han Kim
Department of ECE
University of California, San Diego
La Jolla, CA 92093, USA
yhk@ucsd.edu

Arak Sutivong
McKinsey & Company
87 Wireless Road
Bangkok 10330, Thailand
arak_sutivong@mckinsey.com

***Abstract*— We consider the problem of transmitting data at rate R over a state dependent channel $p(y|x, s)$ with the state information available at the sender and at the same time conveying the information about the channel state itself to the receiver. The amount of state information that can be learned at the receiver is captured by the mutual information $I(S^n; Y^n)$ between the state sequence S^n and the channel output Y^n. The optimal tradeoff is characterized between the information transmission rate R and the state uncertainty reduction rate Δ, when the state information is either causally or noncausally available at the sender. This result is closely related and in a sense dual to a recent study by Merhav and Shamai, which solves the problem of *masking* the state information from the receiver rather than conveying it.**

I. Introduction

A channel $p(y|x, s)$ with noncausal state information at the sender has capacity

$$C = \max_{p(u,x|s)} (I(U; Y) - I(U; S)) \tag{1}$$

as shown by Gelfand and Pinsker [6]. Transmitting at capacity, however, obscures the state information S^n as received by the receiver Y^n. In some instances we wish to convey the state information S^n to Y^n. For example, S^n could be time-varying fading parameters or an original image that we wish to enhance. Another motivation comes from cognitive radio systems [5], [10], [4], [7] with the additional assumption that the secondary user X^n communicates its own message and at the same time facilitates the transmission of the primary user's signal S^n. Here we wish to minimize the receiver's uncertainty about the state by reducing the size of the receiver's list of likely candidates of the state sequence S^n.

More precisely, we study the communication problem depicted in Figure 1. Here the sender has access to the channel

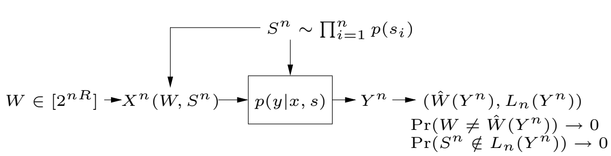

Fig. 1. Pure information transmission versus state uncertainty reduction.

state sequence $S^n = (S_1, S_2, \ldots, S_n)$, independently and identically distributed (i.i.d.) according to $\sim p(s)$ and wishes to transmit a message index $W \in \{1, 2, \ldots, 2^{nR}\}$, independent of S^n, as well as to help the receiver reduce the uncertainty about the channel state in n uses of a state dependent channel $(\mathcal{X} \times \mathcal{S}, p(y|x, s), \mathcal{Y})$. Based on the message W and the channel state S^n, the sender chooses $X^n(W, S^n)$ and transmits it across the channel. Upon observing the channel output Y^n, the receiver guesses $\hat{W} \in \{1, 2, \ldots, 2^{nR}\}$ and forms a list $L_n(Y^n) \subseteq \mathcal{S}^n$ that contains likely candidates of the actual state sequence S^n. The channel state uncertainty reduction rate Δ is given by

$$\Delta = H(S) - \frac{1}{n} \log |L_n|,$$

where $H(S)$ is the per-symbol channel state entropy and $|L_n|$ is the size of the candidate list L_n. In essence, the uncertainty reduction rate Δ captures the difference between the original channel state uncertainty and the residual state uncertainty after observing the channel output. Formally, we define a $(2^{nR}, 2^{n\Delta}, n)$ code as the encoder map

$$X^n : \{1, 2, \ldots, 2^{nR}\} \times \mathcal{S}^n \to \mathcal{X}^n,$$

and decoder maps

$$\hat{W} : \mathcal{Y}^n \to \{1, 2, \ldots, 2^{nR}\},$$
$$L_n : \mathcal{Y}^n \to 2^{\mathcal{S}^n}$$

