[15] S. N. Batalama, M. J. Medley, and I. N. Psaromiligkos, "Adaptive robust spread-spectrum receivers," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 905–917, Jun. 1999.

[16] H. V. Poor and M. Tanda, "Multiuser detection in impulsive channels," *Ann. Telecommun.*, vol. 54, pp. 392–400, Jul./Aug. 1999.

[17] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.

[18] B. Seyfe and A. R. Sharafat, "Non-parametric multiuser detection in non-Gaussian channels," *IEEE Trans. Signal Process.*, to be published.

[19] J. D. Gibson and M. L. Melsa, *Introduction to Non-Parametric Detection With Application*. New York: Academic, 1975.

[20] P. P. Kazakos and D. Kazakos, Eds., *Non-Parametric Methods in Communications*. New York: Marcel Decker, 1977.

[21] S. A. Kassam and J. B. Thomas, "Non-parametric detection, theory and application," in *Benchmark Papers in Electrical Engineering and Computer Science*. New York: Academic, 1980, vol. 23.

[22] T. P. Hettmansperger and J. W. McKean, *Robust Nonparametric Statistical Methods*. New York: Wiley, 1998.

[23] L. A. Jaeckle, "Estimating regression coefficients by minimizing the dispersion of the residuals," *Ann. Math. Statist.*, vol. 43, pp. 1449–1458, 1972.

[24] E. L. Lehmann, *Nonparametric Statistical Methods Based on Ranks*. San Francisco, CA: Holden-Day, 1975.

[25] F. R. Hampel, E. M. Ronchetti, and W. A. Stahl, *Robust Statistics: The Approach Based on Influence Functions*. New York: Wiley, 1986.

[26] X. Wang and H. V. Poor, "Blind multiuser detection: A subspace approach," *IEEE Trans. Info. Theory*, vol. 44, no. 2, pp. 677–690, Mar. 1998.

[27] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Process.*, vol. 44, no. 1, pp. 95–107, Jan. 1995.

[28] ——, "An extension of the PASTd algorithm to both rank and subspace tracking," *IEEE Signal Process. Lett.*, vol. 2, pp. 179–182, Sep. 1995.

[29] ——, "Asymptotic convergence analysis of the projection approximation subspace tracking algorithms," *Signal Process.*, vol. 50, pp. 123–136, 1996.

[30] S. Attallah and K. Abed-Meraim, "Fast algorithms for subspace tracking," *IEEE Signal Process. Lett.*, vol. 8, pp. 203–206, Jul. 2001.

[31] Y. Wen, S. C. Chen, and K. L. Ho, "Robust subspace tracking in impulsive noise," in *Proc. IEEE Int. Conf. Communication (ICC)*, Helsinki, Finland, 2001, pp. 892–896.

[32] T. Gustafsson and C. S. MacInnes, "A class of subspace tracking algorithms based on approximation of noise-subspace," *IEEE Trans. Signal Process.*, vol. 48, no. 11, pp. 3231–3235, Nov. 2000.

[33] K. Abed-Meraim, A. Chekeif, and Y. Hua, "Fast orthonormal PAST algorithm," *IEEE Signal Process. Lett.*, vol. 7, pp. 60–62, Mar. 2000.

Channel Capacity and State Estimation for State-Dependent Gaussian Channels

Arak Sutivong, *Member, IEEE*, Mung Chiang, *Member, IEEE*, Thomas M. Cover, *Fellow, IEEE*, and Young-Han Kim, *Student Member, IEEE*

***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 $E\|S^n - \hat{S}^n\|$ for a state-dependent additive Gaussian channel $Y^n = X^n + S^n + Z^n$ with an independent and identically distributed (i.i.d.) Gaussian state sequence $S^n = (S_1, \ldots, S_n)$ 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 simple power-sharing technique, whereby the transmitter power is appropriately allocated between pure information transmission and state amplification.**

***Index Terms*—Additive Gaussian noise channels, channels with state information, joint source–channel coding, state amplification, state estimation.**

I. 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 [1] in 1958,

Manuscript received October 20, 2002; revised January 12, 2005. The work of A. Sutivong, T. M. Cover, and Y.-H. Kim was supported in part by the National Science Foundation under Grants CCR-9973134, CCR-0311633, by MURI under Grant DAAD-19-99-1-0215, and by the Stanford Networking Research Center (SNRC). The work of M. Chiang was supported by the Hertz Foundation Fellowship and a Stanford Graduate Fellowship. The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory and Its Applications 2000, Honolulu, HI, November 2000 and at the IEEE International Symposium on Information Theory, Washington, DC, June 2001.

A. Sutivong was with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA. He is now with QUALCOMM Inc., San Diego, CA 92121 USA (email: araks@qualcomm.com).

M. Chiang was with the the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA. He is now with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (email: chiangm@princeton.edu).

T. M. Cover is with the Department of Electrical Engineering and the Department of Statistics, Stanford University, Stanford, CA 94305 USA (email: cover@stanford.edu).

Y.-H. Kim is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (email: yhk@stanford.edu).

