

Distortion Minimization in Gaussian Layered Broadcast Coding With Successive Refinement

Chris T. K. Ng, *Member, IEEE*, Deniz Gündüz, *Member, IEEE*, Andrea J. Goldsmith, *Fellow, IEEE*, and Elza Erkip, *Senior Member, IEEE*

Abstract—A transmitter without channel state information wishes to send a delay-limited Gaussian source over a slowly fading channel. The source is coded in superimposed layers, with each layer successively refining the description in the previous one. The receiver decodes the layers that are supported by the channel realization and reconstructs the source up to a distortion. The expected distortion is minimized by optimally allocating the transmit power among the source layers. For two source layers, the allocation is optimal when power is first assigned to the higher layer up to a power ceiling that depends only on the channel fading distribution; all remaining power, if any, is allocated to the lower layer. For convex distortion cost functions with convex constraints, the minimization is formulated as a convex optimization problem. In the limit of a continuum of infinite layers, the minimum expected distortion is given by the solution to a set of linear differential equations in terms of the density of the fading distribution. As the number of channel uses per source symbol tends to zero, the power distribution that minimizes expected distortion converges to the one that maximizes expected capacity.

Index Terms—Broadcast, convex optimization, distortion, fading channels, power allocation, source–channel coding, successive refinement, superposition.

I. INTRODUCTION

IN an ergodic wireless channel, from the source–channel separation theorem [1], it is optimal to first compress the source and incur the associated distortion at a rate equal to the channel capacity, then send the compressed representation over the channel at capacity with asymptotically small error. However, when delay constraints stipulate that the receiver decodes within a single realization of a slowly fading channel, without channel state information (CSI) at the transmitter, the

transmission over a single fading block is non-ergodic and source–channel separation is not necessarily optimal. In this case it is possible to reduce the end-to-end distortion of the reconstructed source by jointly optimizing the source-coding rate and the transmit power allocation based on the characteristics of the source and the channel. In particular, we consider using the layered broadcast coding approach with successive refinement in the transmission of a Gaussian source over a slowly fading channel, in the absence of CSI at the transmitter. First we assume the channel has a finite number of discrete fading states, then we extend the results to continuous fading distributions, for example, Rayleigh fading with diversity from the realization of independent fading paths. The source is coded in layers, with each layer successively refining the description in the previous one. The transmitter simultaneously transmits the codewords of all layers to the receiver by superimposing them with an appropriate power allocation. The receiver successfully decodes the layers supported by the channel realization, and combines the descriptions in the decoded layers to reconstruct the source up to a distortion. In this paper, we are interested in minimizing the expected distortion, and more generally, a convex distortion cost function, of the reconstructed source by optimally allocating the transmit power among the layers of codewords. The system model is applicable to communication systems with real-time traffic where it is difficult for the transmitter to learn the channel condition. For example, in a satellite voice system, it is desirable to consider the efficient transmission of the voice streams over uncertain channels that minimize the end-to-end distortion.

The broadcast strategy is proposed in [2] to characterize the set of achievable rates when the channel state is unknown at the transmitter. In the case of a Gaussian channel under Rayleigh fading, [3] describes the layered broadcast coding approach, and derives the optimal power allocation that maximizes expected capacity when the channel has a single-antenna transmitter and receiver. The layered broadcast approach is extended to multiple-antenna channels and the corresponding achievable rates are presented in [4]. In [5], coding theorems are presented for the broadcast approach with delayed error-free feedback under decoding delay constraints.

In the transmission of a Gaussian source over a Gaussian channel, uncoded transmission is optimal [6] in the special case when the source bandwidth equals the channel bandwidth [7]. For other bandwidth ratios, hybrid digital–analog joint source–channel transmission schemes are studied in [8]–[10]; in these works, the codes are designed to be optimal at a target signal-to-noise ratio (SNR) but degrade gracefully should the

Manuscript received May 22, 2007; revised November 15, 2008. Current version published October 21, 2009. 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, a grant from Intel, and the NSF under Grants 0430885 and 0635177. The work of C. T. K. Ng was supported by a Croucher Foundation Fellowship. The material in this paper was presented in part at the IEEE International Conference on Communications, Glasgow, Scotland, U.K., June 2007, and at the IEEE International Symposium on Information Theory (ISIT), Nice, France, June 2007.

