

# Minimum Expected Distortion in Gaussian Source Coding with Fading Side Information

Chris T. K. Ng, *Member, IEEE*, Chao Tian, *Member, IEEE*,  
Andrea J. Goldsmith, *Fellow, IEEE*, and Shlomo Shamai (Shitz), *Fellow, IEEE*

## Abstract

An encoder, subject to a rate constraint, wishes to describe a Gaussian source under squared error distortion. The decoder, besides receiving the encoder's description, also observes side information consisting of uncompressed source symbol subject to slow fading and noise. The decoder knows the fading realization but the encoder knows only its distribution. The rate-distortion function that simultaneously satisfies the distortion constraints for all fading states was derived by Heegard and Berger. A layered encoding strategy is considered in which each codeword layer targets a given fading state. When the side-information channel has two discrete fading states, the expected distortion is minimized by optimally allocating the encoding rate between the two codeword layers. For multiple fading states, the minimum expected distortion is formulated as the solution of a convex optimization problem with linearly many variables and constraints. Through a limiting process on the primal and dual solutions, it is shown that single-layer rate allocation is optimal when the fading probability density function is continuous and quasiconcave (e.g., Rayleigh, Rician, Nakagami, and log-normal). In particular, under Rayleigh fading, the optimal single codeword layer targets the least favorable state as if the side information was absent.

## Index Terms

Convex optimization, distortion minimization, fading channel, Heegard-Berger, rate-distortion function, side information, source coding.

## I. INTRODUCTION

**I**N lossy data compression, side information at the decoder can help reduce the distortion in the reconstruction of the source [1]. The decoder, however, may have access to the side information only through an unreliable channel. For example, in distributed compression over wireless sensor networks, correlated sensor measurements from a neighboring node may be available to the decoder through a fading wireless channel. In this work, we consider a Gaussian source where the encoder is subject to a rate constraint and the distortion metric is the mean squared error of the reconstruction. In addition to the compressed symbol, we assume that the decoder observes the original symbol through a separate analog fading channel. We assume, similar to the approach in [2], that the fading is quasistatic, and that the decoder knows the fading realization but the encoder knows only its distribution. The rate-distortion function that dictates the rate required to satisfy the distortion constraint associated with each fading state is given by Heegard and Berger in [3]. We consider a layered encoding strategy based on the uncertain fading realization in the side-information channel, and optimize the rate allocation among the possible fading states to minimize the expected distortion. In particular, we formulate the distortion

This work was supported by the US Army under MURI award W911NF-05-1-0246, the ONR under award N00014-05-1-0168, DARPA's ITMANET program under grant 1105741-1-TFIND, and a grant from Intel. The work of C. Ng was supported by a Croucher Foundation Fellowship. The work of S. Shamai has been supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMMunications NEWCOM++. The material in this paper was presented in part at the IEEE Information Theory Workshop, Lake Tahoe, CA, September 2007.

C. Ng was with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: Chris.Ng@ieee.org).

C. Tian is with AT&T Labs-Research, Florham Park, NJ 07932 USA (e-mail: tian@research.att.com).

A. Goldsmith is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: andrea@wsl.stanford.edu).

S. Shamai (Shitz) is with the Department of Electrical Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel (e-mail: sshlomo@ee.technion.ac.il).

minimization as a convex optimization problem, and develop an efficient representation for the Heegard–Berger rate–distortion function under which the optimization problem size is linear in the number of discrete fading states. Furthermore, we identify the conditions under which single-layer rate allocation is expected-distortion–minimizing, and extend these optimality conditions for continuous fading distributions through a limiting process on the primal and dual solutions in the optimization. We show that single-layer rate allocation is optimal for fading distributions with continuous, quasiconcave probability density functions such as Rayleigh, Rician, Nakagami, and log-normal.

When the side-information channel exhibits no fading, the distortion is given by the Wyner–Ziv rate–distortion function [4]. Rate–distortion is considered in [5], [6] when the side information is also available at the encoder, and in [7] when there is a combination of decoder-only and encoder-and-decoder side information. Successive refinement source coding in the presence of side information is considered in [8], [9]. The side-information scalable rate–distortion region is characterized in [10], in which the user with inferior side information decodes an additional layer of the source-coding codeword. Lossless source coding with an unknown amount of side information at the decoder is considered in [11], in which a fixed data block is broadcast to different users in a variable number of channel uses [12]. In [13], [14], expected distortion is minimized in the transmission of a Gaussian source over a slowly fading channel in the absence of channel state information at the transmitter (CSIT). Broadcast transmission with imperfect CSIT is considered in [15]. Another application of source coding with uncertain side information is in systematic lossy source-channel coding [16] over a fading channel without CSIT. For example, when upgrading legacy communication systems, a digital channel may be added to augment an existing analog channel. In this case the analog reception then plays the role of side information in the decoding of the description from the digital channel. In [17], [18], hybrid digital/analog and digital transmission schemes are considered for Wyner–Ziv coding over broadcast channels. The system model studied in this paper is also related to distributed source coding over multiple links [19], [20] where, besides source coding over a finite-capacity reliable link, noisy versions of the source are described through additional backhaul links with infinite capacity but that are subject to random failure. At the decoder, the realized quality of the side information is determined by the number of backhaul links that are successfully connected. Similar models are considered in [21] for distributed unreliable relay communications.

The remainder of the paper is organized as follows. The system model is described in Section II. Section III derives the minimum expected distortion and presents the convex optimization framework when the side-information channel has discrete fading states. Section IV investigates the optimal rate allocation under different fading distributions in the side-information channel. Section V considers the optimality of single-layer rate allocation under discrete fading states as well as continuous fading distributions. Conclusions are given in Section VI.

## II. SYSTEM MODEL

### A. Source Coding with Fading Side-Information Channel

Consider the system model shown in Fig. 1. An encoder wishes to describe a real Gaussian source sequence  $\{X\}$  under a rate constraint of  $R_X$  nats per symbol, where the sequence of random variables are independent identically distributed (i.i.d.) with  $X \sim \mathcal{N}(0, \sigma_X^2)$ . The decoder, in addition to receiving the encoder’s description, observes side information  $Y'$ , where  $Y' = \sqrt{S}X + Z$ , with  $Z \sim \text{i.i.d. } \mathcal{N}(0, 1)$ . Hence the quality of the side information depends on  $S$ , the power gain of the side-information channel. We assume  $S$  is a quasistatic random variable: it is drawn from some cumulative distribution function (cdf)  $F(s)$  at the beginning of each transmission block and remains unchanged through the block. The decoder knows the realization of  $S$ , but the encoder knows only its distribution  $F(s)$ . When the fading distribution is continuous, it is characterized by the probability density function (pdf)  $f(s) = F'(s)$ . The decoder forms an estimate of the source and reconstructs the sequence  $\{\hat{X}\}$ . We are interested in minimizing the expected squared error distortion  $\mathbb{E}[D]$  of the reconstruction, where  $D = (X - \hat{X})^2$ .

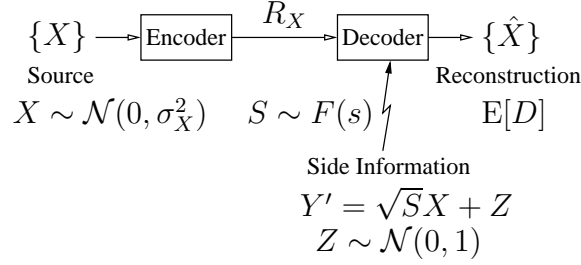


Fig. 1. Source coding with fading side-information channel.

Suppose the side-information channel has  $M$  discrete fading states. Let the probability distribution of  $S$  be given as follows:

$$\Pr\{S = s_i\} = p_i, \quad i = 1, \dots, M, \quad \sum_{i=1}^M p_i = 1 \quad (1)$$

where the  $s_i$ 's are enumerated in ascending order  $0 \leq s_1 < s_2 < \dots < s_M$ . Let  $Y'_i$  denote the side information under fading state  $s_i$

$$Y'_i \triangleq \sqrt{s_i}X + Z, \quad i = 1, \dots, M. \quad (2)$$

Note that the set of side information random variables are stochastically degraded. Let  $\hat{X}_i$  be the reconstruction when side information  $Y'_i$  is available at the decoder, and  $D_i$  be the corresponding squared error distortion. The minimum expected distortion under rate constraint  $R_X$  is then given by

$$E[D]^* = \min_{\mathbf{D}: R(\mathbf{D}) \leq R_X} \mathbf{p}^T \mathbf{D} \quad (3)$$

where  $\mathbf{p} \triangleq [p_1 \dots p_M]^T$ ,  $\mathbf{D} \triangleq [D_1 \dots D_M]^T$ , and  $R(\mathbf{D})$  is the rate-distortion function that simultaneously satisfies the distortion set  $\mathbf{D}$ .

### B. Heegard-Berger Rate-Distortion Function

The rate-distortion function that dictates the rate required to simultaneously satisfy a set of distortion constraints associated with a set of degraded side-information random variables is given by Heegard and Berger in [3] (an alternate form for  $M = 2$  is described in [22]). When the side information random variables satisfy the degradedness condition  $X \leftrightarrow Y_M \leftrightarrow Y_{M-1} \leftrightarrow \dots \leftrightarrow Y_1$ , the rate-distortion function is

$$R_{\text{HB}}(\mathbf{D}) = \min_{W_1^M \in P(\mathbf{D})} \sum_{i=1}^M I(X; W_i | Y_i, W_1^{i-1}) \quad (4)$$

where  $W_1^i$  denotes the vector  $W_1, \dots, W_i$ . The minimization takes place over  $P(\mathbf{D})$ , the set of all  $W_1^M$  jointly distributed with  $X, Y_1^M$  such that:

$$W_1^M \leftrightarrow X \leftrightarrow Y_M \leftrightarrow Y_{M-1} \leftrightarrow \dots \leftrightarrow Y_1 \quad (5)$$

and there exists decoding functions  $\hat{X}_i(Y_i, W_1^i)$ 's under given distortion measures  $d_i$ 's that satisfy

$$E[d_i(X, \hat{X}_i)] \leq D_i, \quad i = 1, \dots, M. \quad (6)$$

As noted in [3], since  $R_{\text{HB}}(\mathbf{D})$  depends on  $X, Y_1^M$  only through the marginal distribution  $p(x, y_i)$ ,  $i = 1, \dots, M$ , the degradedness of the side information need not be physical. We construct  $Y_1^M$  to have the same marginals as  $Y'_1^M$  by setting  $p(y_i|x) = p(y'_i|x)$ ,  $i = 1, \dots, M$ . The rate-distortion function  $R(\mathbf{D})$  in (3) is then given by the Heegard-Berger rate-distortion function (4) with squared error distortion measures  $d_i(X, \hat{X}_i) = (X - \hat{X}_i)^2$ .

### III. MINIMUM EXPECTED DISTORTION

#### A. Gaussian Source under Squared Error Distortion

The Heegard–Berger rate–distortion function  $R_{\text{HB}}(D_1, \dots, D_M)$  for a Gaussian source under squared error distortion is given in [3], [9], [10]. For  $M = 2$ , [3] describes the Gaussian rate–distortion function where the worst fading state corresponds to no side information, and [9] considers side information with different quality levels. The Gaussian Heegard–Berger rate–distortion function is considered in [10] for  $M > 2$ . However, the representations of  $R_{\text{HB}}(D_1, \dots, D_M)$  described in [3], [9] are characterized by exponentially-many distinct regions, and [10] involves optimal Gaussian random variables whose variances are determined by an algorithmic procedure. These characterizations, though complete, are not amenable to efficient minimization of the expected distortion. In this section, we derive a representation for  $R_{\text{HB}}(D_1, \dots, D_M)$  that can be incorporated in an optimization framework. In particular, instead of describing the achievable distortion set by its exponentially-many segments of boundaries, we construct a characterization that comprises a sequence of convex inequalities, each relating the achievable distortion between two adjacent fading states. Consequently, we formulate the distortion minimization as a convex optimization problem where the number of variables and constraints are linear in  $M$ . First we consider the case when the side-information channel has only two discrete fading states ( $M = 2$ ); in Section III-C we extend the analysis to multiple fading states where  $M > 2$ .