with list size

$$|L_n| = 2^{n(H(S)-\Delta)}.$$

The probability of a message decoding error $P_{e,w}^{(n)}$ and the probability of a list decoding error $P_{e,s}^{(n)}$ are defined respectively as $P_{e,w}^{(n)} = \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \Pr(\hat{W} \neq w | W = w)$, $P_{e,s}^{(n)} = \Pr(S^n \notin L_n(Y^n))$, where the message index W is chosen uniformly over $\{1, \ldots, 2^{nR}\}$ and the state sequence S^n is drawn $\sim$ i.i.d. $p(s)$, independent of W. A pair (R, Δ) is said to be achievable if there exists a sequence of $(2^{nR}, 2^{n\Delta}, n)$ codes with $P_{e,w}^{(n)} \to 0$ and $P_{e,s}^{(n)} \to 0$ as $n \to \infty$.

Definition. The tradeoff region $\mathcal{R}^*$ is the closure of all achievable (R, Δ) pairs.

This paper shows that the tradeoff region $\mathcal{R}^*$ can be

1-4244-1429-6/07/$25.00 ©2007 IEEE

characterized as the union of all (R, Δ) pairs satisfying

$$\begin{aligned} R &\le I(U;Y) - I(U;S) \\ \Delta &\le H(S) \\ R + \Delta &\le I(X,S;Y) \end{aligned}$$

for some joint distribution $p(s)p(u,x|s)p(y|x,s)$.

In particular, the maximum uncertainty reduction rate

$$\Delta^* = \sup\{\Delta : (R,\Delta) \text{ is achievable for some } R \ge 0\}$$

is given by

$$\Delta^* = \min\{\max_{p(x|s)} I(X,S;Y),\ H(S)\}. \tag{2}$$

The maximum uncertainty reduction rate Δ^* is achieved by designing the signal X^n to enhance the receiver's estimation of the state S^n while using the remaining pure information bearing freedom in X^n to provide more information about the state.

When the state information is only *causally* available at the transmitter, that is, when the channel input X_i depends on only the past and the current channel channel state S^i, we will show that the tradeoff region $\mathcal{R}^*$ is given as the union of all (R, Δ) pairs satisfying

$$\begin{aligned} R &\le I(U;Y) \\ \Delta &\le H(S) \\ R + \Delta &\le I(X,S;Y) \end{aligned}$$

over all joint distributions $p(s)p(u)p(x|u,s)p(y|x,s)$. Interestingly, the maximum uncertainty reduction rate Δ^* stays the same as the noncausal case (2).

The problem of communication over state-dependent channels with states known at the sender has attracted a great deal of attention. This research area was first pioneered by Shannon [11], Kuznetsov and Tsybakov [8], and Gelfand and Pinsker [6]. Several advancements in both theory and practice have been made over the years. Most notably, Costa [1] studied the now famous "writing on dirty paper" problem and showed that the capacity of an additive white Gaussian noise channel is not affected by additional interference, as long as the entire interference sequence is available at the sender prior to the transmission. This fascinating result has been further extended with strong motivations from applications in digital watermarking and multi-antenna broadcast channels.

In [13], [14], we formulated the problem of simultaneously transmitting pure information and helping the receiver estimate the channel state under a distortion measure. Although the characterization of the optimal rate-distortion tradeoff is still open in general (cf. [12]), a complete solution is given for the Gaussian case (the writing on dirty paper channel) under quadratic distortion [13]. In this particular case, optimality was shown for a simple power-sharing scheme between pure information transmission via Costa's original coding scheme and state amplification via simple scaling.

Recently, Merhav and Shamai [9] considered a related problem of transmitting pure information, but this time under the additional requirement of minimizing the amount of information the receiver can learn about the channel state. In this interesting work, the optimal tradeoff between pure information rate R and the amount of state information E is characterized for both causal and noncausal setups. Furthermore, for the Gaussian noncausal case (writing on dirty paper), the optimal rate-distortion tradeoff is given under quadratic distortion. (This may well be called "writing dirty on paper".)

The current paper thus complements [9] in a dual manner. It is refreshing to note that our notion of uncertainty reduction rate Δ is essentially equivalent to Merhav and Shamai's notion of E; both notions capture the normalized mutual information $I(S^n; Y^n)$. (See the discussion in Section III.) The crucial difference is that Δ is to be maximized while E is to be minimized. Both problems admit single-letter optimal solutions.

