Channel Capacity and State Estimation for State-Dependent Gaussian Channels*

Arak Sutivong[†], Mung Chiang[‡], Thomas M. Cover[§], and Young-Han Kim[¶]
Electrical Engineering Department, Stanford University, CA 94305

October 22, 2002

Abstract

We formulate a problem of state information transmission over a state-dependent channel with states known at the transmitter. In particular, we solve a problem of minimizing the mean-squared channel state estimation error for a state-dependent additive Gaussian channel $Y^n = X^n + S^n + Z^n$ with an i.i.d. Gaussian state sequence S^n known at the transmitter and an unknown i.i.d. additive Gaussian noise Z^n. We show that a simple technique of direct state amplification (i.e., $X^n = \alpha S^n$), where the transmitter uses its entire power budget to amplify the channel state, yields the minimum mean-squared 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. We characterize the optimal tradeoff between the rate R of the independent information that can be reliably transmitted and the mean-squared state estimation error D. We show that any optimal (R, D) tradeoff pair can be achieved via a power-sharing technique, whereby the transmitter power is appropriately allocated between pure information transmission and state amplification.

Keywords: Additive Gaussian noise channels, channels with state information, joint source-channel coding, state amplification, state estimation.

[*] Submitted to *IEEE Trans. Inform. Theory,* October 2002. Part of the work was presented at the International Symposium on Information Theory and Its Applications 2000 (ISITA 2000, [2]), Hawaii, USA, November 2000 and at the International Symposium on Information Theory 2001 (ISIT 2001, [13]), Washington D.C., USA, June 2001.

[†] This work was partially supported by NSF CCR-9973134 and Stanford Networking Research Center (SNRC).

[‡] This work was supported by the Hertz Foundation Fellowship and the Stanford Graduate Fellowship.

[§] This work was partially supported by NSF CCR-9973134 and MURI DAAD-19-99-1-0215.

[¶] This work was partially supported by NSF CCR-9973134 and Stanford Networking Research Center (SNRC).

1 Introduction

In many communication scenarios, the communicating parties typically 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. A particular area of research in communication with state information that has attracted a great deal of attention is a study of transmission over state-dependent channels[1] with state information available at the transmitter. This area of research was considered by Shannon in 1958 [12], Kusnetsov and Tsybakov [9], and Gel'fand and Pinsker [7].

Shown in Fig. 1 is the setup of a channel with state information available only at the sender, as considered by Gel'fand and Pinsker in [7]. Here, the transmitter wishes to send pure information $W \in \{1, 2, \ldots, 2^{nR}\}$, independent of the channel state, in n uses of a discrete memoryless state-dependent channel $p(y|x, s)$ with state $S^n = (S_1, S_2, \ldots, S_n)$ known at the transmitter. Based on pure information W and channel state S^n, the transmitter chooses $X^n(W, S^n)$ and sends it across the channel. Upon receiving Y^n, the receiver guesses $\hat{W}(Y^n) \in \{1, 2, \ldots, 2^{nR}\}$. Applications of this model include multimedia information hiding [10], digital watermarking [1], [3], multi-antenna broadcast, and data storage over memory with defects [8], [9], etc.

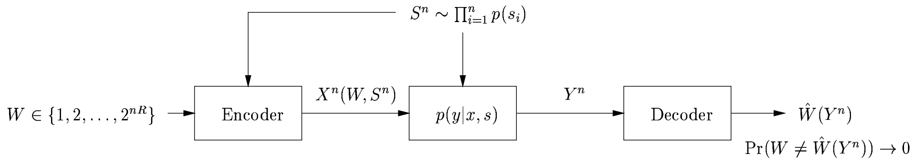
$S^n \sim \prod_{i=1}^n p(s_i)$
$W \in \{1, 2, \ldots, 2^{nR}\}$
Encoder
$X^n(W, S^n)$
$p(y|x, s)$
Y^n
Decoder
$\hat{W}(Y^n)$
$\Pr(W \neq \hat{W}(Y^n)) \to 0$

Figure 1: Pure information transmission over a state-dependent channel with states known at the transmitter.

Most of the existing literature has focused on determining the channel capacity or devising practical capacity-achieving coding techniques. 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. The model of this communication scenario is shown in Fig. 2. In this setup, the transmitter wishes to help the receiver estimate the channel state $S^n = (S_1, S_2, \ldots, S_n)$ of a state-dependent channel $p(y|x, s)$. Based on the channel state S^n, the transmitter transmits $X^n(S^n)$. Upon observing the channel output Y^n, the receiver forms an estimate of the channel state $\hat{S}^n(Y^n) \in \hat{\mathcal{S}}^n$. The channel state estimation error is given by $Ed(S^n, \hat{S}^n) = \frac{1}{n}\sum_{i=1}^n Ed(S_i, \hat{S}_i)$, where $d : \mathcal{S} \times \hat{\mathcal{S}} \to \mathbf{R}$ is a distortion measure between the channel S and its reconstruction $\hat{S}$.

An example of the above communication scenario is an analog-digital hybrid radio system [11], where digital refinement information is overlayed on top of the existing legacy analog transmission, which must be kept intact due to backward compatibility requirements, to help improve the detection and reconstruction of the original analog signal. In this example, the

[1] A state-dependent channel is simply a channel whose output conditional distribution depends on a time-varying random parameter called the channel state.

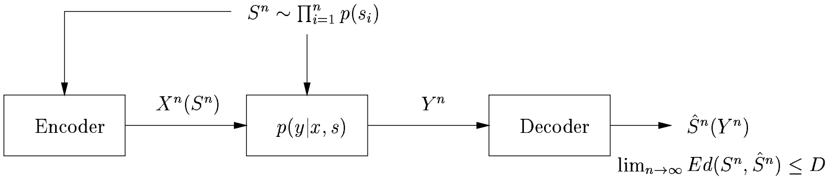
$S^n \sim \prod_{i=1}^n p(s_i)$
Encoder
$X^n(S^n)$
$p(y|x,s)$
Y^n
Decoder
$\hat{S}^n(Y^n)$
$\lim_{n\to\infty} Ed(S^n, \hat{S}^n) \leq D$

Figure 2: State information transmission over a state-dependent channel with states known at the transmitter.

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 work, we attempt to formulate and solve this problem of state information transmission. We consider a specific problem of state information transmission over a state-dependent additive Gaussian channel as shown in Fig. 3. In this setup, the transmitter has access to the channel state $S^n = (S_1, S_2, \ldots, S_n)$, $S_i \sim \mathcal{N}(0, Q)$, and wishes to help reveal it to the receiver. Based on the channel state S^n, the transmitter transmits $X^n(S^n)$, subject to a power constraint P. Upon receiving the output $Y^n = X^n(S^n) + S^n + Z^n$, where $Z^n \sim \mathcal{N}(0, N)$ is an unknown i.i.d. additive Gaussian noise, the receiver forms an estimate $\hat{S}^n(Y^n)$. The goal is to minimize the mean-squared state estimation error $E||S^n - \hat{S}^n||^2$.