Communicated by İ.E. Telatar, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2005.844108

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

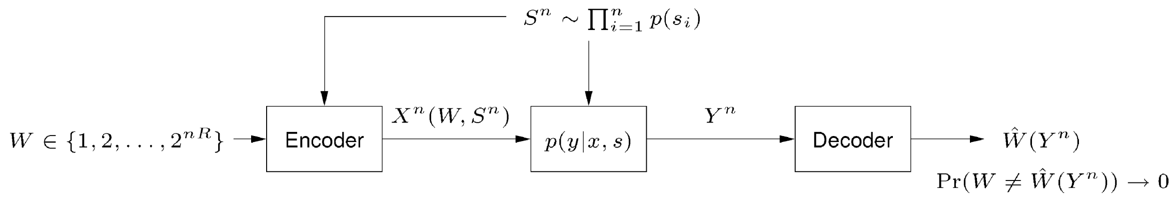

Fig. 1. Pure information transmission over a state-dependent channel with states known at the transmitter.

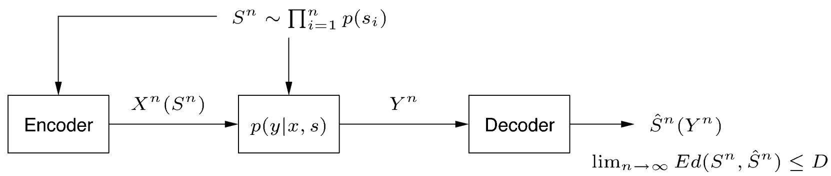

Fig. 2. State information transmission over a state-dependent channel with states known at the transmitter.

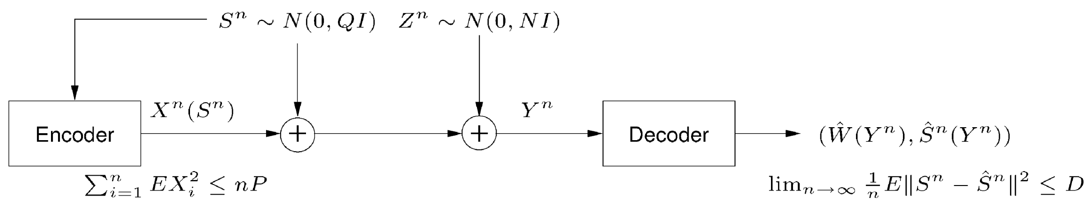

Fig. 3. State information transmission over a state-dependent additive Gaussian channel with states known at the transmitter.

Kusnetsov and Tsybakov [2], Gel'fand and Pinsker [3], and Heegard and El Gamal [4].

Consider a state-dependent channel with state information available only at the sender, as shown Fig. 1. 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 $S^n = (S_1, S_2, \ldots, S_n)$, S_i independent and identically distributed (i.i.d.) $\sim p(s)$, 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 [5], digital watermarking [6], [7], multiple-antenna broadcast, and data storage over memory with defects [2], [4], etc.

Most of the existing literature has focused on determining the channel capacity or devising practical capacity-achieving coding techniques. As shown by Gel'fand and Pinsker [3], and independently by Heegard and El Gamal [4], the capacity of a discrete memoryless state-dependent channel is given by

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

where U is an auxiliary random variable with finite cardinality. In celebrated "writing on dirty paper" [8], Costa considered a memoryless Gaussian state-dependent channel $Y^n = X^n + S^n + Z^n$ and showed that the capacity of the channel is not affected by the presence of the additive state noise S^n as long as the transmitter has full prior knowledge of it. This result has been extended to various setups in [7], [9], [10].

In certain communication scenarios, however, rather than communicating pure information W across the channel, the transmitter may instead wish to help reveal the channel state S^n 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 $\hat{S}^n(Y^n) \in \hat{\mathcal{S}}^n$ of the channel state. 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 $d : \mathcal{S} \times \hat{\mathcal{S}} \to \mathbb{R}$ 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]. Here, digital refinement information is overlayed on top of the existing legacy analog transmission in order to help improve the detection and reconstruction of the original analog signal, which must be kept intact due to backward compatibility requirements. In this example, the existing analog transmission can be viewed as the channel state that the transmitter has access to and wishes to help reveal to the receiver. A key observation here is that the presence of the analog signal affects the channel over which the digital information is transmitted. At the same time, the digital transmission may itself interfere with the existing analog transmission, thereby degrading the quality of the original analog signal—the very thing that the digital information is designed to help improve.

In this correspondence, we study this 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 i.i.d. $\sim N(0, Q)$, 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 an average power constraint P. Upon receiving the output $Y^n = X^n(S^n) + S^n + Z^n$, where $Z_i \sim N(0, N)$, $i = 1, 2, \ldots$, is unknown i.i.d. additive Gaussian noise sequence, the receiver forms an estimate $\hat{S}^n(Y^n)$. The goal is to minimize the mean-squared state estimation error $E\|S^n - \hat{S}^n\|^2$.