When the side-information channel has two fading states, the encoder constructs a source-coding scheme that consists of two layers of codewords. The base layer is designed to be decodable under either channel condition, while the top layer is only decodable under the more favorable channel realization. We derive the rate requirements of the two codeword layers, and optimally allocate the encoding rate  $R_X$  between them to minimize the expected distortion. For  $M = 2$ , the Heegard–Berger rate–distortion function is given by

$$R_{\text{HB}}(D_1, D_2) = \min_{W_1, W_2 \in \mathcal{P}(D_1, D_2)} \{I(X; W_1|Y_1) + I(X; W_2|Y_2, W_1)\}. \quad (7)$$

For a Gaussian source under a squared error distortion measure, a jointly Gaussian codebook is optimal [3], [9], [10]. When  $W_1^M, X$  are jointly Gaussian, the mutual information expressions in (7) evaluate to

$$\begin{aligned} & I(X; W_1|Y_1) + I(X; W_2|Y_2, W_1) \\ &= \frac{1}{2} \log(\text{VAR}[X|Y_1]) - \frac{1}{2} \log \frac{\text{VAR}[X|Y_1, W_1]}{\text{VAR}[X|Y_2, W_1, W_2]} - \frac{1}{2} \log(\text{VAR}[X|Y_2, W_1, W_2]) \end{aligned} \quad (8)$$

$$= -\frac{1}{2} \log(s_1 + \sigma_x^{-2}) - \frac{1}{2} \log(1 + (s_2 - s_1)\text{VAR}[X|Y_1, W_1]) - \frac{1}{2} \log(\text{VAR}[X|Y_2, W_1, W_2]) \quad (9)$$

where  $\log$  is the natural logarithm, and (9) follows from expanding the conditional variance expressions by applying Lemma 1 and Corollary 1. The proof of Lemma 1 is given in Appendix A.

**Lemma 1.** *Let  $X, W_1^k$  be jointly Gaussian random variables. If  $Y = \sqrt{s}X + Z$ , where  $Z \sim \mathcal{N}(0, 1)$  is independent from  $X, W_1^k$ , then*

$$\text{VAR}[X|Y, W_1^k] = (\text{VAR}[X|W_1^k]^{-1} + s)^{-1}. \quad (10)$$

**Corollary 1.** *Let  $Y_j = \sqrt{s_j}X + Z$ ,  $Y_i = \sqrt{s_i}X + Z$ .*

$$\frac{\text{VAR}[X|Y_i, W_1^k]}{\text{VAR}[X|Y_j, W_1^k]} = 1 + (s_j - s_i)\text{VAR}[X|Y_i, W_1^k]. \quad (11)$$

To characterize the Heegard–Berger rate–distortion function  $R_{\text{HB}}(D_1, D_2)$ , we substitute (9) in (7), and minimize over  $W_1, W_2$

$$\begin{aligned} R_{\text{HB}}(D_1, D_2) = & -\frac{1}{2} \log(s_1 + \sigma_x^{-2}) + \min_{W_1} \left\{ -\frac{1}{2} \log(1 + (s_2 - s_1)\text{VAR}[X|Y_1, W_1]) \right. \\ & \left. + \min_{W_2} \left\{ -\frac{1}{2} \log(\text{VAR}[X|Y_2, W_1, W_2]) \right\} \right\}. \end{aligned} \quad (12)$$

Note that  $s_2 > s_1 \geq 0$  by assumption. Accordingly, in the inner minimization in (12),  $R_{\text{HB}}(D_1, D_2)$  is decreasing in  $\text{VAR}[X|Y_2, W_1, W_2]$ . Hence the choice of  $W_2$  is optimal when  $\text{VAR}[X|Y_2, W_1, W_2]$  is increased until one of its upper bound constraints is tight

$$\max_{W_2} \text{VAR}[X|Y_2, W_1, W_2] = \min(\text{VAR}[X|Y_2, W_1], D_2). \quad (13)$$

The optimal  $W_2^*$  that achieves (13) is presented subsequently. The first term in the  $\min(\cdot)$  expression in (13) follows from the non-negativity of the mutual information  $I(X; W_2|Y_2, W_1)$ , and the second one follows from the distortion constraint on  $\hat{X}_2$  as given in (6)

$$\text{VAR}[X|Y_2, W_1, W_2] = \mathbb{E}[(X - \hat{X}_2(Y_2, W_1, W_2))^2] \leq D_2. \quad (14)$$

Applying Corollary 1, the first term in (13) evaluates to

$$\text{VAR}[X|Y_2, W_1] = (\text{VAR}[X|Y_1, W_1]^{-1} + s_2 - s_1)^{-1}. \quad (15)$$

Under optimal  $W_2$ , therefore, the Heegard–Berger rate–distortion function in (12) reduces to

$$\begin{aligned} R_{\text{HB}}(D_1, D_2) = & -\frac{1}{2} \log(s_1 + \sigma_x^{-2}) + \min_{W_1} \left\{ -\frac{1}{2} \log(1 + (s_2 - s_1) \text{VAR}[X|Y_1, W_1]) \right. \\ & \left. - \frac{1}{2} \log \min\left((\text{VAR}[X|Y_1, W_1]^{-1} + s_2 - s_1)^{-1}, D_2\right) \right\}. \end{aligned} \quad (16)$$

The maximization over  $W_1$  in (16) has a similar structure as the one previously considered in (13). Specifically,  $R_{\text{HB}}(D_1, D_2)$  in (16) is decreasing in  $\text{VAR}[X|Y_1, W_1]$ . Hence  $W_1$  is optimal when  $\text{VAR}[X|Y_1, W_1]$  is increased until it meets one of its upper bound constraints

$$\max_{W_1} \text{VAR}[X|Y_1, W_1] = \min(\text{VAR}[X|Y_1], D_1) \quad (17)$$

where the first term in (17) follows from the non-negativity of  $I(X; W_1|Y_1)$ , and the second one from the distortion constraint on  $\hat{X}_1$

$$\text{VAR}[X|Y_1, W_1] = \mathbb{E}[(X - \hat{X}_1(Y_1, W_1))^2] \leq D_1. \quad (18)$$

Next, we consider the construction of  $W_1, W_2$  that achieves the rate–distortion function, namely, jointly Gaussian random variables with conditional variances that satisfy (13), (17). We construct the optimal distribution  $W_1^*, W_2^*$  as follows:

$$W_1^* = \sqrt{a_1}X + N_1 \quad (19)$$

$$W_2^* = \sqrt{a_2}X + N_2 \quad (20)$$

where  $N_i \sim \text{i.i.d. } \mathcal{N}(0, 1)$ ,  $i = 1, 2$ , is independent from  $X, Y_1, Y_2$ , and  $a_1, a_2$  are nonnegative scalars whose values are to be specified. For notational convenience, we define

$$V_1 \triangleq \text{VAR}[X|Y_1, W_1^*] \quad (21)$$

$$= \min((\sigma_X^{-2} + s_1)^{-1}, D_1) \quad (22)$$

where (22) follows from (17). Substitute (19) in (21), and  $a_1$  evaluates to

$$a_1 = V_1^{-1} - \sigma_X^{-2} - s_1. \quad (23)$$

Similarly, to identify the optimal  $W_2^*$ , we define

$$V_2 \triangleq \text{VAR}[X|Y_2, W_1^*, W_2^*] \quad (24)$$

$$= \min((V_1^{-1} + s_2 - s_1)^{-1}, D_2) \quad (25)$$

which follows from (13), (15). Substitute (20) in (24), and  $a_2$  evaluates to

$$a_2 = V_2^{-1} - V_1^{-1} - (s_2 - s_1). \quad (26)$$

To provide an interpretation regarding the source encoding rates under different fading states of the side-information channel, we introduce the notations

$$R_1 \triangleq I(X; W_1^* | Y_1) \quad (27)$$

$$= \frac{1}{2} \log \frac{(\sigma_X^{-2} + s_1)^{-1}}{V_1} \quad (28)$$

$$R_2 \triangleq I(X; W_2^* | Y_2, W_1^*) \quad (29)$$

$$= \frac{1}{2} \log \frac{(V_1^{-1} + s_2 - s_1)^{-1}}{V_2} \quad (30)$$

where (28), (30) follow from expanding the mutual information expressions applying (23), (26). We interpret  $R_1$  as the rate of a source coding base layer that describes  $X$  when the side-information quality is that of  $Y_1$  or better. On the other hand,  $R_2$  is the rate of a top layer that describes  $X$  only when the decoder has the better side information  $Y_2$ . Finally, we substitute (28), (30) in (7) to obtain the two-layer Heegard–Berger rate–distortion function

$$R_{\text{HB}}(D_1, D_2) = R_1 + R_2 \quad (31)$$

$$= -\frac{1}{2} \log(\sigma_X^{-2} + s_1) - \frac{1}{2} \log V_2 - \frac{1}{2} \log(1 + (s_2 - s_1)V_1) \quad (32)$$

where  $V_1, V_2$  are as defined in (21), (24) above. The derivation of (32) depends on the side information only through the marginals  $p(y_i|x)$ 's; therefore, the rate–distortion function applies as well to the stochastically degraded side information  $Y'_M, \dots, Y'_1$ .

### B. Optimal Distortion Trade-off and Rate Allocation

Under a source-coding rate constraint of  $R_X$ , the Heegard–Berger feasible distortion region is described by

$$\mathcal{D}(R_X) \triangleq \{(D_1, D_2) \mid R_{\text{HB}}(D_1, D_2) \leq R_X\}. \quad (33)$$

The distortion regions under different values of  $R_X$  are illustrated in Fig. 2. Setting  $R_{\text{HB}}(D_1, D_2) = R_X$ , the dominant boundary of  $\{(D_1, D_2)\}$  defines the Pareto optimal trade-off curve (shown in bold in Fig. 2) between the two distortion constraints on  $\hat{X}_1$  and  $\hat{X}_2$ , which is given by

$$D_2 = [e^{2R_X}(\sigma_X^{-2} + s_1)(1 + (s_2 - s_1)D_1)]^{-1} \quad (34)$$

over the interval

$$(e^{2R_X}(\sigma_X^{-2} + s_1))^{-1} \leq D_1 \leq (\sigma_X^{-2} + s_1)^{-1}. \quad (35)$$

We find the optimal operating point on the Pareto curve to minimize the expected distortion

$$\mathbb{E}[D]^* = \min_{D_1, D_2: R_{\text{HB}}(D_1, D_2) \leq R_X} p_1 D_1 + p_2 D_2. \quad (36)$$

In Section III-C, it is shown that the above minimization is a convex optimization problem. Hence the Karush–Kuhn–Tucker (KKT) conditions are necessary and sufficient for optimality. Moreover,  $\mathcal{D}(R_X)$ , being the sublevel set of a convex function, is a convex set. After substituting (34) in (36), from the KKT optimality conditions, we obtain the optimal base layer distortion

$$D_1^* = (D_1^* \star)_{[D_1^-, D_1^+]} \quad (37)$$

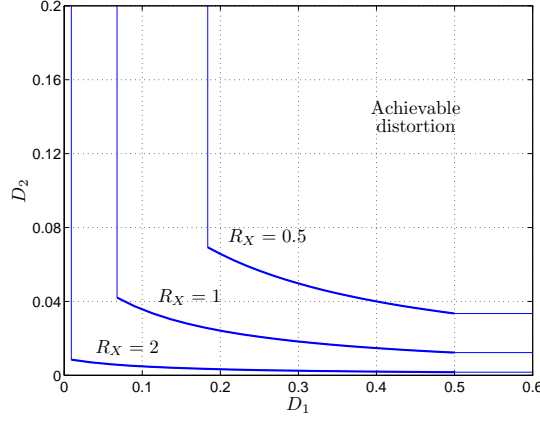


Fig. 2. Achievable distortion under different values of the encoding rate constraint  $R_X$ . For each  $R_X$ , the Pareto optimal trade-off curve is shown in bold.

where  $(x)_{[a,b]}$  denotes the projection

$$(x)_{[a,b]} \triangleq \min(\max(a, x), b) \quad (38)$$

and the distortion and its boundaries are given by

$$D_1^- \triangleq (e^{2R_X}(\sigma_X^{-2} + s_1))^{-1} \quad (39)$$

$$D_1^* \triangleq \frac{1}{s_2 - s_1} \left[ \left( e^{2R_X} \frac{\sigma_X^{-2} + s_1}{s_2 - s_1} \frac{p_1}{p_2} \right)^{-1/2} - 1 \right] \quad (40)$$

$$D_1^+ \triangleq (\sigma_X^{-2} + s_1)^{-1}. \quad (41)$$

The optimal top layer distortion  $D_2^*$  is given by

$$D_2^* = (D_2^*)_{[D_2^-, D_2^+]} \quad (42)$$

where

$$D_2^- \triangleq (e^{2R_X}(\sigma_X^{-2} + s_2))^{-1} \quad (43)$$

$$D_2^* \triangleq (e^{2R_X}(\sigma_X^{-2} + s_1)(s_2 - s_1)p_2/p_1)^{-1/2} \quad (44)$$

$$D_2^+ \triangleq (e^{2R_X}(\sigma_X^{-2} + s_1) + s_2 - s_1)^{-1}. \quad (45)$$

The corresponding optimal rate allocation  $R_1^*, R_2^*$  can be found as given in (28), (30).