The rest of this paper is organized as follows. In the next section, we establish the optimal (R, Δ) region for the case in which the state information S^n is noncausally available at the transmitter before the actual communication. Section III extends the notion of state uncertainty reduction to continuous alphabets, by identifying the list decoding requirement $S^n \in L_n(Y^n)$ with the mutual information rate $\frac{1}{n} I(S^n; Y^n)$. In particular, we characterize the optimal (R, Δ) region for Costa's "writing on dirty paper" channel. Since the intuition gained from the study of noncausal setup carries over to the case in which the transmitter has causal knowledge of the state sequence, the causal case is treated only briefly in Section IV, followed by concluding remarks in Section V.

II. Optimal (R, Δ) Tradeoff: Noncausal Case

In this section, we characterize the optimal tradeoff region between the pure information rate R and the state uncertainty reduction rate Δ with the state information noncausally available at the transmitter, as formulated in Section I.

Theorem 1. *The tradeoff region $\mathcal{R}^*$ for a state-dependent channel $(\mathcal{X} \times \mathcal{S}, p(y|x,s), \mathcal{Y})$ with the state information S^n noncausally known at the transmitter is the union of all (R, Δ) pairs satisfying*

$$R \le I(U;Y) - I(U;S) \tag{3}$$
$$\Delta \le H(S) \tag{4}$$
$$R + \Delta \le I(X,S;Y) \tag{5}$$

for some joint distribution $p(s)p(u,x|s)p(y|x,s)$, where the auxiliary random variable U has cardinality bounded by $|\mathcal{U}| \le |\mathcal{X}| + |\mathcal{S}|$.

It is easy to see that we can recover the Gelfand–Pinsker capacity formula

$$\begin{aligned} C &= \max\{R : (R,\Delta) \in \mathcal{R}^* \text{ for some } \Delta \ge 0\} \\ &= \max_{p(x,u|s)} (I(U;Y) - I(U;S)). \end{aligned}$$

On the other extreme, we have the following result.

Corollary 1. *Under the condition of Theorem 1, the maximum uncertainty reduction rate* $\Delta^* = \max\{\Delta : (R, \Delta) \in \mathcal{R}^*$ *for some* $R \geq 0\}$ *is given by*

$$\Delta^* = \min\{\max_{p(x|s)} I(X, S; Y),\ H(S)\}. \tag{6}$$

Thus, the receiver can learn about the state S^n essentially at the maximum cut-set rate $\max_{p(x|s)} I(X, S; Y)$.

Before we prove Theorem 1, we need the following two lemmas. The first one extends Fano's inequality [3, Lemma 7.9.1] to list decoding.

Lemma 1. *For a sequence of list decoders* $L_n : \mathcal{Y}^n \to 2^{\mathcal{S}^n},\ Y^n \mapsto L_n(Y^n)$ *with list size* $|L_n|$ *fixed for each* n*, let* $P_{e,s}^{(n)} = \Pr(S^n \notin L_n(Y^n))$ *be the sequence of corresponding probabilities of list decoding error. If* $P_{e,s}^{(n)} \to 0$*, then*

$$H(S^n|Y^n) \leq \log |L_n| + n\epsilon_n,$$

where $\epsilon_n \to 0$ *as* $n \to \infty$.

Proof. Readers are referred to [2]. □

The second lemma is crucial to the proof of Theorem 1 and contains a more interesting technique than Lemma 1.

Lemma 2. *Let* $\mathcal{R}$ *be the union of all* (R, Δ) *pairs satisfying* (3)–(5)*. Let* $\mathcal{R}_0$ *be the closure of the union of all* (R, Δ) *pairs satisfying* (3), (4)*, and*

$$R + \Delta \leq I(U, S; Y) \tag{7}$$

for some joint distribution $p(s)p(x, u|s)p(y|x, s)$ *where the auxiliary random variable* U *has finite cardinality. Then*

$$\mathcal{R} = \mathcal{R}_0.$$

Proof. Omitted. Refer to [2]. □

Now we are ready to prove Theorem 1.