C. T. K. Ng is with Bell Laboratories, Alcatel-Lucent, Holmdel, NJ 07733 USA (e-mail: Chris.Ng@alcatel-lucent.com).

D. Gündüz is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA, and the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: dgunduz@princeton.edu).

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

E. Erkip is with the Department of Electrical and Computer Engineering, Polytechnic Institute of New York University, Brooklyn, NY 11201 USA (e-mail: elza@poly.edu).

Communicated by A. Grant, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2009.2030455

realized SNR deviate from the target. In particular, [9] conjectures that no code is simultaneously optimal at different SNRs when the source and channel bandwidths are not equal. In this paper, the code considered is not targeted for a specific fading state; we minimize a convex distortion cost function over the fading distribution of the channel.

In [11], the minimum distortion is investigated in the transmission of a source over two independently fading channels in terms of the distortion exponent, which is defined as the exponential decay rate of the expected distortion in the high-SNR regime. Upper bounds on the distortion exponent and achievable joint source–channel schemes are presented in [12] for a single-antenna quasi-static Rayleigh-fading channel, and later in [13], [14] for multiple-antenna channels. One of the proposed schemes in [13], layered source coding with progressive transmission (LS), is analyzed in terms of expected distortion for a finite number of layers at finite SNR in [15]. The results in [12], [13] show that the broadcast strategy with layered source coding under an appropriate power allocation scheme is optimal for multiple-input single-output (MISO) and single-input multiple-output (SIMO) systems in terms of the distortion exponent. Numerical optimization of the power allocation with constant rate among the layers is examined in [16], while [17] considers the optimization of power and rate allocation and presents approximate solutions in the high-SNR regime. Motivated by the optimality of the broadcast strategy in the high-SNR regime, in this work, first presented in [18], [19], we investigate minimizing a convex distortion cost function, and in particular, the expected distortion, at any arbitrary finite SNR.

In a recent work in [20], the minimization of a linear distortion cost function, i.e., the expected distortion, is considered, and optimal power allocation algorithms are presented for discrete and continuous channel fading. In this paper, we study how the properties of the optimal power allocation are affected by the channel–source bandwidth ratio, operating SNR, channel quantization, diversity order, and the related metric of capacity maximization. Moreover, when the channel has discrete fading states, we also consider the minimization of an arbitrary convex distortion cost function under convex constraints, by formulating the distortion minimization as a convex optimization problem. In minimizing the expected distortion, [20] presents algorithms that calculate the optimal rate vector and power allocation based on a linear distortion cost function; general convex cost functions and constraints on the distortion realizations are not considered. We show that the feasible distortion region is convex in layered broadcast coding with successive refinement, and the minimization of convex distortion cost functions can be efficiently solved with convex optimization numerical techniques. The minimization of a general convex distortion cost function under continuous channel fading distributions, however, remains an open research problem.

The remainder of the paper is organized as follows. The system model is presented in Section II, and the layered broadcast coding scheme with successive refinement is explained in more detail in Section III. Section IV focuses on the optimal power allocation between two layers, with the analysis being extended in Section V to consider minimizing a convex distortion

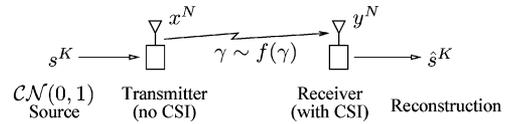


Fig. 1. Source–channel coding without CSI at the transmitter.

cost function over power allocation among multiple discrete layers. The optimal power allocation for discretized Rayleigh fading distributions are presented in Section VI. Minimizing expected distortion under continuous fading distributions are treated in Section VII by studying the limiting process as the channel discretization resolution increases. Section VIII considers the optimal power distribution and minimum expected distortion in Rayleigh-fading channels with diversity, followed by conclusions in Section IX.