As motivation, consider the following problem of signal enhancement in the presence of noise. A Gaussian signal 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

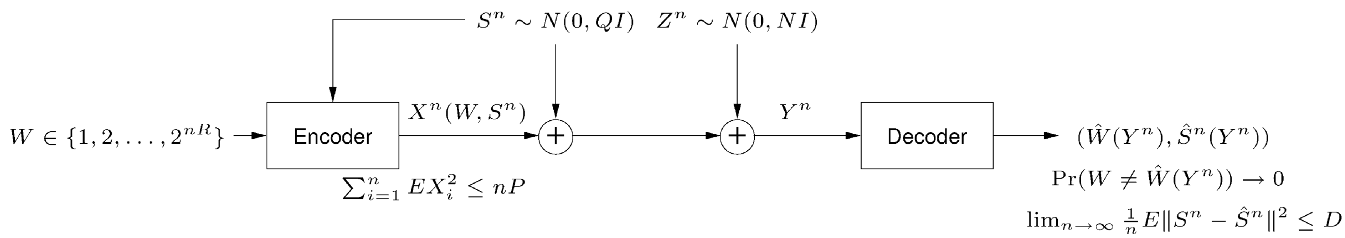

Fig. 4. Pure information and state information transmission over a state-dependent additive Gaussian channel with states known at the transmitter.

$Y^n = (Y_1, Y_2, \ldots, Y_n)$. An informed party who has a precise knowledge of the signal S^n attempts to help enhance the 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, 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 channel model shown in Fig. 3.

We show that a simple technique of direct state amplification is 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 $W \in \{1, 2, \ldots, 2^{nR}\}$ 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. From the channel output $Y^n = X^n(W, S^n) + S^n + Z^n$, the receiver 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.

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 correspondence is organized as follows. In Section II, 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 III, 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 IV and conclude the correspondence in Section V.

II. 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 $Y^n = X^n + S^n + Z^n$, where $S^n \sim N(0, QI)$ is the white Gaussian state, $Z^n \sim N(0, NI)$ is the additive white Gaussian noise, and S^n and Z^n are independent (see Fig. 3). The transmitter has full prior knowledge of the state sequence S^n. Based on the channel state S^n, the transmitter chooses $X^n(S^n)$, subject to a power constraint $\sum_{i=1}^n EX_i^2 \le nP$, and sends it. Upon receiving the channel output Y^n, the receiver forms an estimate $\hat{S}^n$ of the channel state. More formally, an n-block code consists of an encoder map

$$X^n : \mathbb{R}^n \to \mathbb{R}^n$$

yielding codewords $X^n(S^n)$ that satisfy a power constraint

$$\sum_{i=1}^n EX_i^2 \le nP$$

and a decoder map

$$\hat{S}^n : \mathbb{R}^n \to \mathbb{R}^n$$

for estimating the channel state. The mean-squared state estimation error is given by

$$\frac{1}{n} E\|S^n - \hat{S}^n\|^2 = \frac{1}{n} \sum_{i=1}^n E(S_i - \hat{S}_i)^2.$$

We say an estimation error D is *achievable* for a mean-squared error distortion if there exists a sequence of codes $\{(X^n(\cdot), \hat{S}^n(\cdot))\}_{n=1}^\infty$ such that

$$\lim_{n\to\infty} \frac{1}{n} E\|S^n - \hat{S}^n\|^2 \le D.$$

Theorem 1: For the state-dependent additive Gaussian channel $Y^n = X^n(S^n) + S^n + Z^n$, the infimum of achievable estimation errors is given by

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

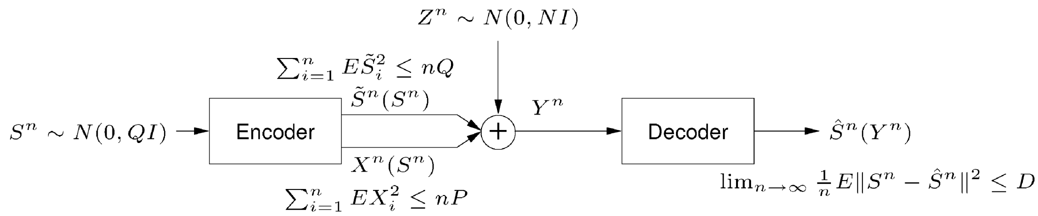

Fig. 5. Relaxed channel model for state information transmission.

Moreover, there exists a sequence of codes $(X^n(\cdot), \hat{S}^n(\cdot))$ that attains D^* for any finite block length n.

A. Proof of 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. Clearly, the power constraint is satisfied. 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 minimum mean-squared error estimate of the state S^n given the output Y^n). The corresponding mean-squared state estimation error is given by

$$\begin{aligned} D &= \frac{1}{n} E\|S^n - \hat{S}^n\|^2 \\ &= E|S_1 - \hat{S}_1|^2 \\ &= Q \frac{N}{\left(\sqrt{Q} + \sqrt{P}\right)^2 + N}. \end{aligned}$$

This result does not depend on the block length n, which proves the last statement of the theorem. □

Incidentally, we can strengthen the expected power constraint $E\sum_i X_i^2 \leq nP$ on the input sequence X^n to a stronger constraint $\sum_i X_i^2 \leq nP$ by the strong law of large numbers and the continuity of the estimation error in the power constraint P, at the expense of losing the finite-block achievability.