Proof sketch of Theorem 1. For the proof of achievability, in the light of Lemma 2, it suffices to prove that any pair (R, Δ) satisfying (3), (4), (7) for some $p(u, x|s)$ is achievable. Since the coding technique is quite standard, we only sketch the proof here. For fixed $p(u, x|s)$, the result of Gelfand–Pinsker [6] shows that the transmitter can send $I(U; Y) - I(U; S)$ bits across the channel. Now we allocate $0 \leq R \leq I(U; Y) - I(U; S)$ bits for sending the pure information and use the remaining $\Gamma = I(U; Y) - I(U; S) - R$ bits for sending the state information by random binning (i.e., sending the random hash index of S^n). At the receiving end, the receiver is able to decode the codeword U^n from Y^n. Using joint typicality of (U^n, Y^n, S^n), the state uncertainty can be first reduced from $H(S)$ to $H(S|Y, U)$. In addition, using $\Gamma = I(U; Y) - I(U; S) - R$ bits of refinement information from the transmitter, we can further reduce the state uncertainty, resulting in the total state uncertainty reduction rate $\Delta = I(U, Y; S) + I(U; Y) - I(U; S) - R = I(U, S; Y) - R$. By varying $0 \leq R \leq I(U; Y) - I(U; S)$, it can be readily seen that all (R, Δ) pairs satisfying

$$\begin{aligned} R &\leq I(U; Y) - I(U; S) \\ \Delta &\leq H(S) \\ R + \Delta &\leq I(U, S; Y) \end{aligned}$$

for any fixed $p(x, u|s)$ are achievable.

For the proof of converse, we have to show that given any sequence of $(2^{nR}, 2^{n\Delta}, n)$ codes with $P_{e,w}^{(n)}, P_{e,s}^{(n)} \to 0$, the (R, Δ) pairs must satisfy

$$\begin{aligned} R &\leq I(U; Y) - I(U; S) \\ \Delta &\leq H(S) \\ R + \Delta &\leq I(X, S; Y) \end{aligned}$$

for some joint distribution $p(s)p(x, u|s)p(y|x, s)$.

The pure information rate R can be readily bounded from the previous work by Gelfand and Pinsker [6, Proposition 3]. On the other hand, since $\log |L_n| = n(H(S) - \Delta)$, we can trivially bound Δ by Lemma 1 as

$$\begin{aligned} n\Delta &\leq nH(S) - H(S^n|Y^n) + n\epsilon'_n \\ &\leq nH(S) + n\epsilon'_n. \end{aligned}$$

Similarly, we can bound $R + \Delta$ as

$$\begin{aligned} n(R + \Delta) &\leq I(W; Y^n) + I(S^n; Y^n) + n\epsilon''_n \\ &\overset{(a)}{\leq} I(W; Y^n|S^n) + I(S^n; Y^n) + n\epsilon''_n \\ &\leq I(W, S^n; Y^n) + n\epsilon''_n \\ &\overset{(b)}{=} I(X^n, S^n; Y^n) + n\epsilon''_n \\ &\overset{(c)}{\leq} \frac{1}{n} \sum_{i=1}^{n} I(X_i, S_i; Y_i) + \epsilon''_n, \end{aligned} \tag{8}$$

where (a) follows since W is independent of S^n and conditioning reduces entropy, (b) follows from the data processing inequality (both directions), and (c) follows from the memorylessness of the channel. Using the usual time-sharing random variable completes the proof. □

III. Extension to Continuous State Space

In the previous section, we characterized the tradeoff region $\mathcal{R}^*$ between the pure information rate R and the state uncertainty reduction rate Δ with noncausal state information at the transmitter. From the definition of uncertainty reduction rate Δ based on the size of the receiver's candidate state sequence list $L_n(Y^n)$ with $\log |L_n(Y^n)| = n(H(S) - \Delta)$, it appears at first that our notion of uncertainty reduction rate Δ is meaningful only when the channel state S has finite cardinality (i.e., $|\mathcal{S}| < \infty$), or at least when $H(S) < \infty$.