The optimal rate allocation and the corresponding minimum expected distortion are plotted in Fig. 3 and Fig. 4, respectively, for  $R_X = 1$ ,  $\sigma_X^2 = 1$ , and  $s_1 = 0$  dB. Note that  $R_2^*$ , the rate allocated to the top layer, is not monotonic with the side-information channel condition. As fading state  $s_2$  improves,  $R_2^*$  increases to take advantage of the better side-information quality. However, when  $s_2$  is large,  $R_2^*$  begins to decline as the expected distortion is dominated by the worst fading state. In addition, the optimal rate allocation is heavily skewed towards the lower layer:  $R_2^* > 0$  only when  $p_2$  is large.

### C. Multiple Discrete Fading States

The rate-distortion function (32) extends directly to the case when the side-information channel has multiple discrete fading states:  $S = s_i$  with probability  $p_i$ , where  $i = 1, \dots, M$ , with  $0 \leq s_1 < \dots < s_M$ , and  $M > 2$ . The Heegard-Berger rate-distortion function for  $M > 2$  can be characterized by a similar

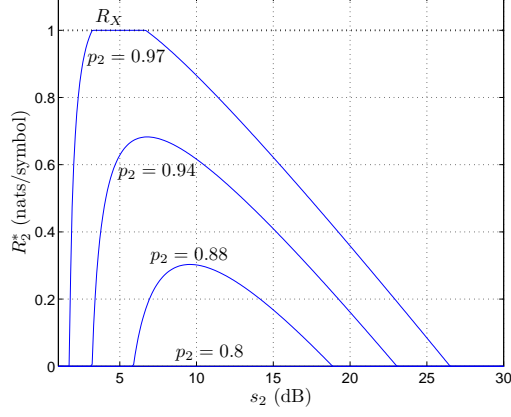


Fig. 3. Optimal rate allocation that minimizes expected distortion. The side-information channel has two discrete fading states ( $s_1 = 0$  dB).

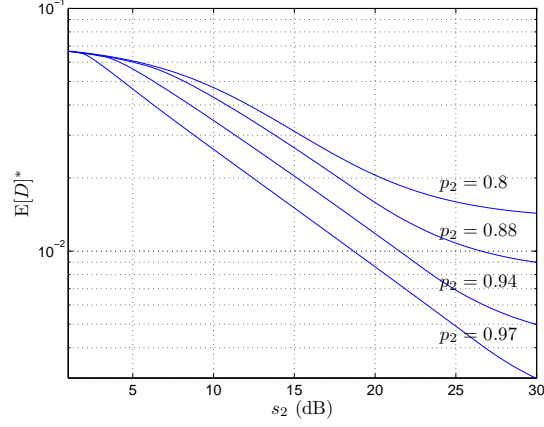


Fig. 4. Minimum expected distortion under optimal rate allocation. The side-information channel has two discrete fading states ( $s_1 = 0$  dB).

representation as that given in (32) for  $M = 2$ . Specifically, we construct the optimal distribution for the auxiliary random variable  $W_i^*$ 's to be given by

$$W_i^* = \sqrt{a_i}X + N_i, \quad i = 1, \dots, M \quad (46)$$

where  $N_i \sim \text{i.i.d. } \mathcal{N}(0, 1)$ , and  $a_i$ 's are nonnegative scalars whose values are to be specified. The rate of the  $i$ th layer is

$$R_i \triangleq I(X; W_i^* | Y_i, W_1^*, \dots, W_{i-1}^*) \quad (47)$$

$$= \frac{1}{2} \log \frac{(V_{i-1}^{-1} + s_i - s_{i-1})^{-1}}{V_i} \quad (48)$$

where

$$V_i \triangleq \text{VAR}[X | Y_i, W_1^*, \dots, W_i^*] \quad (49)$$

$$= \min((V_{i-1}^{-1} + s_i - s_{i-1})^{-1}, D_i) \quad (50)$$

and  $s_0 \triangleq 0$ ,  $V_0 \triangleq \sigma_X^2$  for convenience in notations. In the above, (50) follows from the non-negativity of  $I(X; W_i^* | Y_i, W_1^*, \dots, W_{i-1}^*)$  and the distortion constraint (6). The  $a_i$  that achieves (49) is determined from (50), which evaluates to

$$a_i = V_i^{-1} - V_{i-1}^{-1} - (s_i - s_{i-1}). \quad (51)$$



As  $R_{\text{HB}}(\mathbf{D}) = \sum_{i=1}^M R_i$ , we substitute (48) in (4) to obtain the rate–distortion function

$$R_{\text{HB}}(\mathbf{D}) = -\frac{1}{2} \log(\sigma_X^{-2} + s_1) - \frac{1}{2} \log V_M - \frac{1}{2} \sum_{i=1}^{M-1} \log(1 + (s_{i+1} - s_i)V_i) \quad (52)$$

where the  $V_i$ 's are as given in (50).

Under multiple fading states, however, a closed-form expression for the minimum expected distortion  $\mathbb{E}[D]^*$  does not appear analytically tractable. Nevertheless, the expected distortion minimization in (3) can be formulated as the following convex optimization problem:

$$\text{minimize } J(D_1, \dots, D_M) \quad (53)$$

$$\text{over } D_1, \dots, D_M, V_1, \dots, V_M \in \mathbb{R}_{++} \quad (54)$$

subject to

$$-\frac{1}{2} \log(\sigma_X^{-2} + s_1) - \frac{1}{2} \log V_M - \frac{1}{2} \sum_{i=1}^{M-1} \log(1 + (s_{i+1} - s_i)V_i) \leq R_X \quad (55)$$

$$V_i \leq (V_{i-1}^{-1} + s_i - s_{i-1})^{-1}, \quad i = 1, \dots, M \quad (56)$$

$$V_i \leq D_i, \quad i = 1, \dots, M \quad (57)$$

where  $\mathbb{R}_{++}$  denotes the set of positive real numbers. In (53) above, the cost function  $J(\cdot)$  may be any arbitrary function that is convex in  $D_1, \dots, D_M$ . The constraint (55) prescribes the feasible Heegard–Berger distortion region under the source-coding rate constraint  $R_X$ . The constraints (56) and (57) derive from writing out the two upper bounds for each  $V_i$ , as described in (50), as two separate inequality constraints. The equality in (50) may be written as inequality constraints since there is an optimal solution where for each  $i$  at least one of (56) or (57) is tight. Specifically, the left-hand side of the Heegard–Berger constraint in (55) is monotonically decreasing in  $V_i$ 's. Hence for a given optimal  $\{V_i^*, D_i^*\}$ , if neither (56) nor (57) is tight,  $V_i^*$  may be increased to strictly enlarge the feasible set of  $\{D_1, \dots, D_M, V_1, \dots, V_M\} \setminus \{D_i, V_i\}$ .

**Proposition 1.** *The minimization given in (53)–(57) is a convex optimization problem.*

See Appendix B for the proof of Proposition 1. Convexity implies that a local optimum is globally optimal, and its solution can be efficiently computed by standard convex optimization numerical techniques, for instance, by the interior-point method [23], [24]. Moreover, the optimization problem (53)–(57) has  $2M$  variables and  $2M + 1$  inequality constraints, which are linear in the number of side-information channel fading states  $M$ .

In the case where the cost function  $J(D_1, \dots, D_M)$  is non-decreasing in each component  $D_i$ , the constraints (57) may be taken as tight: if at their optimal values  $V_i^* < D_i^*$ , then  $D_i^*$  may be decreased without violating feasibility nor increasing the cost function. In particular, in the remainder of the paper, we consider minimizing the expected distortion:

$$J(\mathbf{D}) = \mathbb{E}[D] = \sum_{i=1}^M p_i D_i. \quad (58)$$

In this case, the optimization problem can be specified more compactly as

$$\text{minimize } p_1 D_1 + \dots + p_M D_M \quad (59)$$

$$\text{over } D_1, \dots, D_M \quad (60)$$

subject to

$$-\frac{1}{2} \log(\sigma_X^{-2} + s_1) - \frac{1}{2} \log D_M - \frac{1}{2} \sum_{i=1}^{M-1} \log(1 + (s_{i+1} - s_i)D_i) \leq R_X \quad (61)$$

$$D_i \leq (D_{i-1}^{-1} + s_i - s_{i-1})^{-1}, \quad i = 1, \dots, M \quad (62)$$

where in (62) similarly  $D_0 \triangleq \sigma_X^2$ . For convenience in stating the optimization problem, in (61), and in the remainder of the paper,  $\log$  refers to the extended-value logarithmic function, where it takes on the value  $-\infty$  for non-positive arguments. Then the feasibility constraints (61), (62) imply  $D_i > 0$ ,  $i = 1, \dots, M$ , and the domain qualification  $\{D_i \in \mathbb{R}_{++}\}$  is thus omitted from (60). The positivity of  $D_i$  can be shown as follows. Note that (61) implies

$$D_M > 0 \quad (63)$$

$$1 + (s_{i+1} - s_i)D_i > 0, \quad i = 1, \dots, M-1. \quad (64)$$

Consider  $i = M-1$ . Suppose  $D_{M-1} < 0$ , then (64) rearranges to  $(D_{M-1}^{-1} + s_M - s_{M-1})^{-1} < 0$ , which contradicts (62), (63):  $(D_{M-1}^{-1} + s_M - s_{M-1})^{-1} \geq D_M > 0$ . Next, suppose  $D_{M-1} = 0$ . Applying  $\lim_{D_{M-1} \rightarrow 0^+}$  on (62), the inequality becomes:  $D_M \leq 0$ , which contradicts (63). Therefore,  $D_{M-1} > 0$ , and similar arguments apply for  $D_{M-2}, \dots, D_1$ .

#### IV. RATE ALLOCATION UNDER DIFFERENT FADING DISTRIBUTIONS

In this section, we apply the optimization framework developed in Section III-C, and study the optimal rate allocation when the side-information channel is subject to different fading distributions. We first consider the scenario when the side-information channel experiences Rician fading, the pdf of which is given by

$$f_C(s) = \frac{(1+K)e^{-K}}{\bar{S}} \exp\left(-\frac{(1+K)s}{\bar{S}}\right) I_0\left(2\sqrt{\frac{K(1+K)s}{\bar{S}}}\right), \quad s \geq 0 \quad (65)$$

where  $I_0(\cdot)$  is the modified Bessel function of zeroth order, and  $\bar{S}$  is the mean channel power gain. The Rician  $K$ -factor represents the power ratio of the line-of-sight (LOS) component to the non-LOS components. Specifically, (65) reduces to Rayleigh fading for  $K = 0$ , and to no fading (i.e., constant channel power gain of  $\bar{S}$ ) for  $K = \infty$ . We discretize the channel fading pdf into  $M$  states

$$p_i = \Pr\{\text{Side information channel state } s_i \text{ is realized}\} \quad (66)$$

$$= \int_{s_i}^{s_{i+1}} f(s) ds, \quad i = 1, \dots, M \quad (67)$$

where we truncate the pdf at  $s_M$ . The quantized channel power gains are evenly spaced:  $s_i = (i-1)s_M/(M-1)$ ,  $i = 1, \dots, M$ , and  $s_{M+1} \triangleq \infty$ . In the numerical experiments, the convex optimization problems are solved using the primal-dual interior-point algorithm described in [24, Section 11.7]. The optimal rate allocation that minimizes the expected distortion  $\mathbb{E}[D]$  is shown in Fig. 5 and Fig. 6, respectively, for different values of  $K$  and  $R_X$  with  $M = 150$ . For comparison, we also show in the figures the optimal rate allocation under Nakagami fading with the pdf

$$f_N(s) = \frac{(m/\bar{S})^m s^{m-1} e^{-ms/\bar{S}}}{\Gamma(m)}, \quad s \geq 0 \quad (68)$$

where  $\Gamma(\cdot)$  is the gamma function. In Fig. 5 and Fig. 6, the Nakagami parameter  $m$  is set to be:  $m = (K+1)^2/(2K+1)$ , under which the Nakagami distribution (68) is commonly used to approximate the Rician distribution in (65) [25].