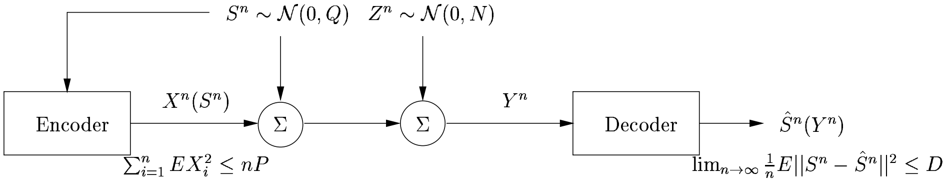
$S^n \sim \mathcal{N}(0, Q)$
$Z^n \sim \mathcal{N}(0, N)$
Encoder
$X^n(S^n)$
Σ
Σ
Y^n
Decoder
$\hat{S}^n(Y^n)$
$\sum_{i=1}^n EX_i^2 \leq nP$
$\lim_{n\to\infty} \frac{1}{n} E||S^n - \hat{S}^n||^2 \leq D$

Figure 3: State information transmission over a state-dependent additive Gaussian channel with states known at the transmitter.

As a motivation, consider the following problem of signal enhancement in the presence of noise. A Gaussian signal $S^n = (S_1, S_2, \ldots, S_n)$, corrupted by an i.i.d. additive Gaussian noise Z^n, is to be reconstructed by a third party based on the (correlated) observation sequence $Y^n = (Y_1, Y_2, \ldots, Y_n)$. An informed party who has a precise knowledge of the signal S^n attempts to help enhance the detection and reconstruction of the signal S^n by sending a signal $X^n(S^n)$, subject to a power constraint. An estimate $\hat{S}^n(Y^n)$ of the signal S^n is formed, subject to a mean-squared error criterion, based on the observation $Y^n = X^n(S^n) + S^n + Z^n$. Natural questions are: (i) what is the optimal enhancing strategy that the informed party should employ? and (ii) what is the corresponding minimum mean-squared estimation error? By recognizing the signal S^n as the channel state, we immediately see that this problem can be analyzed using the framework shown in Fig. 3.

We show that a simple technique of direct state amplification is indeed the optimal state information transmission technique for this setup. In particular, the transmitter sends $X^n = \sqrt{\frac{P}{Q}}S^n$, which helps coherently amplify the presence of the state S^n in the channel. The receiver simply forms an estimate $\hat{S}^n = \frac{Q+\sqrt{PQ}}{(\sqrt{Q}+\sqrt{P})^2+N}Y^n$. The corresponding mean-squared state estimation error is given by $D = Q\frac{N}{\left(\sqrt{Q}+\sqrt{P}\right)^2+N}$.

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 wish to 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. We capture this scenario using the model shown in Fig. 4. In this setup, the sender wishes to send a pure information index $W \in \{1, 2, \ldots, 2^{nR}\}$ as well as help reveal the channel state to the receiver. Based on the pure information W and the state S^n, the transmitter chooses $X^n(W, S^n)$, subject to power constraint P, and transmits it over the channel. The receiver receives $Y^n = X^n(W, S^n) + S^n + Z^n$, decodes $\hat{W}(Y^n) \in \{1, 2, \ldots, 2^{nR}\}$ and forms an estimate $\hat{S}^n(Y^n)$ of the channel state S^n, according to a mean-squared error criterion.

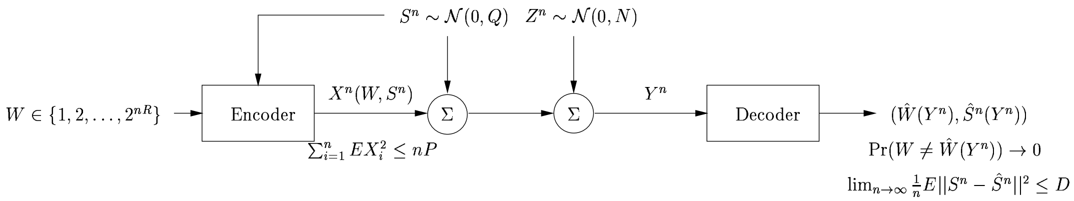
$S^n \sim \mathcal{N}(0,Q)$
$Z^n \sim \mathcal{N}(0,N)$
$W \in \{1,2,\ldots,2^{nR}\}$
Encoder
$X^n(W,S^n)$
$\sum_{i=1}^n EX_i^2 \leq nP$
Σ
Σ
Y^n
Decoder
$(\hat{W}(Y^n), \hat{S}^n(Y^n))$
$\Pr(W \neq \hat{W}(Y^n)) \to 0$
$\lim_{n\to\infty} \frac{1}{n}E\|S^n - \hat{S}^n\|^2 \leq D$

Figure 4: Pure information and state information transmission over a state-dependent additive Gaussian channel with states known at the transmitter.

Naturally, there is a conflict 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. This inherent tradeoff between pure information transmission and state information transmission is what we wish to characterize. In particular, we will characterize the optimal tradeoff between the amount of pure information R that can be reliably communicated and the resulting mean-squared channel state estimation error D. We show that an optimal (R, D) tradeoff pair can be achieved via a power-sharing technique, whereby the transmitter power is appropriately allocated between pure information transmission and state amplification.

This paper is organized as follows. In Section 2, we establish the minimum mean-squared state estimation error corresponding to estimating the channel state of a state-dependent additive Gaussian channel shown in Fig. 3. In Section 3, we characterize the optimal tradeoff between the pure information rate R that can be reliably communicated and the corresponding mean-squared state estimation error D for the setup shown in Fig. 4. We then provide a numerical example in Section 4 and end the paper with concluding remarks in Section 5.

2 Minimum Mean-Squared State Estimation Error

In this section, we establish the minimum mean-squared state estimation error corresponding to estimating the channel state of a state-dependent additive Gaussian channel shown in Fig. 3. Based on the channel state S^n, the transmitter chooses $X^n(S^n)$, subject to a power constraint $\frac{1}{n}\sum_{i=1}^n EX_i^2 \le P$, and sends it. Upon receiving the channel output $Y^n = X^n(S^n) + S^n + Z^n$, the receiver forms an estimate $\hat{S}^n \in \hat{\mathcal{S}}^n$ of the channel state. More formally, a code consists of an encoder map

$$X^n : \mathcal{S}^n \to \mathcal{X}^n,$$

satisfying a power constraint $\frac{1}{n}\sum_{i=1}^n EX_i^2 \le P$ and a decoder map

$$\hat{S}^n : \mathcal{Y}^n \to \hat{\mathcal{S}}^n$$

for estimating the channel state. The mean-squared state estimation error is given by $\frac{1}{n}E||S^n - \hat{S}^n||^2 = \frac{1}{n}\sum_{i=1}^n E(S_i - \hat{S}_i)^2$. An estimation error D is said to be *achievable* for a mean-squared error distortion if there exists a code such that $\lim_{n\to\infty}\frac{1}{n}E||S^n - \hat{S}^n||^2 \le D$.