However, the proof of Theorem 1 (the generalized Fano's inequality in Lemma 1), along with the fact that the optimal region is single-letterizable, reveals that the reduction of list size from $nH(S)$ to $\log |L_n(Y^n)|$ is not fundamental to the notion of state uncertainty reduction rate Δ. What Δ really captures is the normalized mutual information $\frac{1}{n}I(S^n;Y^n)$, which can be defined for arbitrary state space $\mathcal{S}$.

More precisely, we define a $(2^{nR},n)$ code by an encoding function $X^n : \{1,2,\dots,2^{nR}\} \times \mathcal{S}^n \to \mathcal{X}^n$ and a decoding function $\hat{W} : \mathcal{Y}^n \to \{1,2,\dots,2^{nR}\}$. Then, the associated state uncertainty reduction rate with the $(2^{nR},n)$ code is defined as

$$\Delta_I = \frac{1}{n}I(S^n;Y^n),$$

where the mutual information is taken over the joint distribution $p(x^n,s^n,y^n) = p(x^n|s^n)\prod_{i=1}^n\big(p(s_i)p(y_i|x_i,s_i)\big)$ induced by $X^n(W,S^n)$ with the message W distributed uniformly over $\{1,\dots,2^{nR}\}$ independent of S^n. Similarly, the probability of error is defined as $P_e^{(n)} = \Pr(W \neq \hat{W}(Y^n))$. A pair (R,Δ_I) is said to be achievable if there exists a sequence of $(2^{nR},n)$ codes with $P_e^{(n)} \to 0$ and

$$\lim_{n\to\infty}\frac{1}{n}I(S^n;Y^n) \geq \Delta_I.$$

The closure of all achievable (R,Δ_I) pairs is again called the tradeoff region $\mathcal{R}_I^*$. (Here we use the notation $\mathcal{R}_I^*$ and Δ_I instead of $\mathcal{R}^*$ and Δ to distinguish from the original problem formulated in terms of the list size reduction.)

Proposition 1. *The tradeoff region $\mathcal{R}_I^*$ for a state-dependent channel $(\mathcal{X}\times\mathcal{S}, p(y|x,s), \mathcal{Y})$ with the state information S^n noncausally known at the transmitter is the closure of all (R,Δ_I) pairs satisfying*

$$\begin{aligned} R &\leq I(U;Y) - I(U;S) \\ \Delta_I &\leq H(S) \\ R+\Delta_I &\leq I(X,S;Y), \end{aligned}$$

for some joint distribution $p(s)p(u,x|s)p(y|x,s)$ with auxiliary random variable U.

Proof sketch. The proof of the converse follows trivially from the intermediate steps in the proof of the converse for Theorem 1.

For the achievability, we first use a finite partition to quantize the state random variable S into $[S]$. Under this partition, we pick a pair (R,Δ) that is achievable with respect to the original list size reduction problem in Theorem 1. Now from the generalized Fano's inequality (Lemma 1), the achievable uncertainty reduction rate Δ satisfies

$$n\Delta \leq I([S]^n;Y^n) + n\epsilon_n$$

with $\epsilon_n \to 0$ as $n\to\infty$, which implies that

$$\Delta_I = \lim_{n\to\infty}\frac{1}{n}I([S]^n;Y^n) \geq \lim_{n\to\infty}\frac{1}{n}I([S]^n;L_n(Y^n)) = \Delta.$$

Hence, $\mathcal{R}_I^* \supseteq \mathcal{R}^*$ for the given partition. By taking a sequence of partitions with mesh $\to 0$, we have the desired achievability. □

It turns out there is an alternative coding scheme based on the Wyner–Ziv source coding with side information [15] that can also achieve the tradeoff region $\mathcal{R}_I^*$. For the details, refer to [2].

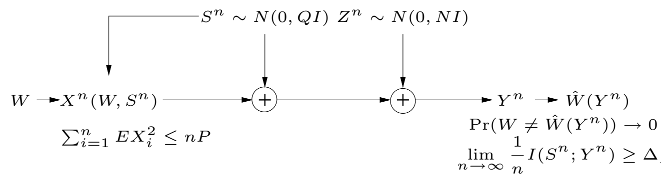

Fig. 2. Writing on dirty paper.