In each case of the numerical results, it is observed that the optimal rate allocation is concentrated at a *single* layer, i.e.,  $R_i^* = R_X$  for some  $i = i^*$  at  $s_{i^*}$ , while  $R_i^* = 0$  for all other  $i \neq i^*$ . The optimal primal and dual variables  $D_i^*, \lambda_i^*$  are plotted in Fig. 7 for the case of Rician fading with  $K = 32$ ,  $\bar{S} = 1$ ,  $R_X = 0.25$ ,  $\sigma_X^2 = 1$ . In this case, the rate allocation concentrates at  $s_{i^*} \approx 0.55$ , and the complementary slackness condition (110) stipulates that the corresponding dual variable be zero:  $\lambda_{i^*} = 0$ . In Fig. 5, under Rayleigh fading ( $K = 0$ ), the optimal rate allocation concentrates at the base layer (i.e.,  $s_{i^*} = 0$ ) of the source code. In the case where the side-information channel has a prominent LOS component, i.e.,

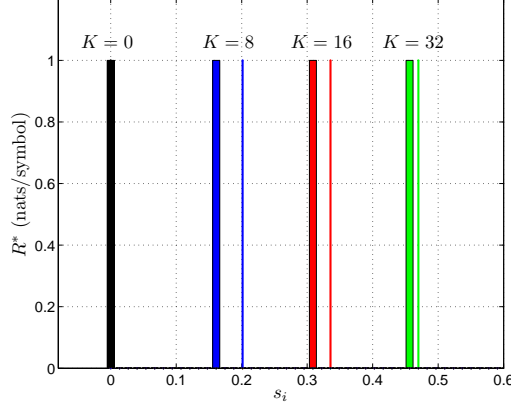


Fig. 5. Optimal rate allocation that minimizes the expected distortion  $E[D]$ . The rate allocation corresponding to Rician fading is shown in bars, and the one corresponding to Nakagami fading with  $m = (K+1)^2/(2K+1)$  is shown in lines. In each case, the optimal rate allocation is concentrated at a single layer. ( $R_X = 1$ ,  $\bar{S} = 1$ ,  $s_M = 2\bar{S}$ ,  $R_X = 1$ ,  $\sigma_X^2 = 1$ ,  $M = 150$ ).

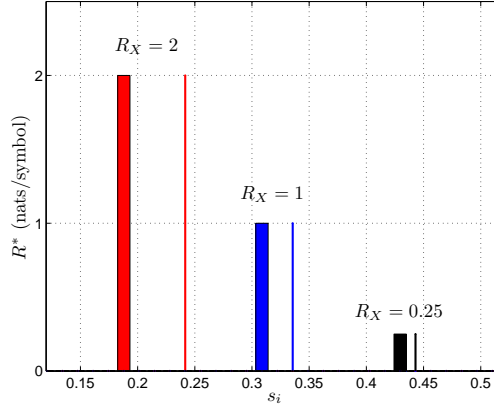


Fig. 6. Optimal rate allocation that minimizes the expected distortion  $E[D]$  with  $K = 16$  under different values of  $R_X$  (the other parameters are the same as those in Fig. 5). In each case, the optimal rate allocation is concentrated at a single layer.

when  $K$  is large,  $s_{i^*}$  increases accordingly as the channel distribution is more concentrated around  $\bar{S}$ . On the other hand, a large source-coding rate  $R_X$  decreases  $s_{i^*}$ , which implies that it is less beneficial to be opportunistic to target possible good channel conditions when  $R_X$  is large. Moreover, for each  $\bar{S}$ , Nakagami fading results in a higher  $s_{i^*}$  than its corresponding Rician fading distribution.

The minimum expectation  $E[D]^*$  that corresponds to the optimal rate allocation is shown in Fig. 8. For comparison, along with  $E[D]^*$ , in Fig. 8 we also show the distortion under different assumptions on the side information. When no side information is available, the distortion is given by the rate–distortion function for a Gaussian source [26]

$$D_{\text{No-SI}}(R_X) = \sigma_X^2 e^{-2R_X}. \quad (69)$$

In the absence of side information,  $D_{\text{No-SI}}$  is an upper bound to  $E[D]^*$ . On the other hand, when  $K = \infty$ , there is no uncertainty in the side-information channel condition with  $S = \bar{S}$ , and the distortion is given by the Wyner–Ziv [1] rate–distortion function

$$D_{\text{WZ}}(R_X) = (\sigma_X^{-2} + \bar{S})^{-1} e^{-2R_X}. \quad (70)$$

In Fig. 8, a larger  $K$  decreases the expected distortion  $E[D]^*$ , and Nakagami fading has a lower  $E[D]^*$  than the corresponding Rician fading distribution. In addition, when  $R_X$  is small,  $E[D]^*$  considerably

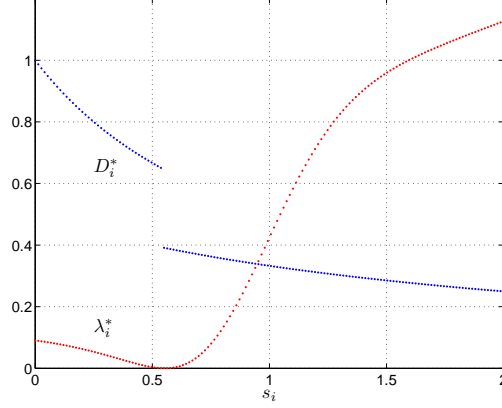


Fig. 7. Optimal primal and dual variables in the expected distortion minimization under Rician fading with  $K = 32$ ,  $\bar{S} = 1$ ,  $R_X = 0.25$ ,  $\sigma_X^2 = 1$ . The rate constraint (61) is tight, and  $\mu^* \approx 0.56$ .

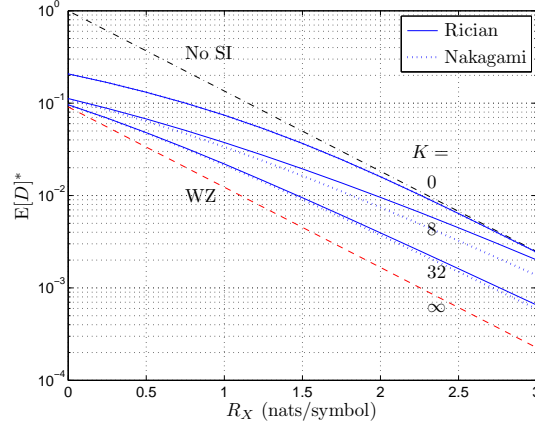


Fig. 8. Minimum expected distortion. The dash-dot line corresponds to the rate-distortion function with no side information (No SI). The dashed line ( $K = \infty$ , i.e., the side-information channel has no fading) corresponds to the Wyner-Ziv (WZ) rate-distortion function. ( $\bar{S} = 10$ ,  $s_M = 2\bar{S}$ ,  $\sigma_X^2 = 1$ ,  $M = 150$ .)

outperforms  $D_{\text{No-SI}}$  where no side information is available, as the reduction in  $\text{VAR}[X]$  from the side information at the decoder is significant. However, when  $R_X$  is large, the improvement of  $E[D]^*$  over  $D_{\text{No-SI}}$  diminishes, as most of the reduction in  $\text{VAR}[X]$  is due to the source-coding rate of  $R_X$ .

In the following, we make a remark on the *distortion exponent*  $\Delta$ , defined similarly as given in [27], which characterizes the rate of exponential decay in distortion at asymptotically large encoding rates:

$$\Delta \triangleq \lim_{R_X \rightarrow \infty} -\frac{\log E[D(R_X)]^*}{2R_X} \quad (71)$$

where  $R_X$  is the source-coding rate, and  $E[D(R_X)]^*$  is the corresponding minimum expected distortion under  $R_X$ . We note that the distortion exponent  $\Delta$  does not depend on the fading distribution  $f(s)$ , since

$$D_{\text{WZ}}(R_X) \leq E[D(R_X)]^* \leq D_{\text{No-SI}}(R_X) \quad (72)$$

$$\lim_{R_X \rightarrow \infty} -\frac{\log D_{\text{WZ}}(R_X)}{2R_X} = \lim_{R_X \rightarrow \infty} -\frac{\log D_{\text{No-SI}}(R_X)}{2R_X} = 1. \quad (73)$$

Therefore, reducing the side-information channel uncertainty (e.g., via deploying multiple antennas or through channel state information feedback) may reduce the expected distortion  $E[D(R_X)]^*$  at finite  $R_X$ , but it does not improve performance in the asymptotic regime in terms of the rate of exponential decay in the distortion as a function of the encoding rate  $R_X$ .

## V. SINGLE-LAYER RATE ALLOCATION

In the numerical experiments presented in Section IV, the optimal rate allocation concentrates at a single codeword layer under a wide class of fading distributions in the side-information channel. Motivated by such observations, in the following, we consider the single-layer rate allocation:

$$\mathcal{R}_{\bar{s}} = \{\text{encoding rate } R_X \text{ at fading state } \bar{s}\} \quad (74)$$

and investigate the conditions under which  $\mathcal{R}_{\bar{s}}$  is indeed expected-distortion-minimizing. First, we consider the case where the side-information channel has infinitely many discrete fading states; based on the discrete fading results, we then consider continuous fading distributions.

### A. Infinitely Many Discrete Fading States

Suppose the side-information channel has infinitely many discrete fading states  $\{s_1, s_2, \dots\}$  with  $0 \leq s_1 < s_2 < \dots$ . Let  $p_i$  denote the probability the side-information channel takes on state  $s_i$ , with  $\sum_{i=1}^{\infty} p_i = 1$ . Let  $\mathbf{D} \triangleq [D_1 \ D_2 \ \dots]^T$  denote the set of distortion variables induced by the rate allocation scheme. Note that the Heegard–Berger rate–distortion function extends to the set of infinitely many degraded fading states. We assume the total encoding rate  $R_X$  is finite, and hence so is the rate  $R_i$  for each fading state. For admissibility, the random coding argument in [3, Section III] applies for each of the fading states, and the converse in [3, Section VII] depends only on the total encoding rate but not the number of fading states. Furthermore, it can be shown using the same approach in [10] that the optimality of Gaussian solution for the Gaussian side information Heegard–Berger problem still holds for this scenario. The expected distortion minimization problem is given by

$$\text{minimize} \quad \sum_{i=1}^{\infty} p_i D_i \quad (75)$$

$$\text{over} \quad D_1, D_2, \dots \quad (76)$$

subject to

$$-\frac{1}{2} \sum_{i=1}^{\infty} \left( \log(D_{i-1}^{-1} + s_i - s_{i-1}) + \log D_i \right) \leq R_X \quad (77)$$

$$D_i \leq (D_{i-1}^{-1} + s_i - s_{i-1})^{-1}, \quad i = 1, 2, \dots \quad (78)$$

The rate constraint (77) follows from (48). Recall that the extended-value logarithm takes on  $-\infty$  for non-positive arguments. The constraint (77) then implies  $D_i > 0$ ; hence the domain  $D_i > 0$  is not explicitly stated in the optimization problem. In Appendix C, it is shown that (75)–(78) is a convex problem, and its KKT optimality conditions and dual function are given in Appendix D. Next, we show that the single-layer rate allocation (74) is expected-distortion-minimizing in (75)–(78) under some conditions.

Consider the single-layer rate allocation (74) with  $\bar{s} = s_i$  for some  $i$ . Relating  $\mathcal{R}_{\bar{s}}$  to the induced distortion set (48), we have

$$\bar{D}_i = \begin{cases} (\sigma_X^{-2} + s_i)^{-1}, & i : s_i < \bar{s} \\ ((\sigma_X^{-2} + \bar{s})e^{2R_X} + s_i - \bar{s})^{-1}, & i : s_i \geq \bar{s} \end{cases} \quad (79)$$

with the single-layer expected distortion given by

$$\mathbb{E}[\bar{D}] = \sum_{i=1}^{\infty} p_i \bar{D}_i. \quad (80)$$

In the notation, the overbar accent is used to represent single-layer rate allocation. To determine optimality, we apply the KKT complementary slackness condition (110) on (79) above, and it stipulates that

$$\bar{\lambda}_i = 0, \quad i : s_i = \bar{s}. \quad (81)$$

Next, we apply the KKT gradient conditions (109) on (79), (81), and the dual variables  $\bar{\lambda}_i$  are given as follows:

$$\bar{\lambda}_i = (\bar{s} - s_i) \frac{\bar{\mu} \sigma_X^{-2} + s_i}{2 \sigma_X^{-2} + \bar{s}} - \sum_{j: s_i \leq s_j < \bar{s}} p_j \left( \frac{\sigma_X^{-2} + s_i}{\sigma_X^{-2} + s_j} \right)^2, \quad i: s_i < \bar{s} \quad (82)$$

$$\begin{aligned} \bar{\lambda}_i = & -(s_i - \bar{s}) \frac{\bar{\mu}}{2} \left( 1 + \frac{s_i - \bar{s}}{(\sigma_X^{-2} + \bar{s}) e^{2R_X}} \right) \\ & + \sum_{j: \bar{s} \leq s_j < s_i} p_j \prod_{k: s_j \leq s_k < s_i} \left( 1 + \frac{s_{k+1} - s_k}{(\sigma_X^{-2} + \bar{s}) e^{2R_X} + s_k - \bar{s}} \right)^2, \quad i: s_i > \bar{s}. \end{aligned} \quad (83)$$

The dual variable  $\bar{\mu}$  is set such that the rate constraint is tight in (111):

$$\frac{\bar{\mu}}{2} = \sum_{j: s_j \geq \bar{s}} p_j \frac{(\sigma_X^{-2} + \bar{s}) e^{2R_X}}{((\sigma_X^{-2} + \bar{s}) e^{2R_X} + s_j - \bar{s})^2}. \quad (84)$$

The single-layer rate allocation optimality conditions are stated in the following proposition, and its proof is given in Appendix E.