Theorem 1 *Consider a state-dependent additive Gaussian channel $Y^n = X^n(W, S^n) + S^n + Z^n$, with noncausal state information $S^n = (S_1, S_2, \ldots, S_n)$ at the transmitter with S_i i.i.d. $\sim \mathcal{N}(0, Q)$, unknown independent noise $Z^n = (Z_1, Z_2, \ldots, Z_n)$ with Z_i i.i.d. $\sim \mathcal{N}(0, N)$, and the transmitter power constraint $\frac{1}{n}\sum_{i=1}^n EX_i^2(W, S^n) \le P$. The minimum mean-squared estimation error of the state S^n at the receiver is given by*

$$D = Q\frac{N}{\left(\sqrt{Q}+\sqrt{P}\right)^2 + N}. \tag{1}$$

2.1 Proof of the Achievability

Based on the channel state S^n, the transmitter sends $X^n = \sqrt{\frac{P}{Q}}S^n$, i.e., a scaled version of the channel state S^n. Upon receiving the channel output $Y^n = X^n(S^n) + S^n + Z^n = \left(1 + \sqrt{\frac{P}{Q}}\right)S^n + Z^n$, the receiver forms an estimate $\hat{S}^n = \frac{Q+\sqrt{PQ}}{(\sqrt{Q}+\sqrt{P})^2+N}Y^n$ (i.e., the best linear estimate of the state S^n given the output Y^n). The corresponding mean-squared state estimation error is given by

$$\begin{aligned} D &= \lim_{n\to\infty}\frac{1}{n}E||S^n - \hat{S}^n||^2 \\ &= Q\frac{N}{\left(\sqrt{Q}+\sqrt{P}\right)^2 + N}. \end{aligned}$$

This completes the proof of the achievability.

2.2 Proof of the Converse

In proving the converse of Theorem 1, we must show that given any sequence of codes $\left(X^n(\cdot), \hat{S}^n(\cdot)\right)$ with $\lim_{n\to\infty} \frac{1}{n}E||S^n - \hat{S}^n||^2 \leq D$, the distortion D must satisfy

$$D \geq Q\frac{N}{\left(\sqrt{Q}+\sqrt{P}\right)^2 + N}.$$

We first define $D_n = \frac{1}{n}E||S^n - \hat{S}^n||^2$ and recognize

$$\frac{1}{2}\log\left(\frac{Q}{D_n}\right) \leq \frac{1}{n}I(S^n;Y^n), \tag{2}$$

with equality for i.i.d. jointly Gaussian random variables (Y_i, X_i, S_i, Z_i). eq. (2) can be seen as follows:

$$\begin{aligned}
\frac{1}{n}I(S^n;Y^n) &= \frac{1}{n}(h(S^n) - h(S^n|Y^n)) \\
&\overset{(a)}{=} \frac{1}{n}(h(S^n) - h(S^n - \hat{S}^n(Y^n)|Y^n)) \\
&\overset{(b)}{\geq} \frac{1}{n}(h(S^n) - h(S^n - \hat{S}^n)) \\
&\overset{(c)}{=} \frac{1}{n}\sum_{i=1}^{n} h(S_i) - h(S^n - \hat{S}^n) \\
&\geq \frac{1}{n}\sum_{i=1}^{n}(h(S_i) - h(S_i - \hat{S}_i)) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{2}\log(2\pi eQ) - h(S_i - \hat{S}_i)\right) \\
&\overset{(d)}{\geq} \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{2}\log(2\pi eQ) - \frac{1}{2}\log\left(2\pi eE(S_i - \hat{S}_i)^2\right)\right) \\
&\overset{(e)}{\geq} \frac{1}{2}\log(2\pi eQ) - \frac{1}{2}\log\left(2\pi e\frac{1}{n}\sum_{i=1}^{n}E(S_i - \hat{S}_i)^2\right) \\
&= \frac{1}{2}\log(2\pi eQ) - \frac{1}{2}\log\left(2\pi e\frac{1}{n}E||S^n - \hat{S}^n||^2\right) \\
&= \frac{1}{2}\log\left(\frac{Q}{\frac{1}{n}E||S^n - \hat{S}^n||^2}\right) \\
&= \frac{1}{2}\log\left(\frac{Q}{D_n}\right),
\end{aligned}$$

where
(a) follows from the fact that $\hat{S}^n(Y^n)$ is a function of Y^n,
(b) since conditioning reduces entropy,
(c) from the i.i.d. assumption of the state S^n sequence,
(d) since the Gaussian distribution maximizes the entropy for a given variance, and
(e) follows from Jensen's inequality.

Now we continue the proof of the converse by writing a chain of inequality

$$\begin{aligned}
\frac{1}{2}\log\left(\frac{Q}{D_n}\right) &\leq \frac{1}{n}I(S^n;Y^n)\\
&= \frac{1}{n}\left(h(Y^n)-h(Y^n|S^n)\right)\\
&\overset{(a)}{=} \frac{1}{n}\left(h(Y^n)-h(Y^n|X^n,S^n)\right)\\
&\leq \frac{1}{n}\sum_{i=1}^{n}\Big(h(Y_i)-h(Y_i|Y^{i-1},X^n,S^n)\Big)\\
&\overset{(b)}{=} \frac{1}{n}\sum_{i=1}^{n}\left(h(Y_i)-h(Y_i|X_i,S_i)\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(h(Y_i)-h(Z_i)\right)\\
&\overset{(c)}{\leq} \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\log\left(\frac{EY_i^2}{N}\right)\\
&\overset{(d)}{\leq} \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\log\left(\frac{\left(\sqrt{P_i}+\sqrt{Q}\right)^2+N}{N}\right)\\
&\overset{(e)}{\leq} \frac{1}{2}\log\left(\frac{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}P_i}+\sqrt{Q}\right)^2+N}{N}\right)\\
&\overset{(f)}{\leq} \frac{1}{2}\log\left(\frac{\left(\sqrt{P}+\sqrt{Q}\right)^2+N}{N}\right)
\end{aligned}$$

where
(a) follows from the fact that X^n depends only on S^n,
(b) since the channel is memoryless,
(c) since the Gaussian distribution maximizes the entropy for a given variance,
(d) with equality when $X_i = \sqrt{\frac{P_i}{Q}}S_i$,
(e) follows from Jensen's inequality, and
(f) follows from the imposed power constraint.

The inequality $\frac{1}{2}\log\left(\frac{Q}{D_n}\right) \leq \frac{1}{2}\log\left(\frac{\left(\sqrt{P}+\sqrt{Q}\right)^2+N}{N}\right)$ for all n implies that $D_n \geq Q\frac{N}{\left(\sqrt{Q}+\sqrt{P}\right)^2+N}$ for all n. Hence, the minimum mean-squared state estimation error D is lower-bounded by $Q\frac{N}{\left(\sqrt{Q}+\sqrt{P}\right)^2+N}$. This completes the proof of the converse.