B. Proof of the Converse

In proving the converse of Theorem 1, we show that given any sequence of codes $\{(X^n(\cdot), \hat{S}^n(\cdot))\}$, the associated distortion $D_n = \frac{1}{n} E\|S^n - \hat{S}^n\|^2$ satisfies

$$D_n \geq Q \frac{N}{\left(\sqrt{Q} + \sqrt{P}\right)^2 + N}, \qquad \text{for all } n.$$

This can be easily proved by the following argument. Consider the problem of transmitting a source over a memoryless channel as depicted in Fig. 5. Here, the transmitter has access to two inputs X^n and $\tilde{S}^n$ with average power constraints

$$\sum_{i=1}^n X_i^2 \leq nP \quad \text{and} \quad \sum_{i=1}^n \tilde{S}_i^2 \leq nQ$$

respectively. The goal is to communicate the memoryless Gaussian source $S^n \sim N(0, QI)$ over the channel $Y^n = X^n + \tilde{S}^n + Z^n$ where the additive Gaussian noise Z_i i.i.d. $\sim N(0, N)$, $i = 1, 2, \ldots, n$, with minimum possible distortion. This setup clearly subsumes the original problem under our consideration as a special case $\tilde{S}_i = S_i$, $i = 1, 2, \ldots$. Now we use the standard source-channel separation theorem [12] for this relaxed problem. Recall that the capacity of the channel in Fig. 5 is

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

which is achieved by transmitting signals X^n and $\tilde{S}^n$ coherently. On the other hand, from rate distortion theory [12, Sec. 13.3.2], the mean-squared error for the Gaussian state sequence S^n can be reduced by a factor of 2^{2C}. Hence, for the relaxed problem, the resulting distortion cannot be smaller than

$$Q2^{-2C} = Q \frac{N}{(\sqrt{P} + \sqrt{Q})^2 + N}.$$

This is clearly a lower bound on the distortion for the original problem, whence we have the desired proof of the converse.

We present another proof, which is more algebraic, but will become useful for the general tradeoff case in the subsequent section. We first 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). Indeed

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

where

(a) follows from the fact that $\hat{S}^n(Y^n)$ is a function of Y^n;

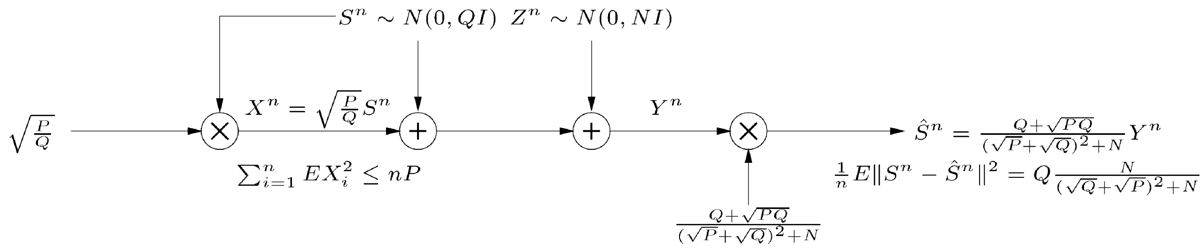

Fig. 6. A state amplification technique.

(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 inequalities

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

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) since $EY_i^2 \leq (\sqrt{P_i}+\sqrt{Q})^2 + N$ with $P_i = EX_i^2$;
(e) from Jensen's inequality; and
(f) follows from the imposed power constraint.

The inequality

$$\frac{1}{2}\log\left(\frac{Q}{D_n}\right) \leq \frac{1}{2}\log\left(\frac{\left(\sqrt{P}+\sqrt{Q}\right)^2+N}{N}\right), \qquad \text{for all } n$$

implies that

$$D_n \geq Q\frac{N}{\left(\sqrt{Q}+\sqrt{P}\right)^2+N}, \qquad \text{for all } n$$

which completes the proof of the converse. □

C. 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 shown by Costa [8]. 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[2] is given by

$$\text{var}(S|Y) = Q\frac{P+N}{Q+P+N}.$$

This is a direct consequence of X^n being statistically uncorrelated with the channel state S^n as observed by Costa [8]. From rate distortion theory, the uncertainty can be further reduced by a factor of 2^{2C}. The resulting mean-squared state estimation error is then given by

$$Q\frac{P+N}{Q+P+N}2^{-2C} = Q\frac{P+N}{Q+P+N}\frac{N}{P+N} = Q\frac{N}{Q+P+N}$$

which is greater than $Q\frac{N}{(\sqrt{Q}+\sqrt{P})^2+N}$, the minimized mean-squared error by the state amplification method.

Instead, the transmitter should use all its power to directly amplify the channel state S^n (i.e., by sending $X^n = \sqrt{\frac{P}{Q}}S^n$). The receiver simply forms 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 $Q\frac{N}{(\sqrt{Q}+\sqrt{P})^2+N}$. There is no codebook involved in this scheme. Furthermore, the encoding/decoding scheme is straightforward, as shown in Fig. 6.

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 $S^n \sim N(0, Q)$ 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 $Z^n \sim N(0, N)$ with an input power constraint P. One optimal technique is for the transmitter to first quantize the source into $\frac{n}{2}\log\left(1+\frac{P}{N}\right)$ bit description, 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 [14], this source can be transmitted uncoded without loss of optimality. Indeed, for each 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 transmitter power requirement). The receiver simply reconstructs $\hat{S}_i = \frac{\sqrt{PQ}}{P+N}Y_i$.