**Proposition 2.** *In the expected distortion minimization problem where the side-information channel has discrete fading states  $\{s_1, s_2, \dots\}$ , the single-layer rate allocation (74) at  $\bar{s} = s_i$  for some  $i$  is optimal if and only if*

$$\bar{\lambda}_i \geq 0, \quad \forall i \quad (85)$$

where the variables  $\bar{\lambda}_i$  are as given in (82), (83).

Approximately, the optimality conditions in (85) correspond to the tail summation in (84) being greater than the partial sum in (82) to the left of  $\bar{s}$ , but smaller than that in (83) to the right of  $\bar{s}$ , after being weighted by their respective factors. Next, we consider the case when the side-information channel is described by a continuous fading distribution, and we show that it admits a remarkably simpler set of sufficient conditions for single-layer rate allocation optimality.

### B. Continuous Fading Distributions

When the side-information channel has a continuous fading distribution, as defined in Proposition 3 below, the corresponding expected distortion can be derived through a discretization and limiting process. The expected distortion under continuous fading is stated in the proposition, and the details of the limiting process can be found in its proof in Appendix F.

**Proposition 3.** *Suppose the side-information channel has a continuous fading distribution, i.e., the cdf  $F(s)$  is absolutely continuous with pdf  $f(s) = F'(s)$ . We further assume that the pdf  $f(s)$  is continuous. Then, under the single-layer rate allocation  $\mathcal{R}_{\bar{s}}$  in (74), the expected distortion is*

$$\mathbb{E}[\bar{D}] = \int_0^{\bar{s}} \frac{f(s)}{\sigma_X^{-2} + s} ds + \int_{\bar{s}}^{\infty} \frac{f(s)}{(\bar{s} + \sigma_X^{-2}) e^{2R_X} + s - \bar{s}} ds. \quad (86)$$

Next, Proposition 4 describes a sufficient condition for single-layer rate allocation optimality under continuous fading, and its proof is given in Appendix G. In the proposition, note that the optimality condition (87) is motivated by considering the limiting case of the optimality conditions (85) in Proposition 2.

**Proposition 4.** *In Proposition 3,  $\mathcal{R}_{\bar{s}}$  is expected-distortion-minimizing if*

$$\bar{\lambda}(s) \geq 0, \quad s \geq 0 \quad (87)$$

where

$$\bar{\lambda}(s) = (\bar{s} - s) \frac{\bar{\mu} \sigma_X^{-2} + s}{2 \sigma_X^{-2} + \bar{s}} - \int_s^{\bar{s}} f(t) \left( \frac{\sigma_X^{-2} + s}{\sigma_X^{-2} + t} \right)^2 dt, \quad 0 \leq s < \bar{s} \quad (88)$$

$$\begin{aligned} \bar{\lambda}(s) = & -(s - \bar{s}) \frac{\bar{\mu}}{2} \left( 1 + \frac{s - \bar{s}}{(\sigma_X^{-2} + \bar{s}) e^{2R_X}} \right) \\ & + \int_{\bar{s}}^s f(t) \left( 1 + \frac{s - t}{(\sigma_X^{-2} + \bar{s}) e^{2R_X} + t - \bar{s}} \right)^2 dt, \quad s \geq \bar{s} \end{aligned} \quad (89)$$

and

$$\frac{\bar{\mu}}{2} = \int_{\bar{s}}^{\infty} f(s) \frac{(\sigma_X^{-2} + \bar{s}) e^{2R_X}}{((\sigma_X^{-2} + \bar{s}) e^{2R_X} + s - \bar{s})^2} ds. \quad (90)$$

### C. Quasiconcave Probability Density Functions

In the following, we show that fading distributions with continuous, quasiconcave pdfs satisfy the single-layer rate allocation optimality conditions (87) in Proposition 4. A function  $g(x)$  is quasiconcave, or referred to as unimodal, if its superlevel sets  $\{x \mid g(x) \geq \alpha\}$ , for all  $\alpha$ , are convex. Notably, most common wireless channel fading distributions have pdfs that are continuous and quasiconcave: e.g., Rayleigh, Rician, Nakagami, and log-normal.

First, we set out the procedure to identify  $\bar{s}^*$ , the state at which the encoding rate  $R_X$  is to be concentrated. Let  $f(s)$  denote the pdf of the fading distribution of the side-information channel, and we assume that  $f(s)$  is continuous and quasiconcave. We denote the superlevel set of  $f(s)$  by the interval  $[s_a, s_b] \triangleq \{s \mid f(s) \geq \alpha\}$ , for some nonnegative scalar  $\alpha \geq 0$ . Specifically, we choose an  $\alpha^*$  such that the following relationship holds:

$$\int_{s_a(\alpha^*)}^{\infty} \frac{f(s) - \alpha^*}{((\sigma_X^{-2} + s_a(\alpha^*)) e^{2R_X} + s - s_a(\alpha^*))^2} ds = 0. \quad (91)$$

Note that the left-hand side of (91) varies continuously from positive to negative as a candidate  $\alpha$  ranges from 0 to  $\max f(s)$ ; therefore, there exists an  $\alpha^*$  that equates the expression to zero. We set  $\bar{s}^*$  to be the left endpoint of the superlevel set induced by  $\alpha^*$ :

$$\bar{s}^* = s_a(\alpha^*). \quad (92)$$

With the single-layer rate allocation target  $\bar{s}^*$  properly defined, the following proposition identifies the class of fading distributions under which this rate allocation is optimal; its proof is presented in Appendix H.

**Proposition 5.** *Suppose the side-information channel has a continuous fading distribution with pdf  $f(s)$ . If  $f(s)$  is continuous and quasiconcave, then the single-layer rate allocation  $\mathcal{R}_{\bar{s}^*}$ , with  $\bar{s}^*$  as given in (92), is expected-distortion-minimizing.*

As an example, consider minimizing the expected distortion under Rician fading ( $K = 32$ ,  $R_X = 0.25$ ,  $\bar{S} = 1$ ,  $\sigma_X^2 = 1$ ), for which the KKT optimality conditions are illustrated in Fig. 9. At optimality,  $\alpha^* = \bar{\mu}/2$ . In the figure,  $[s_a, s_b]$  is the  $\bar{\mu}/2$ -superlevel set of  $f(s)$ , and the regions between  $f(s)$  and  $\bar{\mu}/2$  are shaded and labeled (A) and (B), respectively, for  $s_a \leq s \leq s_b$  and  $s > s_b$ . The choice of  $\alpha^*$  in (91), which leads to  $\bar{s}^*$  in (92), corresponds to the area of (A), weighted by  $w_2^{-1}(s)$ , being equal to the area of (B), weighted by  $w_2(s)$ , where  $w_2(s)$  is as defined in (143) in Appendix H. Solving (91) numerically, the resulting single-layer rate allocation target  $\bar{s}^*$  is plotted in Fig. 10 for different values of  $K$  and  $R_X$ .

As a special case of fading distributions with continuous and quasiconcave pdfs, let us consider Rayleigh fading. Its pdf is given by

$$f_R(s) = (1/\bar{S}) e^{s/\bar{S}}, \quad s \geq 0 \quad (93)$$

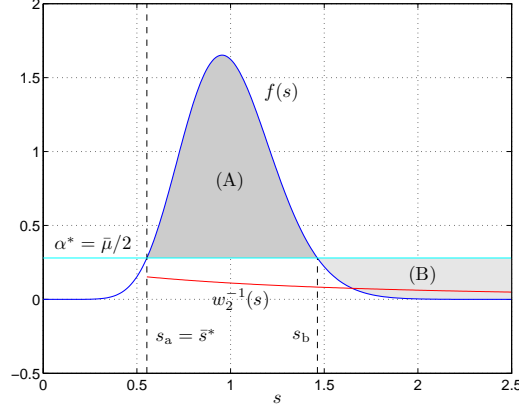


Fig. 9. KKT optimality conditions under Rician fading ( $K = 32$ ,  $R_X = 0.25$ ,  $\bar{S} = 1$ ,  $\sigma_X^2 = 1$ ). Note that the area of (A), weighted by  $w_2^{-1}(s)$ , is equal to the area of (B), weighted by  $w_2(s)$ , where  $w_2(s)$  is as defined in (143) in Appendix H.

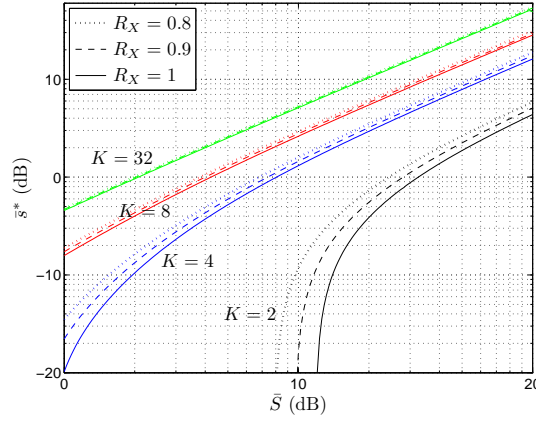


Fig. 10. Optimal single-layer rate allocation target  $\bar{s}^*$  under Rician fading ( $\bar{S} = 1$ ,  $\sigma_X^2 = 1$ ).

where  $\bar{S}$  is the average channel power gain. Recognizing that any nonempty superlevel set of  $f_R(s)$  begins at  $s_a = 0$ , we have the following corollary:

**Corollary 2.** *When the side-information channel is under Rayleigh fading (93), the single-layer rate allocation target  $\bar{s}^* = 0$  is optimal, and the corresponding minimum expected distortion is*

$$\mathbb{E}[\bar{D}_R]^* = \int_0^\infty \frac{(1/\bar{S})e^{s/\bar{S}}}{\sigma_X^{-2}e^{2R_X} + s} ds = (1/\bar{S})e^{C/\bar{S}}\mathbb{E}_1(C/\bar{S}) \quad (94)$$

where  $C \triangleq \sigma_X^{-2}e^{2R_X}$ , and  $\mathbb{E}_1(\cdot)$  is the exponential integral  $\mathbb{E}_1(x) \triangleq \int_x^\infty \frac{e^{-t}}{t} dt$ .

Therefore, under Rayleigh fading, the source-coding scheme does not depend on  $\bar{S}$ ,  $R_X$ , and  $\sigma_X^2$ . It is optimal to concentrate the entire encoding rate  $R_X$  at the base layer  $\bar{s}^* = 0$ , i.e., the source is encoded as if the side information was absent.

## VI. CONCLUSIONS

We studied the problem of optimal rate allocation and distortion minimization for Gaussian source coding under squared error distortion, when the uncompressed source symbol is also conveyed over a fading side-information channel. The encoder knows the fading channel distribution but not its realization. A layered encoding strategy is used, with each codeword layer targeting the realization of a given



fading state. When the side-information channel has two discrete fading states, we derived closed-form expressions for the optimal rate allocation among the fading states and the corresponding minimum expected distortion. The optimal rate allocation is conservative: rate is allocated to the higher layer only if the better fading state is highly probable. For the case of multiple discrete fading states, the minimum expected distortion was shown to be the solution of a convex optimization problem. We derived an efficient representation for the Heegard–Berger rate–distortion function, under which the number of variables and constraints in the optimization problem is linear in the number of fading states.