Now we consider Costa's writing on dirty paper model depicted in Figure 2 as the canonical example of continuous state-dependent channel. Here the channel output is given by $Y^n = X^n + S^n + Z^n$, where $X^n(W,S^n)$ is the channel input subject to a power constraint $\sum_{i=1}^n EX_i^2 \leq nP$, $S^n \sim N(0,QI)$ is the additive white Gaussian state, and $Z^n \sim N(0,NI)$ is the white Gaussian noise. We assume that S^n and Z^n are independent.

Proposition 2. *The tradeoff region $\mathcal{R}_I^*$ for the Gaussian channel depicted in Figure 2 is characterized by the boundary points $(R(\gamma),\Delta_I(\gamma))$, $0\leq\gamma\leq 1$, where*

$$R(\gamma) = \frac{1}{2}\log\left(1+\frac{\gamma P}{N}\right) \quad (9)$$

$$\Delta_I(\gamma) = \frac{1}{2}\log\left(1+\frac{\left(\sqrt{Q}+\sqrt{(1-\gamma)P}\right)^2}{\gamma P+N}\right). \quad (10)$$

Proof sketch. The achievability follows from Proposition 1 with trivial extension to the input power constraint. In particular, we use the simple power sharing scheme proposed in [13], where a fraction γ of the input power is used to transmit the pure information using Costa's writing on dirty paper coding technique, while the remaining $(1-\gamma)$ fraction of the power is used to amplify the state. In other words,

$$X = V + \sqrt{(1-\gamma)\frac{P}{Q}}\,S$$

with $V \sim N(0,\gamma P)$ independent of S, and

$$U = V + \alpha S$$

with

$$\alpha = \frac{\gamma P}{\gamma P+N}\sqrt{\frac{(1-\gamma)P+Q}{Q}}.$$

Evaluating $R = I(U;Y) - I(U;S)$ and $\Delta_I = I(S;Y)$ for each γ, we recover (9) and (10).

The proof of converse requires a little more work, but is essentially same as that of [13, Theorem 2], which we do not repeat here. □

As two extreme points of the (R,Δ_I) tradeoff region, we have on one hand Costa's writing on dirty paper result

$$C = \frac{1}{2}\log\left(1+\frac{P}{N}\right),$$

and on the other hand

$$\Delta_I^* = \frac{1}{2}\log\left(1 + \frac{\left(\sqrt{P}+\sqrt{Q}\right)^2}{N}\right),$$

which is achieved by taking $X = (\sqrt{P/Q})S$.

IV. Optimal (R, Δ) Tradeoff: Causal Case

The previous two sections considered the case in which the transmitter has complete knowledge of the state sequence S^n prior to the actual communication. In this section, we consider another model in which the transmitter learns the state sequence on the fly, hence the encoding function

$$X_i : \{1, 2, \ldots, 2^{nR}\} \times \mathcal{S}^i \to \mathcal{X}, \qquad i = 1, 2, \ldots, n,$$

depends causally on the state sequence.

Theorem 2. *The tradeoff region $\mathcal{R}^*$ for a state-dependent channel $(\mathcal{X} \times \mathcal{S}, p(y|x,s), \mathcal{Y})$ with the state information S^n causally known at the transmitter is the union of all (R, Δ) pairs satisfying*

$$R \le I(U;Y) \tag{11}$$

$$\Delta \le H(S) \tag{12}$$

$$R + \Delta \le I(X, S; Y) \tag{13}$$

for some joint distribution $p(s)p(u)p(x|u,s)p(y|x,s)$, where the auxiliary random variable U has cardinality bounded by $|\mathcal{U}| \le |\mathcal{X}|^{|\mathcal{S}|}$.

Since the proof of Theorem 2 is essentially identical to that of Theorem 1, we skip it here.