2.3 Discussion

To minimize the mean-squared 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 is, in fact, suboptimal. The channel capacity is $C = \frac{1}{2}\log\left(1+\frac{P}{N}\right)$ as obtained by Costa in [4]. When using the channel in such a way that the pure information rate is maximized, the initial state estimation error (from directly observing the channel output) is given by $Q\frac{P+N}{Q+P+N}$. This is a direct consequence of X^n being statistically uncorrelated with the channel state S^n as observed by Costa in [4]. From the rate-distortion theory, the uncertainty can be further reduced by a factor of 2^{2C} [5, Section 13.3.2]. The resulting mean-squared state estimation error is then given by $Q\frac{P+N}{Q+P+N}2^{-2C} = Q\frac{P+N}{Q+P+N}\frac{N}{P+N} = Q\frac{N}{Q+P+N} \geq Q\frac{N}{(\sqrt{Q}+\sqrt{P})^2+N}$, which is clearly suboptimal.

Instead, the transmitter should use all its power to directly amplify the channel state S^n (i.e., by sending $X^n = \sqrt{\frac{P}{Q}}S^n$). At the receiving end, the receiver simply sets the estimate as $\hat{S}^n = \frac{Q+\sqrt{PQ}}{\left(\sqrt{Q}+\sqrt{P}\right)^2+N}Y^n$. The resulting mean-squared state estimation error is given by $Q\frac{N}{(\sqrt{Q}+\sqrt{P})^2+N}$. Note that there is no codebook involved in this scheme. Furthermore, the encoding/decoding scheme is straightforward as shown in Fig. 5.

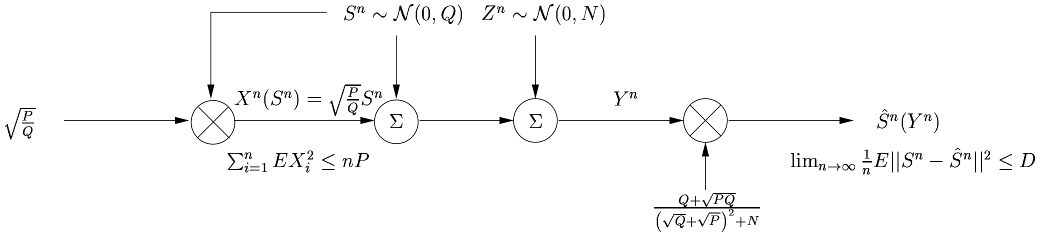
$S^n \sim \mathcal{N}(0,Q)$
$Z^n \sim \mathcal{N}(0,N)$
$\sqrt{\frac{P}{Q}}$
$X^n(S^n) = \sqrt{\frac{P}{Q}}S^n$
Σ
Σ
Y^n
$\hat{S}^n(Y^n)$
$\sum_{i=1}^n EX_i^2 \leq nP$
$\lim_{n\to\infty}\frac{1}{n}E\|S^n - \hat{S}^n\|^2 \leq D$
$\frac{Q+\sqrt{PQ}}{\left(\sqrt{Q}+\sqrt{P}\right)^2+N}$

Figure 5: A state amplification technique.

The optimality of this simple scaling technique is somewhat reminiscent of the technique used in transmitting a Gaussian source over an additive white Gaussian noise channel. More specifically, a Gaussian source $S^n \sim \mathcal{N}(0,Q)$ is to be conveyed (subject to a mean-squared error criterion) over an additive white Gaussian noise channel, $Y^n = X^n + Z^n$, where $Z^n \sim \mathcal{N}(0,N)$ with an input power constraint P. One optimal technique is for the transmitter to first quantize the source using $\frac{1}{2}\log\left(1+\frac{P}{N}\right)$ bits (i.e., the channel capacity) and then send the description over the channel. The resulting mean-squared reconstruction error is given by $Q\frac{P}{P+N}$. Alternatively, as established by Gallager [6], this source can be transmitted uncoded without loss of optimality. Specifically, upon receiving a source symbol S_i at time i the transmitter sends $X_i = \sqrt{\frac{P}{Q}}S_i$, which is merely a scaled version of the source symbol (to meet the transmit power requirement). The receiver simply reconstructs $\hat{S}_i = \frac{\sqrt{QP}}{P+N}Y_i$. The resulting distortion is also given by $Q\frac{P}{P+N}$. This uncoded technique is optimal and can be easily implemented in practice.

3 Optimal (R, D) Region

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 over the channel (Fig. 4). Based on the pure information index $W \in \{1,2,\ldots,2^{nR}\}$ and the channel state S^n, the transmitter chooses $X^n(W,S^n)$, subject to power constraint P, and transmits it over the channel. The receiver receives $Y^n = X^n(W,S^n)+S^n+Z^n$, decodes $\hat{W}(Y^n) \in \{1,2,\ldots,2^{nR}\}$

and forms an estimate $\hat{S}^n(Y^n) \in \hat{\mathcal{S}}^n$ of the channel state S^n, according to a mean-squared error criterion. More formally a $(2^{nR}, n)$ code consists of an encoder map

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

yielding codewords $X^n(1, \cdot), X^n(2, \cdot), \dots, X^n(2^{nR}, \cdot)$ satisfying the expected power constraint P, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} EX_i^2(w, S^n) \le P \qquad w = 1, 2, \dots, 2^{nR},$$

and decoder maps

$$\begin{aligned} \hat{W} &: \mathcal{Y}^n \to \{1, 2, \dots, 2^{nR}\}, \\ \hat{S}^n &: \mathcal{Y}^n \to \hat{\mathcal{S}}^n \end{aligned}$$

for pure information decoding and state estimation.

The probability of a message decoding error and the mean-squared state estimation error are given by $P_e^{(n)} = \frac{1}{2^{nR}} \sum_{i=1}^{2^{nR}} \Pr(\hat{W} \neq i | W = i)$ and $Ed(S^n, \hat{S}^n) = \frac{1}{n} E||S^n - \hat{S}^n(Y^n)||^2$, respectively. An (R, D) pair is said to be *achievable* if there exists a sequence of $(2^{nR}, n)$ codes such that the mean-squared state estimation error $\lim_{n\to\infty} \frac{1}{n} E||S^n - \hat{S}^n(Y^n)||^2 \le D$ and the probability of error $P_e^{(n)} = \frac{1}{2^{nR}} \sum_{i=1}^{2^{nR}} \Pr(\hat{W}(Y^n) \neq i | W = i) \to 0$, where W is uniformly distributed over $\{1, 2, \dots, 2^{nR}\}$ for each block length n. We wish to characterize the optimal (R, D) tradeoff region, which is given by the closure of the convex hull of all achievable (R, D) pairs.