II. SYSTEM MODEL

Consider the system model illustrated in Fig. 1: A transmitter wishes to send a Gaussian source over a wireless channel to a receiver, at which the source is to be reconstructed up to a distortion. Let the source be denoted by s , which is a sequence of independent and identically distributed (i.i.d.) zero-mean circularly symmetric complex Gaussian (ZMCSCG) random variables with unit variance: $s \in \mathbb{C} \sim \mathcal{CN}(0, 1)$. The transmitter and the receiver each have a single antenna and the channel is described by

$$y = Hx + n \quad (1)$$

where $x \in \mathbb{C}$ is the transmit signal, $y \in \mathbb{C}$ is the received signal, and $n \in \mathbb{C} \sim \mathcal{CN}(0, 1)$ is i.i.d. unit-variance ZMCSCG noise.

Suppose the distribution of the channel power gain is described by the probability density function (pdf) $f(\gamma)$, where $\gamma \triangleq |h|^2$ and $h \in \mathbb{C}$ is a realization of H . We first consider fading distributions with a finite number of discrete fading states; subsequently, we generalize to continuous fading distributions. The receiver has perfect CSI but the transmitter has only channel distribution information (CDI), i.e., the transmitter knows the pdf $f(\gamma)$ but not its instantaneous realization. The channel is modeled by a quasi-static block fading process: H is realized i.i.d. at the onset of each fading block and remains unchanged over the block duration. We assume decoding at the receiver is *delay limited*; namely, delay constraints preclude coding across fading blocks but dictate that the receiver decodes at the end of each block. Hence, the transmission over a single fading block is nonergodic.

Suppose each fading block spans N channel uses, over which the transmitter describes K of the source symbols. We define the *bandwidth ratio* as $b \triangleq N/K$, which relates the number of channel uses per source symbol. At the transmitter there is a power constraint on the transmit signal $\mathbb{E}[|x|^2] \leq P$, where the expectation is taken over repeated channel uses over the duration of each fading block. We assume K is large enough to consider the source as ergodic, and N is large enough to design codes that achieve the instantaneous channel capacity of a given fading state with negligible probability of error. At the receiver,

the channel output y is used to reconstruct an estimate \hat{s} of the source. The distortion D is measured by the mean-squared error $E[\|s - \hat{s}\|^2]$ of the estimator, where the expectation is taken over the K -sequence of source symbols and the noise distribution. The instantaneous distortion of the reconstruction depends on the fading realization of the channel; we are interested in minimizing the expected distortion $E[D]$, where the expectation is over the fading distribution, and more generally, a convex distortion cost function with convex constraints in terms of the possible distortion realizations.

III. LAYERED BROADCAST CODING WITH SUCCESSIVE REFINEMENT

To characterize the set of achievable rates when the channel state is unknown at the transmitter, a broadcast strategy is described in [2]. The transmitter designs its codebook by imagining it is communicating with an ensemble of virtual receivers. Each virtual receiver corresponds to a fading state: the realization of the fading state is taken as the channel gain of the virtual receiver. The realized rate at the original receiver is given by the decodable rate of the realized virtual receiver. A fading channel without transmitter CSI, therefore, can be modeled as a broadcast channel (BC). In particular, the capacity region of the BC defines the maximal set of achievable rates among the virtual receivers, which, in terms of the original fading channel, is the maximal set of realized rates among the fading states. In this work, we derive the optimal operating point in the BC capacity region that minimizes a convex distortion cost function, and in particular, the expected distortion $E[D]$, of the reconstructed source.

For fading Gaussian channels, a layered broadcast coding approach is described in [3], [4]. In the layered broadcast approach, the virtual receivers are ordered according to their channel strengths: for single-antenna channels, the channel strength of a virtual receiver is given by the channel power gain of its corresponding fading state. We interpret each codeword intended for a virtual receiver as a *layer* of code, and the transmitter sends the superposition of all layers to the virtual receivers. The capacity region of a single-antenna Gaussian BC is achievable by successive decoding [21], in which each virtual receiver decodes, in addition to its own layer, all the layers below it (the ones with weaker channel strengths). Hence, each layer represents the *additional* information over its lower layer that becomes decodable by the original receiver should the layer be realized.

The layered broadcast approach fits particularly well with the successive refinability [22], [23] of a Gaussian source. Successive refinability states that if a source is first described at rate R_1 , then subsequently refined at rate R_2 , the overall distortion is the same as if the source were described at rate $R_1 + R_2$ in the first place. As the Gaussian source is successively refinable, naturally, each layer in the broadcast approach can be used to carry refinement information of a lower layer. Concatenation of broadcast channel coding with successive refinement source coding is shown in [12], [13] to be optimal in terms of the distortion exponent for MISO/SIMO systems.