As one extreme point of the tradeoff region $\mathcal{R}^*$, we recover the Shannon capacity formula [11] for channels with causal side information at the transmitter as follows:

$$C = \max_{p(u)p(x|u,s)} I(U;Y), \tag{14}$$

which is in general less than (1). On the other hand, the maximum uncertainty reduction rate Δ^* is identical to that for the noncausal case given in Corollary 1.

Corollary 2. *Under the condition of Theorem 2, the maximum uncertainty reduction rate Δ^* is given by*

$$\Delta^* = \min\{\max_{p(x|s)} I(X, S; Y),\ H(S)\}. \tag{15}$$

Thus, the receiver can learn about the state essentially at the maximum cut-set rate, even under the causality constraint.

V. Concluding Remarks

Because the channel is state dependent, the receiver is able to learn something about the channel state from directly observing the channel output. Thus, to help the receiver narrow down the uncertainty about the channel state at the highest rate possible, the sender must jointly optimize between facilitating state estimation and transmitting refinement information, rather than merely using the channel capacity to send the state description. In particular, the transmitter should summarize the state information in such a way that the summary information results in the maximum uncertainty reduction when coupled with the receiver's initial estimate of the state. More generally, by taking away some resources used to help the receiver reduce the state uncertainty, the transmitter can send additional pure information to the receiver and trace the entire (R, Δ) tradeoff region.

There are three surprises here. First, the receiver can learn about the channel state and the independent message at a maximum cut-set rate $I(X, S; Y)$ over all joint distributions $p(x, s)$ consistent with the given state distribution $p(s)$. Second, to help the receiver reduce the uncertainty in the initial estimate of the state (namely, to increase the mutual information from $I(S;Y)$ to $I(X, S; Y)$), the transmitter can allocate the achievable information rate $I(U;Y) - I(U;S)$ in two alternative methods—random binning and its dual, random covering. Thirdly, as far as the sum rate $R + \Delta$ and the maximum uncertainty reduction rate Δ^* are concerned, there is no cost associated with restricting the encoder to learn the state sequence on the fly.

References

[1] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 439–441, 1983.

[2] T. M. Cover, Y.-H. Kim, and A. Sutivong, "Channel capacity and state uncertainty reduction for state-dependent channels," preprint, February 2007. [Online]. Available: http://arxiv.org/abs/cs.IT/0703005/

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

[4] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Trans. Inf. Theory*, vol. IT-52, no. 5, pp. 1813–1827, 2006.

[5] Federal Communications Commission, Cognitive Radio Technologies Proceeding (CRTP), ET Docket, no. 03-108. [Online]. Available: http://www.fcc.gov/oet/cognitiveradio/

[6] S. I. Gelfand and M. S. Pinsker, "Coding for channel with random parameters," *Problems Control Inform. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[7] A. Jovičić and P. Viswanath, "Cognitive radio: an information-theoretic perspective," submitted to *IEEE Trans. Inf. Theory,* 2006. [Online]. Available: http://arxiv.org/abs/cs.IT/0604107/

[8] A. V. Kuznetsov and B. S. Tsybakov, "Coding in a memory with defective cells," *Problemy Peredachi Informatsii*, vol. 10, no. 2, pp. 52–60, 1974.

[9] N. Merhav and S. Shamai, "Information rates subjected to state masking," submitted to *IEEE Trans. Inf. Theory,* 2006. [Online]. Available: http://www.ee.technion.ac.il/people/merhav/papers/p101.pdf

[10] J. Mitolla, III, "Cognitive radio: an integrated agent architecture for software defined radio," Ph.D. Thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2000.

[11] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop.*, vol. 2, pp. 289–293, 1958.

[12] A. Sutivong, "Channel capacity and state estimation for state-dependent channels," Ph.D. Thesis, Stanford University, Mar. 2003.

[13] A. Sutivong, M. Chiang, T. M. Cover, and Y.-H. Kim, "Channel capacity and state estimation for state-dependent Gaussian channels," *IEEE Trans. Inf. Theory*, vol. IT-51, no. 4, pp. 1486–1495, 2005.

[14] 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*, Lausanne, Switzerland, June/July 2002, p. 226.

[15] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–10, 1976.