[2]One may hope to further reduce the state estimation error by using the decoded message $\boldsymbol{U}^n(\hat{\boldsymbol{W}})$ in addition to the channel output $\boldsymbol{Y}^n$. However, it is easy to check that, under the capacity-achieving distribution, $\boldsymbol{U}$ and $\boldsymbol{S}$ are conditionally independent given $\boldsymbol{Y}$, thus, $\mathbf{var}(\boldsymbol{S}|\boldsymbol{Y}) = \mathbf{var}(\boldsymbol{S}|\boldsymbol{Y},\boldsymbol{U})$ and it does not help to use the decoded message. See [13] for comparison.

The resulting distortion is also given by $Q\frac{P}{P+N}$. This uncoded technique is optimal and can be easily implemented in practice.

III. 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, independent of the state, over the channel $Y^n = X^n + S^n + Z^n$, where $S^n \sim N(0, QI)$ is the white Gaussian state, $Z^n \sim N(0, NI)$ is the additive white Gaussian noise, and S^n and Z^n are independent (see Fig. 4). The transmitter has full prior knowledge of the state sequence S^n. Based on the message 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 $\hat{S}^n(Y^n) \in \mathbb{R}^n$ of the channel state S^n. More formally, a $(2^{nR}, n)$ code consists of an encoder map

$$X^n : \{1, 2, \ldots, 2^{nR}\} \times \mathbb{R}^n \to \mathbb{R}^n$$

yielding codewords $(X^n(1, S^n), X^n(2, S^n), \ldots, X^n(2^{nR}, S^n))$ that satisfy the expected power constraint

$$E\sum_{i=1}^{n} X_i^2(W, S^n) \le nP, \qquad W = 1, 2, \ldots, 2^{nR}$$

and decoder maps

$$\begin{aligned} \hat{W} &: \mathbb{R}^n \to \{1, 2, \ldots, 2^{nR}\} \\ \hat{S}^n &: \mathbb{R}^n \to \mathbb{R}^n \end{aligned}$$

for pure information decoding and state estimation.

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

$$P_e^{(n)} = \frac{1}{2^{nR}} \sum_{i=1}^{2^{nR}} \Pr(\hat{W}(Y^n) \ne i | W = i)$$

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 probability of errror $P_e^{(n)} \to 0$ and the mean-squared state estimation error $Ed(S^n, \hat{S}^n) \le D$ 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.

The reader is advised to compare this (R, D) tradeoff with the one in classical rate distortion theory. The latter characterizes the description rate of a given source to meet a distortion constraint, while the former characterizes the *tradeoff* between the rate of reliable pure information transmission and the distortion of channel state estimation at the same receiver of the channel. In a sense, this is a tradeoff problem of fundamental quantities in information theory and estimation theory.

Theorem 2: For the state-dependent additive Gaussian channel $Y^n = X^n(W, S^n) + S^n + Z^n$, the optimal (R, D) tradeoff region 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 $0 \le \gamma \le 1$.

In the following, we will focus on the boundary points of the optimal (R, D) tradeoff region given by

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

A. Proof of Achievability

The achievability proof of Theorem 2 is based on power sharing, whereby the transmitter power P is allocated between pure information transmission and state amplification. The scheme works as follows. Fix a power allocation parameter $0 \le \gamma \le 1$ and divide the transmitter power into γP and $(1-\gamma)P$. Based on the state S^n, generate a state-amplification signal

$$X_s^n = \sqrt{\frac{(1-\gamma)P}{Q}} S^n$$

(which 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 $(\sqrt{Q} + \sqrt{(1-\gamma)P})^2$. To send pure information $W \in \{1, 2, \ldots, 2^{nR}\}$, apply the "writing on dirty paper" coding technique used by Costa [8], with the state $S_i + X_{s,i}$ i.i.d. $\sim N\left(0, \left(\sqrt{Q} + \sqrt{(1-\gamma)P}\right)^2\right)$, unknown noise Z_i i.i.d. $\sim N(0, N)$, 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, as shown in [8]. 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 $0 \le \gamma \le 1$, we are able to trade off between the state estimation error and the amount of pure information that can be reliably transmitted. $\square$

Note that we can achieve the same (R, D) tradeoff region under the stronger power constraint $\sum_i X_i^2 \le nP$. As before, this can be easily shown by the strong law of large numbers and the continuity of (5) in P.

B. Proof of Converse

In proving the converse of Theorem 2, we show that given any sequence of $(2^{nR}, n)$ codes with the probability of error $P_e^{(n)} \to 0$ and the asymptotic estimation error

$$D = \liminf_{n\to\infty} \frac{1}{n} E\|S^n - \hat{S}^n\|^2$$

the (R, D) pair satisfies

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

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

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

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

Note that $(R(\gamma), D(\gamma))$ pairs, $0 \le \gamma \le 1$, 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 $0 \le \gamma \le 1$. 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

$$\liminf_{n\to\infty} \frac{1}{n} E\|S^n - \hat{S}^n\|^2 \le D$$

R and D satisfy, for all $\mu \ge 0$

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

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

Recall from (2) in Section II that

$$\frac{1}{2}\log\left(\frac{Q}{D_n}\right) \le \frac{1}{n}I(S^n; Y^n), \quad \text{where } D_n = \frac{1}{n}E\|S^n - \hat{S}^n\|^2.$$

Thus, for $0 \le \mu \le 1$, we can bound the weighted sum $R + \frac{\mu}{2}\log\left(\frac{Q}{D_n}\right)$ as shown in (6) at the bottom of the page, where

(a) follows from the fact that W is uniform over $\{1, 2, \ldots, 2^{nR}\}$;
(b) since W and S^n are independent;
(c) from Fano's inequality with $\epsilon_n \to 0$ as $n \to \infty$;
(d) from the data processing inequality;
(e) from the chain rule and the fact that conditioning reduces entropy; and
(f) since the channel is a discrete memoryless channel.