Theorem 2 *Consider a state-dependent additive Gaussian channel* $Y^n = X^n(W, S^n) + S^n + Z^n$, *with noncausal state information* $S^n = (S_1, S_2, \dots, S_n)$ *at the transmitter with* S_i *i.i.d.* $\sim \mathcal{N}(0, Q)$, *unknown independent noise* $Z^n = (Z_1, Z_2, \dots, Z_n)$ *with* Z_i *i.i.d.* $\sim \mathcal{N}(0, N)$, *and the transmitter power constraint* $\frac{1}{n}\sum_{i=1}^{n} EX_i^2(W, S^n) \le P$. *The optimal* (R, D) *tradeoff region for this channel is given by the closure of the convex hull of all* (R, D) *pairs satisfying*

$$R \le \frac{1}{2}\log\left(1 + \frac{\gamma P}{N}\right) \tag{3}$$

$$D \ge Q\frac{(\gamma P + N)}{\left(\sqrt{Q} + \sqrt{(1-\gamma)P}\right)^2 + \gamma P + N} \tag{4}$$

for some $0 \le \gamma \le 1$. *Equivalently, the optimal* (R, D) *tradeoff pairs are given by*

$$(R, D) = \left(\frac{1}{2}\log\left(1 + \frac{\gamma P}{N}\right), Q\frac{(\gamma P + N)}{\left(\sqrt{Q} + \sqrt{(1-\gamma)P}\right)^2 + \gamma P + N}\right), \quad 0 \le \gamma \le 1. \tag{5}$$

3.1 Proof of the Achievability

The achievability proof of Theorem 2 is based on the idea of power-sharing, whereby the transmitter power P is allocated between pure information transmission and state amplification.

Intuitively, the scheme works as follows. Fix a power allocation parameter $0 \leq \gamma \leq 1$ and divide the transmitter power into γP and $(1-\gamma)P$. Based on the state S^n, generate a state-amplification signal $X_s^n = \sqrt{\frac{(1-\gamma)P}{Q}} S^n$ (note that this consumes power $(1-\gamma)P$). This signal X_s^n will be used to directly amplify the existing state S^n, thereby effectively changing the power of the state from Q to $\left(\sqrt{Q}+\sqrt{(1-\gamma)P}\right)^2$. To send pure information $W \in \{1, 2, \ldots, 2^{nR}\}$, apply a similar technique used by Costa in [4] with the state i.i.d. $\sim \mathcal{N}\left(0, \left(\sqrt{Q}+\sqrt{(1-\gamma)P}\right)^2\right)$, unknown noise Z^n i.i.d. $\sim \mathcal{N}(0, N)$, and the transmitter power γP. Call the signal carrying pure information X_w^n. Then send $X_w^n + X_s^n$ over the channel. The received signal is $Y^n = X_w^n + X_s^n + S^n + Z^n = X_w^n + \left(1 + \sqrt{\frac{(1-\gamma)P}{Q}}\right) S^n + Z^n$.

Using this technique, pure information can be transmitted at the rate $R = \frac{1}{2}\log\left(1 + \frac{\gamma P}{N}\right)$ bits, which is achievable as shown by Costa in [4]. The receiver forms an estimate $\hat{S}^n(Y^n) = \frac{Q+\sqrt{(1-\gamma)PQ}}{\left(\sqrt{Q}+\sqrt{(1-\gamma)P}\right)^2 + \gamma P + N} Y^n$. The resulting mean-squared channel state estimation error D is given by $Q\frac{(\gamma P + N)}{\left(\sqrt{Q}+\sqrt{(1-\gamma)P}\right)^2 + \gamma P + N}$. By varying the power allocation parameter $0 \leq \gamma \leq 1$, we are able to trade off between the amount of pure information that can be reliably transmitted and the state estimation error.

3.2 Proof of the Converse

In proving the converse of Theorem 2, we must show that given any sequence of $(2^{nR}, n)$ codes with $P_e^{(n)} \to 0$ and $\lim_{n\to\infty} \frac{1}{n} E||S^n - \hat{S}^n||^2 \leq D$, then R and D must satisfy

$$\begin{aligned} R &\leq \frac{1}{2}\log\left(1 + \frac{\gamma P}{N}\right) \\ D &\geq Q\frac{(\gamma P + N)}{\left(\sqrt{Q}+\sqrt{(1-\gamma)P}\right)^2 + \gamma P + N} \end{aligned}$$

for some $0 \leq \gamma \leq 1$. To this end, define

$$\begin{aligned} R(\gamma) &= \frac{1}{2}\log\left(1 + \frac{\gamma P}{N}\right), \text{ and} \\ D(\gamma) &= Q\frac{(\gamma P + N)}{\left(\sqrt{Q}+\sqrt{(1-\gamma)P}\right)^2 + \gamma P + N}. \end{aligned}$$

Note that the $(R(\gamma), D(\gamma))$ pairs, $0 \leq \gamma \leq 1$, are the Pareto optimal tradeoff pairs of the (R, D) region stated in Theorem 2. Furthermore $R(\gamma)$ and $D(\gamma)$ are monotonic and strictly concave functions in $0 \leq \gamma \leq 1$. Hence, we can equivalently establish the converse of Theorem 2 by showing that given any sequence of $(2^{nR}, n)$ codes with $P_e^{(n)} \to 0$ and $\lim_{n\to\infty} \frac{1}{n} E||S^n - \hat{S}^n||^2 \leq D$, then R and D must satisfy, for all $\mu \geq 0$,

$$R + \mu\frac{1}{2}\log\left(\frac{Q}{D}\right) \leq R(\gamma_\mu) + \mu\frac{1}{2}\log\left(\frac{Q}{D(\gamma_\mu)}\right),$$

where $0 \leq \gamma_\mu \leq 1$ is chosen to maximize $R(\gamma) + \mu \frac{1}{2} \log\left(\frac{Q}{D(\gamma)}\right)$ for a given value of μ.

Recall from Equation (2) in Section 2 that $\frac{1}{2}\log\left(\frac{Q}{D_n}\right) \leq \frac{1}{n} I(S^n; Y^n)$, where $D_n = \frac{1}{n} E||S^n - \hat{S}^n||^2$. Thus, for $0 \leq \mu \leq 1$, we can bound the weighted sum $R + \mu \frac{1}{2}\log\left(\frac{Q}{D_n}\right)$ as follows:

$$\begin{aligned}
R + \mu \frac{1}{2} \log\left(\frac{Q}{D_n}\right) &\leq R + \frac{\mu}{n} I(S^n; Y^n) \\
&= \mu(R + \frac{1}{n} I(S^n; Y^n)) + (1-\mu) R \\
&\overset{(a)}{\leq} \frac{\mu}{n}\left(H(W) + I(S^n; Y^n)\right) + \frac{(1-\mu)}{n} H(W) \\
&\overset{(b)}{=} \frac{\mu}{n}\left(H(W|S^n) + I(S^n; Y^n)\right) + \frac{(1-\mu)}{n} H(W|S^n) \\
&\overset{(c)}{\leq} \frac{\mu}{n}\left(I(W; Y^n|S^n) + I(S^n; Y^n)\right) + \frac{(1-\mu)}{n} I(W; Y^n|S^n) + \epsilon_n \\
&\overset{(d)}{\leq} \frac{\mu}{n}\left(I(X^n; Y^n|S^n) + I(S^n; Y^n)\right) + \frac{(1-\mu)}{n} I(X^n; Y^n|S^n) + \epsilon_n \\
&= \frac{\mu}{n} I(X^n, S^n; Y^n) + \frac{(1-\mu)}{n} I(X^n; Y^n|S^n) + \epsilon_n \\
&= \frac{\mu}{n}\left(h(Y^n) - h(Y^n|X^n, S^n)\right) + \frac{(1-\mu)}{n}\left(h(Y^n|S^n) - h(Y^n|X^n, S^n)\right) + \epsilon_n \\
&\overset{(f)}{\leq} \frac{\mu}{n}\sum_{i=1}^{n}(h(Y_i) - h(Y_i|Y^{i-1}, X^n, S^n)) + \frac{(1-\mu)}{n}\sum_{i=1}^{n}(h(Y_i|S_i) - h(Y_i|Y^{i-1}, X^n, S^n)) + \epsilon_n \\
&\overset{(g)}{=} \frac{\mu}{n}\sum_{i=1}^{n}\left(h(Y_i) - h(Y_i|X_i, S_i)\right) + \frac{(1-\mu)}{n}\sum_{i=1}^{n}\left(h(Y_i|S_i) - h(Y_i|X_i, S_i)\right) + \epsilon_n \\
&= \frac{\mu}{n}\sum_{i=1}^{n}\left(h(Y_i) - h(Z_i)\right) + \frac{(1-\mu)}{n}\sum_{i=1}^{n}\left(h(Y_i|S_i) - h(Z_i)\right) + \epsilon_n \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\mu h(Y_i) + (1-\mu) h(Y_i|S_i) - h(Z_i)\right) + \epsilon_n \\
&\overset{(h)}{\leq} \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\log\left(\frac{\left(\left(\sqrt{Q} + \sqrt{(1-\gamma)P_i}\right)^2 + \gamma P_i + N\right)^{\mu}(\gamma P_i + N)^{1-\mu}}{N}\right) + \epsilon_n \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\log\left(\frac{\left(\sqrt{Q} + \sqrt{(1-\gamma)P_i}\right)^2 + \gamma P_i + N}{N}\right)^{\mu} + \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\log\left(\frac{\gamma P_i + N}{N}\right)^{1-\mu} + \epsilon_n \\
&\overset{(i)}{\leq} \frac{1}{2}\log\left(\frac{\left(\sqrt{Q} + \sqrt{(1-\gamma)P}\right)^2 + \gamma P + N}{N}\right)^{\mu} + \frac{1}{2}\log\left(\frac{\gamma P + N}{N}\right)^{1-\mu} + \epsilon_n \\
&= \frac{1}{2}\log\left(1 + \frac{\gamma P}{N}\right) + \frac{\mu}{2}\log\left(\frac{\left(\sqrt{Q} + \sqrt{(1-\gamma)P}\right)^2 + \gamma P + N}{(\gamma P + N)}\right) + \epsilon_n \qquad (6) \\
&\overset{(j)}{\leq} \frac{1}{2}\log\left(1 + \frac{\gamma_\mu P}{N}\right) + \frac{\mu}{2}\log\left(\frac{\left(\sqrt{Q} + \sqrt{(1-\gamma_\mu)P}\right)^2 + \gamma_\mu P + N}{(\gamma_\mu P + N)}\right) + \epsilon_n
\end{aligned}$$

$$\overset{(k)}{=} \quad R(\gamma_\mu) + \mu\frac{1}{2}\log\left(\frac{Q}{D(\gamma_\mu)}\right) + \epsilon_n,$$

where
(a) with equality when W is uniform over $\{1, 2, \ldots, 2^{nR}\}$,
(b) follows from the fact that W and S^n are independent,
(c) from Fano's inequality,
(d) from the data processing inequality,
(f) from the chain rule and the fact that conditioning reduces entropy,
(g) since the channel is a discrete memoryless channel,
(h) with $\{X_i\}$ Gaussian (see Lemma 1 in the Appendix); then represent $X_i = V_i + \sqrt{(1-\gamma)\frac{P_i}{Q}}S_i$, where $V_i \sim \mathcal{N}(0, \gamma P_i)$ independent of S_i, with $0 \le \gamma \le 1$ chosen such that the resulting covariance of (Y_i, X_i, S_i, Z_i) is the same as that induced by the code,
(i) follows from Jensen's inequality and the power constraint requirement,
(j) where γ_μ is chosen to maximize the expression in eq. (6) for a fixed μ, and
(k) follows from the definitions of $R(\gamma)$ and $D(\gamma)$.

Because the expression in eq. (6) is a strictly concave function, and we are maximizing over a convex constraint set of γ, there is a unique optimal γ_μ for a given μ. In short, for all $0 \le \mu \le 1$,

$$R + \mu\frac{1}{2}\log\left(\frac{Q}{D_n}\right) \quad \le \quad R(\gamma_\mu) + \mu\frac{1}{2}\log\left(\frac{Q}{D(\gamma_\mu)}\right) + \epsilon_n, \tag{7}$$

where $0 \le \gamma_\mu \le 1$ is chosen to maximize the expression in eq. (6) for a given $0 \le \mu \le 1$. Furthermore, it is straightforward to see that for all $\mu \ge 1$, $\gamma = 0$ maximizes the expression in eq. (6).

As a result, in the limit as $\epsilon_n \to 0$ and $\lim_{n\to\infty} D_n = \lim_{n\to\infty}\frac{1}{n}E||S^n - \hat{S}^n||^2 \le D$, for all $\mu \ge 0$,

$$\begin{aligned} R + \mu\frac{1}{2}\log\left(\frac{Q}{D}\right) &\le \lim_{n\to\infty}\left(R + \mu\frac{1}{2}\log\left(\frac{Q}{D_n}\right)\right), \\ &\overset{(a)}{\le} \lim_{n\to\infty}\left(R(\gamma_\mu) + \mu\frac{1}{2}\log\left(\frac{Q}{D(\gamma_\mu)}\right) + \epsilon_n\right), \\ &= R(\gamma_\mu) + \mu\frac{1}{2}\log\left(\frac{Q}{D(\gamma_\mu)}\right), \end{aligned}$$

where (a) follows from eq. (7). This establishes the converse of Theorem 2.