Next, we considered single-layer rate allocation, and identified the conditions under which such allocation is expected-distortion–minimizing, for the respective cases of discrete as well as continuous fading in the side-information channel. Under continuous fading, we showed that these optimality conditions are satisfied by distributions with continuous and quasiconcave pdfs, e.g., Rayleigh, Rician, Nakagami, and log-normal. Moreover, for Rayleigh fading, the optimal rate allocation concentrates at the base layer: i.e., the source is encoded as if the side information was absent.

In this paper, we focused on fading distributions for which the optimal rate allocation consists of a single codeword layer. For fading distributions with continuous pdfs that are not quasiconcave, we conjecture that the expected-distortion–minimizing rate allocation remains discrete but may comprise multiple codeword layers. By contrast, a continuum of codeword layers is in general necessary when maximizing expected capacity or minimizing expected distortion over a slowly fading channel [2], [13], [14]. In a broader context, the techniques for source coding under fading side-information channels may be applied to improve quantize-and-forward schemes [28] in wireless network transmissions, where the side information represents the auxiliary signals forwarded by a cooperating user as received via a fading channel. In those cases where different distortion measures other than squared error are considered, however, different conclusions regarding the optimal number of source-coding layers may result.

## APPENDIX A PROOF OF LEMMA 1

*Proof:* The lemma follows from the minimum mean square error (MMSE) estimate of Gaussian random variables. Let  $X, \mathbf{W}$ , where  $\mathbf{W} \triangleq [W_1 \dots W_k]^T$ , be distributed as

$$\begin{bmatrix} \mathbf{W} \\ X \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_{\mathbf{W}} \\ \mu_X \end{bmatrix}, \begin{bmatrix} \Sigma_{\mathbf{W}} & \Sigma_{\mathbf{W}X} \\ \Sigma_{\mathbf{W}X}^T & \sigma_X^2 \end{bmatrix}\right). \quad (95)$$

The conditional distribution is Gaussian, and the corresponding variance is

$$\text{VAR}[X|Y, \mathbf{W}] = \sigma_X^2 - \begin{bmatrix} \Sigma_{\mathbf{W}X} \\ \sqrt{s}\sigma_X^2 \end{bmatrix}^T \begin{bmatrix} \Sigma_{\mathbf{W}} & \sqrt{s}\Sigma_{\mathbf{W}X} \\ \sqrt{s}\Sigma_{\mathbf{W}X}^T & s\sigma_X^2 + 1 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma_{\mathbf{W}X} \\ \sqrt{s}\sigma_X^2 \end{bmatrix} \quad (96)$$

$$= \frac{\sigma_X^2 - \Sigma_{\mathbf{W}X}^T \Sigma_{\mathbf{W}}^{-1} \Sigma_{\mathbf{W}X}}{1 + s(\Sigma_{\mathbf{W}X}^T \Sigma_{\mathbf{W}}^{-1} \Sigma_{\mathbf{W}X})} \quad (97)$$

$$= (\text{VAR}[X|\mathbf{W}]^{-1} + s)^{-1}. \quad (98)$$

■

## APPENDIX B PROOF OF PROPOSITION 1

*Proof:* Each of the inequality constraints in (55)–(57) is convex: i.e., it is of the form

$$c_x(D_1, \dots, D_M, V_1, \dots, V_M) \leq c_c(D_1, \dots, D_M, V_1, \dots, V_M) \quad (99)$$

where  $c_x(\cdot)$  is convex in  $D_1, \dots, V_M$ , and  $c_c(\cdot)$  is concave in  $D_1, \dots, V_M$ . In particular, in (56), the right-hand side of each inequality constraint depends on only  $V_{i-1}$ . Being twice-differentiable, its concavity can be verified by the second-order condition

$$\frac{d^2}{dV_{i-1}^2}(V_{i-1}^{-1} + s_i - s_{i-1})^{-1} = \frac{-2(s_i - s_{i-1})}{(1 + (s_i - s_{i-1})V_{i-1})^3} \quad (100)$$

which is negative since  $s_i > s_{i-1}$ ,  $V_{i-1} > 0$  as given in the problem formulation, for  $i = 2, \dots, M$ . Therefore, in (53)–(57), we minimize a convex function subject to a set of convex inequality constraints, which is a convex optimization problem. ■

## APPENDIX C

### CONVEX OPTIMIZATION PROBLEM (75)–(78)

We show that (75)–(78) is a convex problem. For convenience, let us denote the left-hand side of (77) as follows:

$$\Phi(\mathbf{D}) \triangleq -\frac{1}{2} \sum_{i=1}^{\infty} \left( \log(D_{i-1}^{-1} + s_i - s_{i-1}) + \log D_i \right). \quad (101)$$

We show that  $\Phi(\mathbf{D})$  is convex in  $\mathbf{D}$ . Let  $\mathbf{D}^{(1)}, \mathbf{D}^{(2)} \succ 0$  be two sets of element-wise positive distortion variables. By the mean-value theorem, for some  $\gamma \in [0, 1]$ , we have

$$\begin{aligned} \Phi(\mathbf{D}^{(2)}) &= \Phi(\mathbf{D}^{(1)}) + \sum_{i=1}^{\infty} \frac{\partial \Phi(\mathbf{D}^{(1)})}{\partial D_i} (D_i^{(2)} - D_i^{(1)}) \\ &\quad + \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{\partial^2 \Phi(\mathbf{D}^{(1)} + \gamma(\mathbf{D}^{(2)} - \mathbf{D}^{(1)}))}{\partial D_i \partial D_j} (D_i^{(2)} - D_i^{(1)}) (D_j^{(2)} - D_j^{(1)}) \end{aligned} \quad (102)$$

$$\geq \Phi(\mathbf{D}^{(1)}) + \sum_{i=1}^{\infty} \frac{\partial \Phi(\mathbf{D}^{(1)})}{\partial D_i} (D_i^{(2)} - D_i^{(1)}) \quad (103)$$

which corresponds to the first-order convexity condition. In (103), the inequality follows from:

$$\frac{\partial^2 \Phi}{\partial D_i \partial D_j} = \begin{cases} \frac{1}{2} \left( \frac{s_{i+1} - s_i}{1 + (s_{i+1} - s_i) D_i} \right)^2, & i = j \\ 0, & i \neq j \end{cases} \quad (104)$$

which shows that the last term in (102) is nonnegative. Each constraint in (78) is shown to be a convex set in (100). Convexity is preserved under the intersection of a possibly infinite number of convex sets [24, Section 2.3.1]. The linear objective function in (75) is convex; therefore, (75)–(78) is a convex optimization problem.

## APPENDIX D

### KKT OPTIMALITY CONDITIONS AND DUAL FUNCTION FOR (75)–(78)

In the following, we characterize the KKT optimality conditions [29] for the expected distortion minimization problem (75)–(78). First, we form the Lagrangian

$$L(\mathbf{D}, \boldsymbol{\lambda}, \mu) = L_1(\mathbf{D}, \boldsymbol{\lambda}, \mu) + L_2(\mathbf{D}, \boldsymbol{\lambda}, \mu) + L_3(\mathbf{D}, \boldsymbol{\lambda}, \mu) \quad (105)$$

$$L_1(\mathbf{D}, \boldsymbol{\lambda}, \mu) \triangleq \sum_{i=1}^{\infty} p_i D_i \quad (106)$$

$$L_2(\mathbf{D}, \boldsymbol{\lambda}, \mu) \triangleq \sum_{i=1}^{\infty} \lambda_i \left( D_i - (D_{i-1}^{-1} + s_i - s_{i-1})^{-1} \right) \quad (107)$$

$$L_3(\mathbf{D}, \boldsymbol{\lambda}, \mu) \triangleq \mu \left( -\frac{1}{2} \sum_{i=1}^{\infty} \left( \log(D_{i-1}^{-1} + s_i - s_{i-1}) + \log D_i \right) - R_X \right) \quad (108)$$

where  $\boldsymbol{\lambda} \triangleq [\lambda_1 \ \lambda_2 \ \dots]^T$ , and  $\mu, \boldsymbol{\lambda}$  are the Lagrange multipliers, or dual variables, associated with inequalities (77), (78), respectively. At optimality, the gradient of the Lagrangian vanishes:

$$0 = \frac{\partial L}{\partial D_i} = p_i + \lambda_i - \frac{\lambda_{i+1}}{(1 + (s_{i+1} - s_i)D_i)^2} - \frac{\mu}{2} \frac{s_{i+1} - s_i}{1 + (s_{i+1} - s_i)D_i}, \quad i = 1, 2, \dots \quad (109)$$

and the complementary slackness conditions hold:

$$0 = \lambda_i \left( D_i - (D_{i-1}^{-1} + s_i - s_{i-1})^{-1} \right), \quad i = 1, 2, \dots \quad (110)$$

$$0 = \mu \left( -\frac{1}{2} \sum_{i=1}^{\infty} \left( \log(D_{i-1}^{-1} + s_i - s_{i-1}) + \log D_i \right) - R_X \right). \quad (111)$$

The primal feasibility conditions are given by (77), (78), and the dual feasibility conditions are

$$\mu \geq 0, \quad \lambda_i \geq 0, \quad i = 1, 2, \dots \quad (112)$$

Together, (77), (78), and (109)–(112) are the necessary and sufficient conditions for optimality in the convex problem (75)–(78).

The dual function of the expected distortion minimization problem is given by

$$G(\boldsymbol{\lambda}, \mu) = \inf_{\mathbf{D}} L(\mathbf{D}, \boldsymbol{\lambda}, \mu). \quad (113)$$

Let  $\hat{\mathbf{D}} \triangleq [\hat{D}_1 \ \hat{D}_2 \ \dots]^T$  denote the Lagrangian-minimizing  $\mathbf{D}$  in (113). For  $\boldsymbol{\lambda} \succeq 0$ ,  $\mu \geq 0$ ,  $\hat{\mathbf{D}}$  can be determined from the KKT gradient conditions (109):

$$\hat{D}_i(\boldsymbol{\lambda}, \mu) = \frac{\mu(s_{i+1} - s_i) - 4(p_i + \lambda_i) + \sqrt{\mu^2(s_{i+1} - s_i)^2 + 16(p_i + \lambda_i)\lambda_{i+1}}}{4(s_{i+1} - s_i)(p_i + \lambda_i)}, \quad (114)$$

$$i = 1, 2, \dots$$

The dual function provides a lower bound to the solution of the optimization problem. Let  $J^*$  denote the minimum expected distortion in (75)–(78). For any  $\tilde{\mathbf{D}}$  that is primal feasible by (77), (78), and any dual feasible  $\tilde{\boldsymbol{\lambda}} \succeq 0$ ,  $\tilde{\mu} \geq 0$ , we have

$$G(\tilde{\boldsymbol{\lambda}}, \tilde{\mu}) \leq J^* \leq J(\tilde{\mathbf{D}}). \quad (115)$$

Furthermore, since the optimization problem is convex and Slater's condition holds, the duality gap is zero:

$$G(\boldsymbol{\lambda}^*, \mu^*) = J^* = J(\mathbf{D}^*) \quad (116)$$

where  $\mathbf{D}^*$  and  $\boldsymbol{\lambda}^*, \mu^*$  respectively denote the optimal primal and dual variables that satisfy the KKT optimality conditions (77), (78), and (109)–(112).

## APPENDIX E PROOF OF PROPOSITION 2

*Proof:* In the convex optimization problem (75)–(78), the single-layer rate allocation  $\mathcal{R}_{\bar{s}}$  is primal feasible by (77), (78). Through construction of the dual variables  $\bar{\lambda}_i, \bar{\mu}$  in (81)–(84), the KKT gradient conditions (109) and the complementary slackness conditions (110), (111) are satisfied. The dual feasibility condition  $\bar{\mu} \geq 0$  follows from the non-negativity of each term in the summation in (84). Besides (81), dual feasibility of  $\bar{\lambda}_i$  as given in (85) are the remaining necessary and sufficient conditions for optimality. ■

## APPENDIX F

### PROOF OF PROPOSITION 3

*Proof:* We assume the side-information channel has a fading distribution with a continuous pdf  $f(s)$ , and we partition the continuum of fading state into subintervals  $[u_0, u_1], [u_1, u_2], \dots$  of lengths  $\Delta u_1, \Delta u_2, \dots$ , with  $\Delta u_i \triangleq u_i - u_{i-1}$ ,  $u_0 = 0$ . Let us assume  $\bar{s} > 0$ ; in the case of  $\bar{s} = 0$ , it is interpreted as  $\lim_{\bar{s} \rightarrow 0^+}$ . The partition boundaries are chosen such that  $u_{\bar{i}} = \bar{s}$  for some index  $\bar{i}$ . The fading realization falls within each subinterval with probability  $p_i = \int_{u_{i-1}}^{u_i} f(s) ds$ , for  $i = 1, 2, \dots$ . We consider two sets of discretized fading states. In the first set, for each subinterval, we discretize the fading state to its worst-case realization:  $\{s_i\} = \mathbf{u}_w \triangleq \{u_0, u_1, \dots\}$ , i.e., we have fading state  $u_{i-1}$  with probability  $p_i$ . In the second set, in a similar manner, for each subinterval we discretize the fading state to its best-case realization:  $\{s_i\} = \mathbf{u}_b \triangleq \{u_1, u_2, \dots\}$ , where probability  $p_i$  is associated with fading state  $u_i$ .