The inequality (g) needs some explanation. As is proved in Lemma 1 in the Appendix , for each i and for any $0 \le \mu \le 1$, we have

$$\mu h(Y_i) + (1-\mu)h(Y_i|S_i) \le \frac{\mu}{2}\log\left(2\pi e E Y_i^2\right) + \frac{1-\mu}{2}\log\left(2\pi e\left(EY_i^2 - \frac{(ES_iY_i)^2}{ES_i^2}\right)\right) \tag{7}$$

with equality when X_i is jointly Gaussian with S_i and Z_i. Now we represent

$$X_i = V_i + \sqrt{(1-\gamma_i)\frac{P_i}{Q}}S_i$$

where $V_i \sim N(0, \gamma P_i)$ is independent of S_i, with $0 \le \gamma_i \le 1$ chosen such that the resulting covariance of (Y_i, X_i, S_i, Z_i) is the same as

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

that induced by the code. Evaluating the right-hand side of (7) under the jointly Gaussian (Y_i, X_i, S_i, Z_i), we get

$$\begin{aligned}\mu h(Y_i) &+ (1-\mu)h(Y_i|S_i)\\ &\le \frac{\mu}{2}\log\left(2\pi e\left(\sqrt{Q}+\sqrt{(1-\gamma_i)P_i}\right)^2+\gamma_i P_i+N\right)\\ &\quad+\frac{1-\mu}{2}\log\left(2\pi e((\gamma_i P_i+N)\right)\end{aligned}$$

from which the inequality (g) follows immediately.

We continue our chain of inequality by choosing $0\le\gamma\le 1$ such that $\gamma P=(\sum_{i=1}^n \gamma_i P_i)/n$. Observe that $(1-\gamma)P=(\sum_{i=1}^n (1-\gamma_i)P_i)/n$. Now we have (8) at the bottom of the page, where

- (h) follows from Jensen's inequality and the power constraint requirement;
- (i) since γ_μ is chosen to maximize the expression (8) for a fixed μ; and
- (j) follows from the definitions of $R(\gamma)$ and $D(\gamma)$.

Since the expression in (8) is a strictly concave function in γ, and we are maximizing over $0\le\gamma\le 1$, there is a unique optimal γ_μ for a given μ. In short, for all $0\le\mu\le 1$

$$R+\mu\frac{1}{2}\log\left(\frac{Q}{D_n}\right)\le R(\gamma_\mu)+\mu\frac{1}{2}\log\left(\frac{Q}{D(\gamma_\mu)}\right)+\epsilon_n \tag{9}$$

where $0\le\gamma_\mu\le 1$ is chosen to maximize the expression (8) 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 (8).

As a result, in the limit as $\epsilon_n\to 0$ and

$$\liminf_{n\to\infty} D_n=\liminf_{n\to\infty}\frac{1}{n}E\|S^n-\hat{S}^n\|^2=D$$

we have for all $\mu\ge 0$

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

This establishes the converse of Theorem 2. □

Note that when $\mu=0$, the above argument gives a direct proof for the converse part of the coding theorem in Costa's "writing on dirty paper" [8].

C. 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 $0\le\gamma\le 1$, we can trade off between the pure information rate R and the mean-squared estimation error D. In particular, $\gamma=0$ corresponds to the case where the transmitter uses the entire power budget to amplify the channel state, leaving no resources 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 case 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 $D=Q\frac{P+N}{Q+P+N}\le Q$, which suggests that the receiver is still 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 interesting 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