3.3 Discussion

As given in Theorem 2, the optimal (R, D) tradeoff pairs are given by

$$(R, D) = \left(\frac{1}{2}\log\left(1 + \frac{\gamma P}{N}\right), Q\frac{(\gamma P + N)}{\left(\sqrt{Q} + \sqrt{(1-\gamma)P}\right)^2 + \gamma P + N}\right), \quad 0 \le \gamma \le 1.$$

By varying the power allocation parameter $0 \leq \gamma \leq 1$, we are able to trade off between the pure information rate R and the mean-squared estimation error D. In particular, $\gamma = 0$ corresponds to the scenario where the transmitter uses the entire power budget to amplify the channel state, leaving no resources left for pure information transmission. The corresponding optimal (R, D) tradeoff pair is given by

$$(R, D) = \left(0, Q\frac{N}{\left(\sqrt{Q} + \sqrt{P}\right)^2 + N}\right).$$

On the other hand, $\gamma = 1$ corresponds to the scenario where the transmitter wishes to send only pure information while ignoring the state estimation error. The optimal (R, D) tradeoff pair is given by

$$(R, D) = \left(\frac{1}{2}\log\left(1 + \frac{P}{N}\right), Q\frac{P + N}{Q + P + N}\right).$$

The resulting mean-squared state estimation error $D = Q\frac{P+N}{Q+P+N} \leq Q$, which suggests that the receiver is able to learn something about the channel state on its own even though the transmitter does not attempt to help convey any state information.

There is an intriguing relationship between the transmitted signal X^n and the state S^n associated with each point on the optimal tradeoff curve. In particular, a different point on the curve reflects a different degree of correlation between the transmitted signal X^n and the state S^n. As discussed in Section 2, when the goal is only to minimize the mean-squared state estimation error, the transmitter directly amplifies the state by sending the signal X^n in the direction of the state S^n; i.e., the signal X^n is chosen to be completely correlated with the state S^n as shown in Fig. 6a. Furthermore, as observed by Costa in [4], when the goal is only to maximize the pure information rate, the transmitter uses the signal X^n to nudge the state S^n, in the direction of the desired codeword. Interestingly, at this operating point, the resulting X^n is statistically uncorrelated with the state S^n as shown in Fig. 6b. To achieve a particular (R, D) tradeoff pair on the boundary of the optimal tradeoff region, the transmitter employs a power-sharing technique, whereby the transmitter power is appropriately allocated between pure information transmission and state amplification. A different (R, D) tradeoff pair on the boundary reflects a different degree of correlation between the transmitted signal X^n and the state S^n (Fig. 6c).

4 Numerical Example

A specific numerical example is given in this section. Consider an additive Gaussian channel $Y^n = X^n(W, S^n) + S^n + Z^n$, with S_i i.i.d. $\sim \mathcal{N}(0, 1)$, noise Z_i i.i.d. $\sim \mathcal{N}(0, 1)$, and transmitter power constraint $P = 1$. The optimal (R, D) tradeoff region is shown in Fig. 7.

Consider first the scenario where the transmitter wishes to help the receiver minimize the channel state estimation error. In this case, the transmitter uses all its power to amplify the state S^n by transmitting $X^n = \sqrt{\frac{P}{Q}}S^n = S^n$. The corresponding mean-squared state estimation

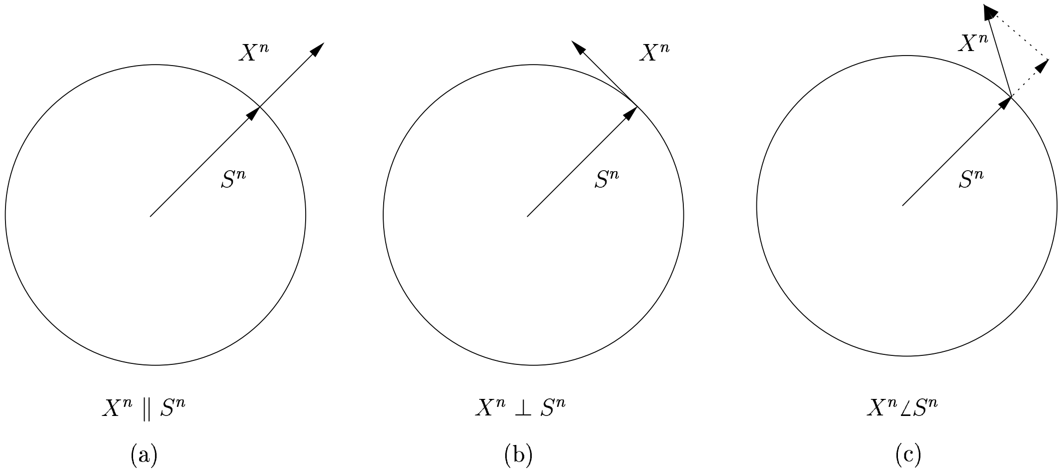

Figure 6: A diagram relating the transmitted signal X^n and the channel state S^n. (a) Pure state amplification, (b) pure information transmission, and (c) combination of state amplification and information transmission.

error is given by $Q\frac{N}{\left(\sqrt{P}+\sqrt{Q}\right)^2+N} = \frac{1}{5}$. Furthermore, since the transmitter power is used entirely to amplify the channel state, no pure information can be conveyed. Note that if the transmitter attempts to maximize the pure information rate and then use it in refining the receiver's initial estimate, then the resulting mean-squared estimation error is given by $Q\frac{N}{Q+P+N} = \frac{1}{3}$, which is clearly suboptimal.

On the other hand, when the transmitter's goal is to transmit only pure information, by applying Costa's transmission technique [4], pure information can be transmitted at the rate given by $\frac{1}{2}\log(1+\frac{P}{N}) = \frac{1}{2}$ bits. Interestingly, due to the state-dependent nature of the channel, the receiver is able to learn something about the state from observing the channel output. Specifically, the receiver is able to reduce the uncertainty about the state from $Q = 1$ to $Q\frac{P+N}{Q+P+N} = \frac{2}{3}$.

A point on the boundary of the tradeoff region is obtained by varying the amount of power used in transmitting pure information and amplifying the state. All (R, D) pairs above the tradeoff curve in Fig. 7 are achievable. As a comparison, a tradeoff region based on a time-sharing technique is shown.