Next, we apply the single-layer rate allocation  $\mathcal{R}_{\bar{s}}$  to the worst-case states  $\mathbf{u}_w$ , the continuous distribution  $f(s)$ , and the best-case states  $\mathbf{u}_b$ . Since the decoder may arbitrarily add noise to the side-information channel, the expected distortions under  $\mathcal{R}_{\bar{s}}$  are ordered as follows:

$$J_b(\mathcal{R}_{\bar{s}}) \leq J_c(\mathcal{R}_{\bar{s}}) \leq J_w(\mathcal{R}_{\bar{s}}) \quad (117)$$

where  $J_b(\mathcal{R}_{\bar{s}})$ ,  $J_c(\mathcal{R}_{\bar{s}})$ ,  $J_w(\mathcal{R}_{\bar{s}})$  represent the expected distortions under  $\mathcal{R}_{\bar{s}}$  for the best-case states, continuous distribution, and worst-case states, respectively.

Finally, consider the inequalities in (117) in the limit of small subintervals as  $\|\mathbf{u}\| \rightarrow 0$ , where  $\|\mathbf{u}\| \triangleq \max \Delta u_i$ . As  $f(s)$  is continuous, by the mean-value theorem, we have

$$p_i = \int_{u_{i-1}}^{u_i} f(s) ds = (u_i - u_{i-1})f(u_i^*) \quad (118)$$

for some  $u_i^* \in [u_{i-1}, u_i]$  that achieves the mean value of  $f(\cdot)$  within the partition  $[u_{i-1}, u_i]$ . Apply (118) to (79), (80), we have

$$\lim_{\|\mathbf{u}\| \rightarrow 0} J_b(\mathcal{R}_{\bar{s}}) = \lim_{\|\mathbf{u}\| \rightarrow 0} \sum_{i=1}^{\bar{i}-1} \frac{(u_i - u_{i-1})f(u_i^*)}{\sigma_X^{-2} + u_i} + \sum_{i=\bar{i}}^{\infty} \frac{(u_i - u_{i-1})f(u_i^*)}{(\sigma_X^{-2} + \bar{s})e^{2R_X} + u_i - \bar{s}} \quad (119)$$

$$= \int_0^{\bar{s}} \frac{f(u)}{\sigma_X^{-2} + u} du + \int_{\bar{s}}^{\infty} \frac{f(u)}{(\sigma_X^{-2} + \bar{s})e^{2R_X} + u - \bar{s}} du \quad (120)$$

$$= \lim_{\|\mathbf{u}\| \rightarrow 0} \sum_{i=1}^{\bar{i}-1} \frac{(u_i - u_{i-1})f(u_i^*)}{\sigma_X^{-2} + u_{i-1}} + \sum_{i=\bar{i}}^{\infty} \frac{(u_i - u_{i-1})f(u_i^*)}{(\sigma_X^{-2} + \bar{s})e^{2R_X} + u_{i-1} - \bar{s}} \quad (121)$$

$$= \lim_{\|\mathbf{u}\| \rightarrow 0} J_w(\mathcal{R}_{\bar{s}}). \quad (122)$$

Since from (119)–(122) the limits of  $J_b(\mathcal{R}_{\bar{s}})$ ,  $J_w(\mathcal{R}_{\bar{s}})$  coincide and equal (120), from (117)  $J_c(\mathcal{R}_{\bar{s}})$  also equals this limit, which is given by (86) in the main body of the text. ■

## APPENDIX G

### PROOF OF PROPOSITION 4

*Proof:* Consider the set of best-case and worst-case discretized fading states described in Appendix F. We have the following inequalities:

$$G_b(\bar{\lambda}(\mathbf{u}_b), \bar{\mu}) \stackrel{(a)}{\leq} J_b^* \stackrel{(b)}{\leq} J_c^* \stackrel{(c)}{\leq} J_w^* \stackrel{(d)}{\leq} J_w(\mathcal{R}_{\bar{s}}) \quad (123)$$

where  $J_b^*$ ,  $J_c^*$ ,  $J_w^*$  denote the minimum expected distortion under the best-case states, continuous distribution, and worst-case states, respectively;  $G_b(\cdot)$  is the dual function under the best-case states;  $\bar{\lambda}(\mathbf{u}_b)$  denotes applying  $\bar{\lambda}(\cdot)$  on  $\mathbf{u}_b$  element-wise; and  $J_w(\mathcal{R}_{\bar{s}})$  is as defined in (117). In (123), (b), (c) follow from the ordering on the sets of the fading states, and (a), (d) follow from the duality bounds (115):  $\bar{\lambda}(\mathbf{u}_b)$

is dual feasible by the assumption  $\bar{\lambda}(s) \geq 0$ ;  $\bar{\mu}$  is dual feasible since the integrand in (90) is nonnegative; and  $\mathcal{R}_{\bar{s}}$ , being a valid rate allocation, induces a set of distortion variables that is primal feasible.

Next, we consider  $G_b(\bar{\lambda}(\mathbf{u}_b), \bar{\mu})$  in the limit of  $\|\mathbf{u}\| \rightarrow 0$ . Writing (113) in terms of (105), recall that

$$G_b(\bar{\lambda}(\mathbf{u}_b), \bar{\mu}) = L_1(\hat{\mathbf{D}}(\bar{\lambda}(\mathbf{u}_b), \bar{\mu}), \bar{\lambda}(\mathbf{u}_b), \bar{\mu}) + L_2(\hat{\mathbf{D}}(\bar{\lambda}(\mathbf{u}_b), \bar{\mu}), \bar{\lambda}(\mathbf{u}_b), \bar{\mu}) + L_3(\hat{\mathbf{D}}(\bar{\lambda}(\mathbf{u}_b), \bar{\mu}), \bar{\lambda}(\mathbf{u}_b), \bar{\mu}) \quad (124)$$

where  $\hat{\mathbf{D}}(\bar{\lambda}(\mathbf{u}_b), \bar{\mu})$  is as given in (114). Note that each  $\hat{D}_i(\bar{\lambda}(\mathbf{u}_b), \bar{\mu})$  depends on  $\bar{\lambda}(u_i)$  and  $\bar{\lambda}(u_{i+1})$ . Since  $\bar{\lambda}(s)$  is continuous and differentiable over  $s > 0$ , its values at the partition boundaries are related through the mean-value theorem:

$$\bar{\lambda}(u_{i+1}) = \bar{\lambda}(u_i) + (u_{i+1} - u_i)\bar{\lambda}'(u_i^\dagger) \quad (125)$$

for some  $u_i^\dagger \in [u_i, u_{i+1}]$ . Consider the first term in (124):

$$\lim_{\|\mathbf{u}\| \rightarrow 0} L_1(\hat{\mathbf{D}}(\bar{\lambda}(\mathbf{u}_b), \bar{\mu}), \bar{\lambda}(\mathbf{u}_b), \bar{\mu}) = \lim_{\|\mathbf{u}\| \rightarrow 0} \sum_{i=1}^{\infty} p_i \hat{D}_i(\bar{\lambda}(\mathbf{u}_b), \bar{\mu}) \quad (126)$$

$$= \int_0^\infty f(u) \frac{\bar{\lambda}'(u) - f(u) + \bar{\mu}/2}{2\bar{\lambda}(u)} du \quad (127)$$

$$= \int_0^{\bar{s}} \frac{f(u)}{\sigma_X^{-2} + u} du + \int_{\bar{s}}^\infty \frac{f(u)}{(\sigma_X^{-2} + \bar{s})e^{2R_X} + u - \bar{s}} du \quad (128)$$

$$= \lim_{\|\mathbf{u}\| \rightarrow 0} J_w(\mathcal{R}_{\bar{s}}) \quad (129)$$

where (127) follows from applying (118), (125) to (126); and (129) follows from (120). In (127),  $\bar{\lambda}(\cdot)$  is as defined in (88), (89), and its derivative is given as follows:

$$\bar{\lambda}'(s) = \frac{\bar{\mu}\bar{s} - 2s - \sigma_X^{-2}}{2(\sigma_X^{-2} + \bar{s})} + f(s) - \int_s^{\bar{s}} 2f(t) \frac{\sigma_X^{-2} + s}{(\sigma_X^{-2} + t)^2} dt, \quad 0 \leq s < \bar{s} \quad (130)$$

$$\begin{aligned} \bar{\lambda}'(s) &= -\frac{\bar{\mu}}{2} - \frac{\bar{\mu}(s - \bar{s})}{(\sigma_X^{-2} + \bar{s})e^{2R_X}} + f(s) \\ &\quad + \int_{\bar{s}}^s 2f(t) \frac{(\sigma_X^{-2} + \bar{s})e^{2R_X} + s - \bar{s}}{((\sigma_X^{-2} + \bar{s})e^{2R_X} + t - \bar{s})^2} dt, \quad s \geq \bar{s}. \end{aligned} \quad (131)$$

Substituting (88), (89), (130), (131) into (127), note that the factor in the integrand (defined as  $D(s)$  below) simplifies to the following expressions which are independent of  $f(s)$ :

$$D(s) \triangleq \frac{\bar{\lambda}'(s) - f(s) + \bar{\mu}/2}{2\bar{\lambda}(s)} \quad (132)$$

$$= \begin{cases} (\sigma_X^{-2} + s)^{-1}, & 0 \leq s < \bar{s} \\ ((\sigma_X^{-2} + \bar{s})e^{2R_X} + s - \bar{s})^{-1}, & s \geq \bar{s}. \end{cases} \quad (133)$$

For the second term in (124), note that  $\bar{\lambda}(u_{\bar{i}}) = 0$ , and hence  $i = \bar{i}$  can be excluded from the following summation:

$$\begin{aligned} &\lim_{\|\mathbf{u}\| \rightarrow 0} L_2(\hat{\mathbf{D}}(\bar{\lambda}(\mathbf{u}_b), \bar{\mu}), \bar{\lambda}(\mathbf{u}_b), \bar{\mu}) \\ &= \lim_{\|\mathbf{u}\| \rightarrow 0} \sum_{i \neq \bar{i}} \bar{\lambda}(u_i) \left( \hat{D}_i(\bar{\lambda}(\mathbf{u}_b), \bar{\mu}) - \left( \hat{D}_{i-1}(\bar{\lambda}(\mathbf{u}_b), \bar{\mu})^{-1} + u_i - u_{i-1} \right)^{-1} \right) \end{aligned} \quad (134)$$

$$= \int_{\bar{s}}^\infty \bar{\lambda}(u) (D'(u) + D(u)^2) du \quad (135)$$

$$= 0 \quad (136)$$

where  $\hat{D}_0(\cdot) \triangleq \sigma_X^2$ , and (118), (125) are applied in (134); the notation  $\int_{\sqrt{s}} \cdot du$  in (135) is defined as

$$\int_{\sqrt{s}} g(u) du \triangleq \int_0^{\bar{s}^-} g(u) du + \int_{\bar{s}^+}^\infty g(u) du \quad (137)$$

convergence of the partition summation follows from  $f(s)$ ,  $\bar{\lambda}(s)$ ,  $\bar{\lambda}'(s)$  being continuous when the mean-value theorem is applied to  $\hat{D}_i(\cdot)$  in (114); and (136) follows from  $D'(s) + D(s)^2 = 0$  over the intervals  $0 < s < \bar{s}$  and  $s > \bar{s}$ . Finally, consider the third term in (124):

$$\begin{aligned} & \lim_{\|\mathbf{u}\| \rightarrow 0} L_3(\hat{\mathbf{D}}(\bar{\lambda}(\mathbf{u}_b), \bar{\mu}), \bar{\lambda}(\mathbf{u}_b), \bar{\mu}) \\ &= \lim_{\|\mathbf{u}\| \rightarrow 0} \bar{\mu} \left( -\frac{1}{2} \sum_{i=1}^{\infty} \left( \log \left( \hat{D}_{i-1}(\bar{\lambda}(\mathbf{u}_b), \bar{\mu})^{-1} + u_i - u_{i-1} \right) + \log \hat{D}_i(\bar{\lambda}(\mathbf{u}_b), \bar{\mu}) \right) - R_X \right) \end{aligned} \quad (138)$$

$$= \frac{\bar{\mu}}{2} \int_{\sqrt{s}} D(u) + \frac{D'(u)}{D(u)} du + \bar{\mu}(R_X - R_X) \quad (139)$$

$$= 0 \quad (140)$$

where (118), (125) are applied in (138); the last term in (139) follows from noting that  $\hat{D}_{i-1}(\cdot) \rightarrow (\sigma_X^{-2} + \bar{s})^{-1}$  and  $\hat{D}_i(\cdot) \rightarrow ((\sigma_X^{-2} + \bar{s})e^{2R_X})^{-1}$ ; the notation  $\int_{\sqrt{s}} \cdot du$  is as given in (137), and its convergence follows from continuity of  $f(s)$ ,  $\bar{\lambda}(s)$ ,  $\bar{\lambda}'(s)$ ; (140) follows from  $D(s) + D'(s)/D(s) = 0$  over the intervals  $0 < s < \bar{s}$  and  $s > \bar{s}$ . Combining (124), (129), (136), (140), we have:  $\lim_{\Delta u \rightarrow 0} G_b(\bar{\lambda}(\mathbf{u}_b), \bar{\mu}) = \lim_{\Delta u \rightarrow 0} J_w(\mathcal{R}_{\bar{s}})$ . Since the inequalities in (123) coincide in the limit, the minimum expected distortion  $J_c^*$  is given by (128), which is achievable by  $\mathcal{R}_{\bar{s}}$  by Proposition 3. ■