$$\begin{aligned}R+\frac{\mu}{2}\log\left(\frac{Q}{D_n}\right)&\le\frac{1}{n}\sum_{i=1}^n\frac{1}{2}\log\left(\frac{\left(\left(\sqrt{Q}+\sqrt{(1-\gamma_i)P_i}\right)^2+\gamma_i P_i+N\right)^\mu(\gamma_i P_i+N)^{1-\mu}}{N}\right)+\epsilon_n\\ &=\frac{1}{n}\sum_{i=1}^n\frac{1}{2}\log\left(\frac{\left(\sqrt{Q}+\sqrt{(1-\gamma_i)P_i}\right)^2+\gamma_i P_i+N}{N}\right)^\mu\\ &\quad+\frac{1}{n}\sum_{i=1}^n\frac{1}{2}\log\left(\frac{\gamma_i P_i+N}{N}\right)^{1-\mu}+\epsilon_n\\ &\overset{\text{(h)}}{\le}\frac{1}{2}\log\left(\frac{\left(\sqrt{Q}+\sqrt{(1-\gamma)P}\right)^2+\gamma P+N}{N}\right)^\mu+\frac{1}{2}\log\left(\frac{\gamma P+N}{N}\right)^{1-\mu}+\epsilon_n\\ &=\frac{1}{2}\log\left(1+\frac{\gamma P}{N}\right)+\frac{\mu}{2}\log\left(\frac{\left(\sqrt{Q}+\sqrt{(1-\gamma)P}\right)^2+\gamma P+N}{(\gamma P+N)}\right)+\epsilon_n\\ &\overset{\text{(i)}}{\le}\frac{1}{2}\log\left(1+\frac{\gamma_\mu P}{N}\right)+\frac{\mu}{2}\log\left(\frac{\left(\sqrt{Q}+\sqrt{(1-\gamma_\mu)P}\right)^2+\gamma_\mu P+N}{(\gamma_\mu P+N)}\right)+\epsilon_n\\ &\overset{\text{(j)}}{=}R(\gamma_\mu)+\mu\frac{1}{2}\log\left(\frac{Q}{D(\gamma_\mu)}\right)+\epsilon_n\end{aligned}\tag{8}$$

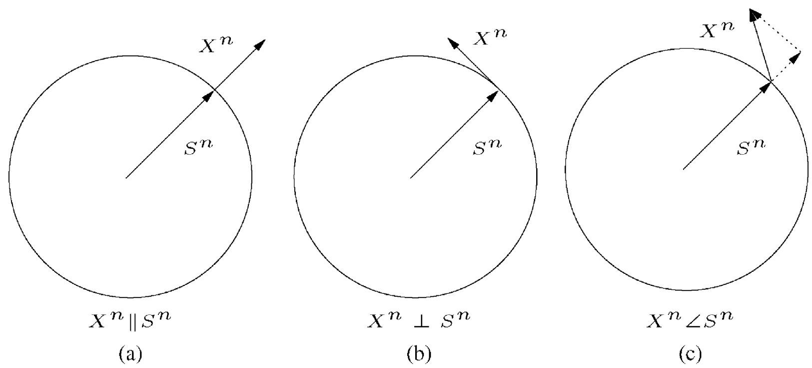

Fig. 7. A diagram relating the transmitted signal $\boldsymbol{X}^n$ and the channel state $\boldsymbol{S}^n$. (a) Pure state amplification, (b) pure information transmission, and (c) combination of state amplification and information transmission.

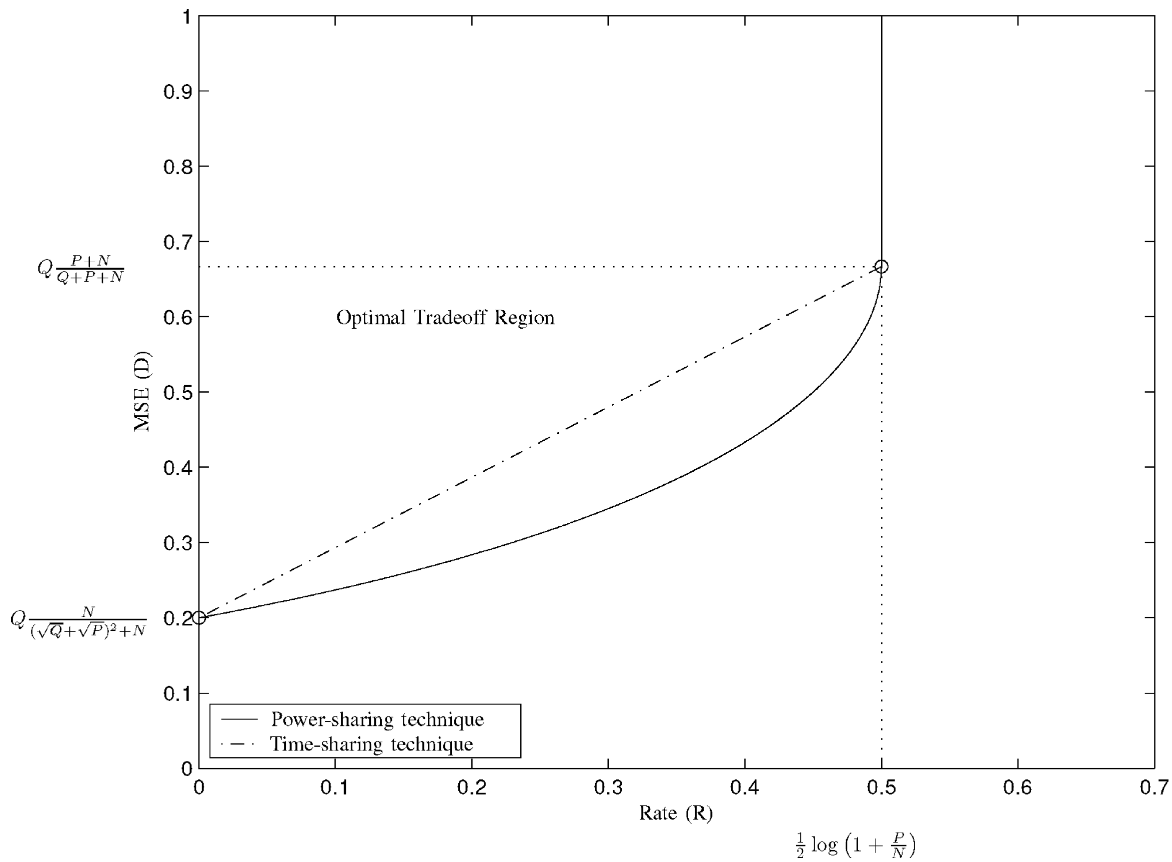

Fig. 8. Optimal $(\boldsymbol{R}, \boldsymbol{D})$ tradeoff region for a state-dependent additive Gaussian channel with states known at the transmitter.