5 Concluding Remarks

We characterize the optimal tradeoff between pure information transmission and channel state estimation for an additive Gaussian channel with state information at the sender. When the goal is to minimize the state estimation error at the receiver, the optimal transmission technique is to use all sender's power to amplify the channel state. Interestingly, because of the state-dependent nature of the channel, state information transmission is more than merely using the channel capacity to carry the channel state description. On the other hand, when the goal is

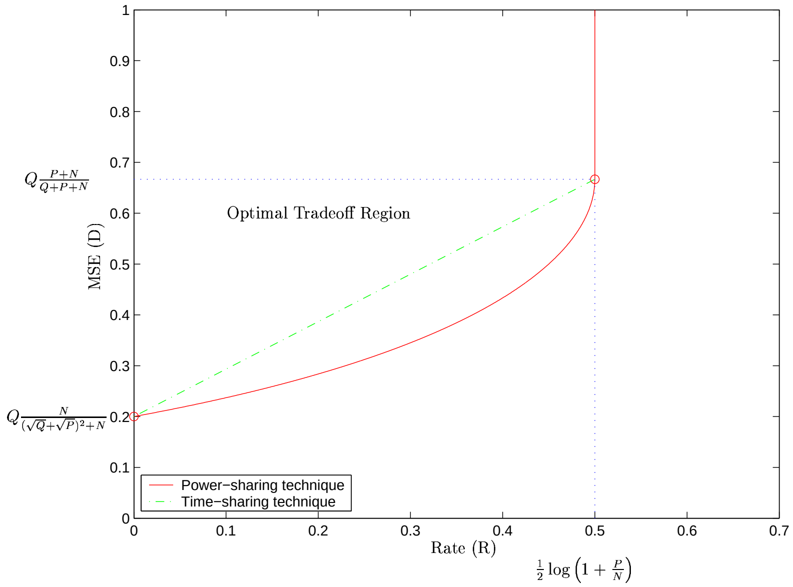

Figure 7: Optimal (R, D) tradeoff region for a state-dependent additive Gaussian channel with states known at the transmitter.

to transmit only pure information, the sender can use the 'writing on dirty paper' technique [4]. Pure information transmission, however, obscures the receiver's view of the channel state, thereby increasing the state estimation error. For this intrinsic conflict, we show that a simple power-sharing technique achieves the optimal tradeoff.

Appendix

Lemma 1 *Let $Y = X + S + Z$ where S and Z are independent zero-mean Gaussian random variables and X is an arbitrary zero-mean random variable correlated with S and Z, with a fixed covariance matrix K_{XSZ}. Then, for any $0 \le \mu \le 1$,*

$$\mu\, h(Y) + (1-\mu)\, h(Y|S) \le \frac{\mu}{2}\log\left(2\pi e\, EY^2\right) + \frac{1-\mu}{2}\log\left(2\pi e\left(EY^2 - \frac{(ESY)^2}{ES^2}\right)\right),$$

with equality when X is jointly Gaussian with S and Z.

Remark: Lemma 1 states that, for a given choice of the input X, there exists a Gaussian input $\hat{X}$ with the same covariance K_{XSZ}, which dominates X in $\mu h(Y) + (1-\mu)h(Y|S)$.

Proof of Lemma 1: For a fixed $0 \le \mu \le 1$, we have the following chain of inequalities:

$$\begin{aligned}\mu h(Y) + (1-\mu)h(Y|S) &\overset{(a)}{\le} \frac{\mu}{2}\log\left(2\pi e\, EY^2\right) + (1-\mu)h(Y|S)\\ &\overset{(b)}{=} \frac{\mu}{2}\log\left(2\pi e\, EY^2\right) + (1-\mu)\int h(Y|S=s)p(s)ds\end{aligned}$$

$$
\begin{aligned}
&\overset{(c)}{\leq} \frac{\mu}{2}\log\left(2\pi e\, EY^2\right) + (1-\mu)\int \frac{1}{2}\log\left(2\pi e \mathrm{Var}(Y|S=s)\right)p(s)ds \\
&\overset{(d)}{\leq} \frac{\mu}{2}\log\left(2\pi e\, EY^2\right) + \frac{1-\mu}{2}\log\left(2\pi e\left(\int \mathrm{Var}(Y|S=s)p(s)ds\right)\right) \\
&= \frac{\mu}{2}\log\left(2\pi e\, EY^2\right) + \frac{1-\mu}{2}\log\left(2\pi e E(\mathrm{Var}(Y|S))\right) \\
&\overset{(e)}{=} \frac{\mu}{2}\log\left(2\pi e\, EY^2\right) + \frac{1-\mu}{2}\log\left(2\pi e E\left(Y-E(Y|S)\right)^2\right) \\
&\overset{(f)}{\leq} \frac{\mu}{2}\log\left(2\pi e\, EY^2\right) + \frac{1-\mu}{2}\log\left(2\pi e\left(EY^2 - \frac{(ESY)^2}{ES^2}\right)\right),
\end{aligned}
$$

where
(a) since the Gaussian distribution maximizes the entropy for a given variance,
(b) from the definition of conditional entropy,
(c) since the Gaussian conditional distribution maximizes the conditional entropy,
(d) from Jensen's inequality,
(e) from the definition of a conditional variance, and
(f) since the linear MMSE error is larger than the MMSE error.

Now consider a zero-mean random variable X jointly Gaussian with S and Z with the same covariance matrix K_{XSZ}. It follows that every inequality above becomes an equality; specifically,
(a) since $Y = X + S + Z$ is Gaussian,
(c) since Y is conditionally Gaussian given S,
(d) since the conditional variance $\mathrm{Var}(Y|S=s)$ is a constant for each s, and
(f) since the MMSE estimate is linear for jointly Gaussian random variables Y and S.

References

[1] B. Chen and G. W. Wornell, “Quantization index modulation: a class of provably good methods for digital watermarking and information embedding,” *IEEE Transactions on Information Theory,* IT-47:1423–1443, May 2001.

[2] M. Chiang, A. Sutivong, and T. M. Cover, “Channel capacity and state estimation,” *Proceedings of International Symposium on Information Theory and Its Applications*, Honolulu, Hawaii, November 2000, pp. 838–840.

[3] A. Cohen and A. Lapidoth, “The Gaussian watermarking game,” *IEEE Transactions on Information Theory,* IT-48:1639-1667, June 2002.

[4] M. H. M. Costa, “Writing on dirty paper,” *IEEE Trans. Inform. Theory,* IT-29:439–411, May 1983.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory,* New York: Wiley, 1991.

[6] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[7] S. I. Gel’fand and M. S. Pinsker, “Coding for channel with random parameters,” *Problems of Control and Information Theory,* vol. 9, no. 1 pp. 19–31, 1980.

[8] C. Heegard and A. El Gamal, “On the capacity of computer memories with defects,” *IEEE Transactions on Information Theory,* IT-29:731–739, September 1983.

[9] A. V. Kusnetsov and B. S. Tsybakov, “Coding in a memory with defective cells,” translated from *Prob. Peredach. Inform.,* vol. 10, no. 2, pp. 52–60, April-June 1974.

[10] J. A. O’Sullivan and P. Moulin, “Information-theoretic analysis of information hiding,” preprint, available at http://www.ifp.uiuc.edu/ moulin/paper.html, 2001.

[11] H. C. Papadopoulos and C.-E. W. Sundberg, “Simultaneous broadcasting of analog FM and digital audio signals by means of adaptive precanceling techniques,” *IEEE Transactions on Communications,* vol. 46, pp. 1233–1242, September 1998.

[12] C. E. Shannon, “Channels with side information at the transmitter,” *IBM Journal of Research and Development,* vol. 2, pp. 289–293, 1958.

[13] A. Sutivong, T. M. Cover, and M. Chiang, “Trade-off between message and state information rates,” *Proceedings of International Symposium on Information Theory*, Washington D.C., June 2001, p. 303.