## APPENDIX H PROOF OF PROPOSITION 5

*Proof:* We show that the single-layer rate allocation satisfies the optimality conditions (87) in Proposition 4. It will be useful to compare  $f(s)$  against  $\bar{\mu}/2$ , so we bring  $\bar{\mu}/2$  inside in the integrals in (88), (89), and rewrite the expressions as

$$\frac{\bar{\lambda}(s)}{w_n(s)} = \int_{\bar{s}}^s \frac{f(t) - \bar{\mu}/2}{w_n(t)} dt \quad (141)$$

where  $n = 1$  for  $0 \leq s < \bar{s}$ ,  $n = 2$  for  $s \geq \bar{s}$ , with  $w_n(s) \geq 0$  given as follows:

$$w_1(s) = (\sigma_X^{-2} + s)^2 \quad (142)$$

$$w_2(s) = ((\sigma_X^{-2} + \bar{s})e^{2R_X} + s - \bar{s})^2. \quad (143)$$

We now show the non-negativity of (141). Solving for  $\alpha^*$  in (91) and comparing its solution against the right-hand side of (90), we recognize that  $\alpha^* = \bar{\mu}/2$ . Thus for  $0 \leq s < \bar{s}^*$ , being outside of the superlevel set  $[s_a, s_b]$ , we have  $f(s) < \bar{\mu}/2$  in the integrand in (141): hence  $\bar{\lambda}(s) > 0$ . For  $s \geq \bar{s}^*$  but  $s \leq s_b$ , we have  $f(s) \geq \bar{\mu}/2$ : hence  $\bar{\lambda}(s) \geq 0$ . Finally, for  $s > s_b$ , note that  $\bar{\lambda}(s)/w_2(s)$  is monotonically decreasing, but it never descends below zero:  $\lim_{s \rightarrow \infty} \bar{\lambda}(s)/w_2(s) = 0$ , which is a consequence of how  $\alpha^*$  is constructed as specified in (91). ■

## ACKNOWLEDGMENT

The authors thank Erik Ordentlich for providing detailed feedback on the manuscript and his valuable technical suggestions during the review process.

## REFERENCES

- [1] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.
- [2] S. Shamai (Shitz) and A. Steiner, "A broadcast approach for a single-user slowly fading MIMO channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2617–2635, Oct. 2003.
- [3] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 727–734, Nov. 1985.
- [4] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder—II: General sources," *Inform. Contr.*, vol. 38, pp. 60–80, Jul. 1978.
- [5] R. M. Gray, "A new class of lower bounds to information rates of stationary sources via conditional rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 480–489, Jul. 1973.
- [6] R. Zamir, "The rate loss in the Wyner–Ziv problem," *IEEE Trans. Inf. Theory*, vol. 42, no. 6, pp. 2073–2084, Nov. 1996.
- [7] M. Fleming and M. Effros, "On rate-distortion with mixed types of side information," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1698–1705, Apr. 2006.
- [8] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner–Ziv problem," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1636–1654, Aug. 2004.
- [9] C. Tian and S. N. Diggavi, "On multistage successive refinement for Wyner–Ziv source coding with degraded side informations," *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2946–2960, Aug. 2007.
- [10] —, "Side-information scalable source coding," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5591–5608, Dec. 2008.
- [11] M. Feder and N. Shulman, "Source broadcasting with unknown amount of receiver side information," in *Proc. IEEE Inform. Theory Workshop*, Bangalore, India, Oct. 2002, pp. 127–130.
- [12] N. Shulman and M. Feder, "Static broadcasting," in *Proc. IEEE Int. Symp. on Inform. Theory*, Sorrento, Italy, Jun. 2000, p. 23.
- [13] C. Tian, A. Steiner, S. Shamai (Shitz), and S. N. Diggavi, "Successive refinement via broadcast: Optimizing expected distortion of a Gaussian source over a Gaussian fading channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 2903–2918, Jul. 2008.
- [14] C. T. K. Ng, D. Gündüz, A. J. Goldsmith, and E. Erkip, "Distortion minimization in gaussian layered broadcast coding with successive refinement," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5074–5086, Nov. 2009.
- [15] A. Steiner and S. Shamai (Shitz), "Achievable rates with imperfect transmitter side information using a broadcast transmission strategy," *IEEE Trans. Wireless Commun.*, vol. 7, no. 3, pp. 1043–1051, Mar. 2008.
- [16] 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.
- [17] D. Gunduz, J. Nayak, and E. Tuncel, "Wyner–Ziv coding over broadcast channels using hybrid digital/analog transmission," in *Proc. IEEE Int. Symp. on Inform. Theory*, Toronto, Canada, Jul. 2008, pp. 1543–1547.
- [18] J. Nayak, E. Tuncel, and D. Gündüz, "Wyner–Ziv coding over broadcast channels: Digital schemes," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1782–1799, Apr. 2010.
- [19] P. Ishwar, R. Puri, K. Ramchandran, and S. S. Pradhan, "On rate-constrained distributed estimation in unreliable sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 765–775, Apr. 2005.
- [20] J. Chen and T. Berger, "Robust distributed source coding," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3385–3398, Aug. 2008.
- [21] O. Simeone, O. Somekh, E. Erkip, H. V. Poor, and S. Shamai (Shitz), "Robust communication via decentralized processing with unreliable backhaul links," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4187–4201, Jul. 2011.
- [22] A. H. Kaspi, "Rate-distortion function when side-information may be present at the decoder," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 2031–2034, Nov. 1994.
- [23] J. Renegar, *A Mathematical View of Interior-Point Methods in Convex Optimization*. Philadelphia, PA: MPS-SIAM, 2001.
- [24] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.
- [25] G. L. Stüber, *Principles of Mobile Communication*. Norwell, MA: Kluwer Academic Publishers, 2000.
- [26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY: Wiley-Interscience, 1991.
- [27] J. N. Laneman, E. Martinian, G. W. Wornell, and J. G. Apostolopoulos, "Source-channel diversity for parallel channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3518–3539, Oct. 2005.
- [28] M. Katz and S. Shamai (Shitz), "Cooperative schemes for a source and an occasional nearby relay in wireless networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5138–5160, Nov. 2009.
- [29] D. G. Luenberger, *Optimization by Vector Space Methods*. New York, NY: Wiley-Interscience, 1969.

**Chris T. K. Ng** (S'99–M'07) received the B.A.Sc. degree in engineering science from the University of Toronto, Toronto, ON, Canada. He received the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA. Dr. Ng was a Member of Technical Staff at Bell Labs, Alcatel-Lucent, in Holmdel, NJ. From 2007 to 2008, he was a Postdoctoral Researcher in the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, Cambridge, MA. His research interests include cooperative communications, joint source-channel coding, cross-layer wireless network design, optimization, and network information theory. Dr. Ng was a recipient of the 2007 IEEE International Symposium on Information Theory Best Student Paper Award, and a recipient of a Croucher Foundation Fellowship in 2007.

**Chao Tian** (S'00–M'05) received the B.E. degree in Electronic Engineering from Tsinghua University, Beijing, China, in 2000 and the M.S. and Ph.D. degrees in Electrical and Computer Engineering from Cornell University, Ithaca, NY in 2003 and 2005, respectively.

Dr. Tian was a postdoctoral researcher at Ecole Polytechnique Federale de Lausanne (EPFL) from 2005 to 2007. He joined AT&T Labs–Research, Florham Park, New Jersey in 2007, where he is now a Senior Member of Technical Staff. His research interests include multi-user information theory, joint source-channel coding, quantization design and analysis, as well as image/video coding and processing. Dr. Tian is currently an associated editor for IEEE Signal Processing Letters.

**Andrea Goldsmith** is a professor of Electrical Engineering at Stanford University, and was previously an assistant professor of Electrical Engineering at Caltech. She co-founded Accelera Mobile Broadband, Inc. and Quantenna Communications Inc., and has previously held industry positions at Maxim Technologies, Memorylink Corporation, and AT&T Bell Laboratories. Dr. Goldsmith is a Fellow of the IEEE and of Stanford, and she has received several awards for her work, including the IEEE Communications Society and Information Theory Society joint paper award, the National Academy of Engineering Gilbreth Lecture Award, the IEEE Wireless Communications Technical Committee Recognition Award, the Alfred P. Sloan Fellowship, and the Silicon Valley/San Jose Business Journal's Women of Influence Award. Her research includes work on wireless information and communication theory, multihop wireless networks, cognitive radios, sensor networks, distributed control systems, "green" wireless system design, and applications of communications and signal processing to biology and neuroscience. She is author of the book "Wireless Communications" and co-author of the books "MIMO Wireless Communications" and "Principles of Cognitive Radio," all published by Cambridge University Press. She received the B.S., M.S. and Ph.D. degrees in Electrical Engineering from U.C. Berkeley.

Dr. Goldsmith has served as associate editor for the IEEE Transactions on Information Theory and as editor for the Journal on Foundations and Trends in Communications and Information Theory and in Networks. She previously served as an editor for the IEEE Transactions on Communications and for the IEEE Wireless Communications Magazine, as well as guest editor for several IEEE journal and magazine special issues. Dr. Goldsmith participates actively in committees and conference organization for the IEEE Information Theory and Communications Societies and has served on the Board of Governors for both societies. She is a Distinguished Lecturer for both societies, served as the President of the IEEE Information Theory Society in 2009, founded and chaired the student committee of the IEEE Information Theory society, and currently chairs the Emerging Technology Committee and is a member of the Strategic Planning Committee in the IEEE Communications Society. At Stanford she received the inaugural University Postdoc Mentoring Award, served as Chair of its Faculty Senate, and currently serves on its Faculty Senate and on its Budget Group.

**Shlomo Shamai (Shitz)** received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, in 1975, 1981 and 1986 respectively.

During 1975–1985 he was with the Communications Research Labs, in the capacity of a Senior Research Engineer. Since 1986 he is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, where he is now a Technion Distinguished Professor, and holds the William Fondiller Chair of Telecommunications. His research interests encompasses a wide spectrum of topics in information theory and statistical communications.

Dr. Shamai (Shitz) is an IEEE Fellow, and the recipient of the 2011 Claude E. Shannon Award. He is the recipient of the 1999 van der Pol Gold Medal of the Union Radio Scientifique Internationale (URSI), and a co-recipient of the 2000 IEEE Donald G. Fink Prize Paper Award, the 2003, and the 2004 joint IT/COM societies paper award, the 2007 IEEE Information Theory Society Paper Award, the 2009 European Commission FP7, Network of Excellence in Wireless COMMunications (NEWCOM++) Best Paper Award, and the 2010 Thomson Reuters Award for International Excellence in Scientific Research. He is also the recipient of 1985 Alon Grant for distinguished young scientists and the 2000 Technion Henry Taub Prize for Excellence in Research. He has served as Associate Editor for the Shannon Theory of the IEEE Transactions on Information Theory, and has also served twice on the Board of Governors of the Information Theory Society. He is a member of the Executive Editorial Board of the IEEE Transactions on Information Theory.