degree of correlation between the transmitted signal X^n and the state S^n. As discussed in Section II, 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. 7(a). Furthermore, as observed by Costa [8], 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. At this operating point, the resulting X^n is statistically uncorrelated with the state S^n as shown in Fig. 7(b). 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. As D increases, the correlation between the transmitted signal X^n and the state S^n decreases (Fig. 7(c)).

IV. 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. $\sim N(0,1)$, noise Z_i i.i.d. $\sim N(0,1)$, and transmitter power constraint $P = 1$. The optimal (R, D) tradeoff region is shown in Fig. 8.

Consider first the case 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 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. As a comparison, if the transmitter attempts, in a suboptimal way, 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

$$Q\frac{N}{Q+P+N}=\frac{1}{3}.$$

On the other hand, when the transmitter's goal is to transmit pure information only, by applying Costa's "writing on dirty paper" coding technique, pure information can be transmitted at the rate given by $\frac{1}{2}\log(1+\frac{P}{N})=\frac{1}{2}$ bits per transmission. Due to the state-dependent nature of the channel, the receiver is still able to learn something about the state from observing the channel output; the uncertainty about the state is reduced 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. 8 are achievable. As a comparison, a tradeoff region based on a naive time-sharing technique is shown.

V. Concluding Remarks

When the goal is to minimize the state estimation error at the receiver, the optimal transmission technique is to use all the sender's power to amplify the channel state. It would be strictly suboptimal to use the channel to send the description of the state. On the other hand, when the goal is to transmit only pure information, the sender can use the "writing on dirty paper" coding technique. But pure information transmission obscures the receiver's view of the channel state, thereby increasing the state estimation error. For this intrinsic conflict, 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 $0\le\mu\le 1$

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

with equality when X is jointly Gaussian with S and Z. In other words, for any 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: Let $\alpha=(ESY)/(ES^2)$. For a fixed $0\le\mu\le 1$, we have the following chain of inequalities:

$$\begin{aligned}\mu h(Y)&+(1-\mu)h(Y|S)\\ &\overset{(a)}{\le}\frac{\mu}{2}\log\left(2\pi e\, EY^2\right)+(1-\mu)h(Y|S)\\ &\overset{(b)}{=}\frac{\mu}{2}\log\left(2\pi e\, EY^2\right)+(1-\mu)h(Y-\alpha S|S)\\ &\overset{(c)}{\le}\frac{\mu}{2}\log\left(2\pi e\, EY^2\right)+(1-\mu)h(Y-\alpha S)\\ &\overset{(d)}{\le}\frac{\mu}{2}\log\left(2\pi e\, EY^2\right)+\frac{1-\mu}{2}\log\left(2\pi e\left(EY^2-\frac{(ESY)^2}{ES^2}\right)\right)\end{aligned}$$

where

(a) follows because the Gaussian distribution maximizes the entropy for a given variance;

(b) since the translation does not change the differential entropy;

(c) since conditioning reduces entropy; and

(d) since the Gaussian distribution maximizes the entropy for a given variance and $E(Y-\alpha S)^2=EY^2-\frac{(ESY)^2}{ES^2}$.

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, we have equality in (a) and (d) since $Y=X+S+Z$ is Gaussian and equality in (c) because $Y-\alpha S$ and S are jointly Gaussian and orthogonal to each other, thus mutually independent. □

Acknowledgment

The authors wish to thank anonymous reviewers for many fruitful suggestions, especially the alternative proof of the converse Theorem 1 and simpler arguments for the proof of Lemma 1.

References

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

[2] A. V. Kusnetsov and B. S. Tsybakov, "Coding in a memory with defective cells," *Probl. Pered. Inform.*, vol. 10, no. 2, pp. 52–60, Apr./Jun. 1974. Translated from Russian.

[3] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Contr. Inform. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[4] C. Heegard and A. El Gamal, "On the capacity of computer memories with defects," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 5, pp. 731–739, Sep. 1983.

[5] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 563–593, Mar. 2003.

[6] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[7] A. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1639–1667, Jun. 2002.

[8] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 439–411, May 1983.

[9] W. Yu, A. Sutivong, D. Julian, T. M. Cover, and M. Chiang, "Writing on colored paper," in *Proc. IEEE Int Symp. Information Theory (ISIT 2001)*, Washington, DC, Jun. 2001, p. 302.

[10] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250–1276, Jun. 2002.

[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 Trans. Commun.*, vol. 46, no. 9, pp. 1233–1242, Sep. 1998.

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

[13] S. Shamai (Shitz), S. Verdú, and R. Zamir, "Systematic lossy source/channel coding," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 564–579, Mar. 1998.

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