We apply the layered broadcast approach and successive refinability to perform source-channel coding as outlined

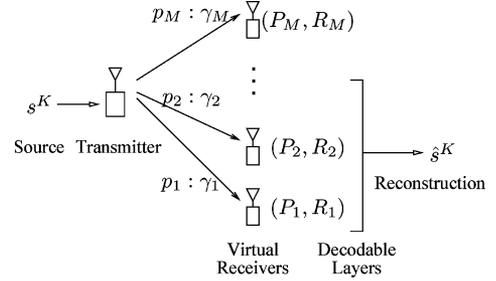


Fig. 2. Layered broadcast coding with successive refinement.

in Fig. 2. First we assume the fading distribution has M nonzero discrete states: the channel power gain realization is $\gamma_i > 0$ with probability p_i , for $i = 1, \dots, M$; and we denote $p_0 \triangleq \Pr\{\gamma = 0\}$, where we have $0 \leq p_i \leq 1$ and $\sum_{i=0}^M p_i = 1$. Accordingly there are M virtual receivers and the transmitter sends the sum of M layers of codewords. Let layer i denote the layer of codeword intended for virtual receiver i , and we order the layers as $\gamma_M > \dots > \gamma_1 > 0$. We refer to layer M as the highest layer and layer 1 as the lowest layer. Each layer successively refines the description of the source s from the layer below it, and the codewords in different layers are independent. Let P_i be the transmit power allocated to layer i , then the transmit symbol x can be written as

$$x = \sqrt{P_1}x_1 + \sqrt{P_2}x_2 + \dots + \sqrt{P_M}x_M \quad (2)$$

where x_1, \dots, x_M are i.i.d. ZMCSCG random variables with unit variance.

With successive decoding, each virtual receiver first decodes and cancels the lower layers before decoding its own layer; the undecodable higher layers are treated as noise. Thus, the rate R_i (bits per channel use) intended for virtual receiver i is

$$R_i = \log \left(1 + \frac{\gamma_i P_i}{1 + \gamma_i \sum_{j=i+1}^M P_j} \right), \quad i = 1, \dots, M \quad (3)$$

where \log is to base 2, and the term $\gamma_i \sum_{j=i+1}^M P_j$ represents the interference power from the higher layers. Suppose γ_k is the realized channel power gain, then the original receiver can decode layer k and all the layers below it. Hence, the realized rate $R_{\text{rlz}}^{(k)}$ at the original receiver is $R_1 + \dots + R_k$.

From the rate distortion function of a unit-variance complex Gaussian source [21], the mean squared distortion is 2^{-bR} when the source is described at a rate of bR bits per symbol. Thus, the realized distortion $D_{\text{rlz}}^{(k)}$ of the reconstructed source \hat{s} is

$$D_{\text{rlz}}^{(k)} = 2^{-bR_{\text{rlz}}^{(k)}} = 2^{-b(R_1 + \dots + R_k)} \quad (4)$$

where the last equality follows from successive refinability. The expected distortion $E[D]$ is obtained by averaging over the probability mass function (pmf) of the fading distribution

$$E[D] = \sum_{i=0}^M p_i D_{\text{rlz}}^{(i)} = \sum_{i=0}^M p_i 2^{-b \sum_{j=1}^i R_j} \quad (5)$$

where $D_{\text{rlz}}^{(0)} \triangleq 1$.

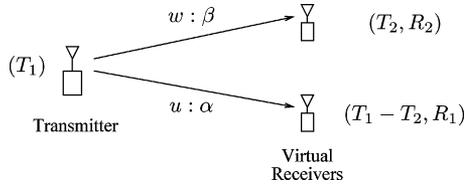


Fig. 3. Power allocation between two layers.

We begin by deriving the optimal power allocation P_1^*, \dots, P_M^* among the layers to find the minimum expected distortion $\mathbb{E}[D]^*$. Subsequently, we generalize to consider minimizing a convex distortion cost function. If a layer has an expected power gain of zero (i.e., $p_i \gamma_i = 0$), the layer is allocated zero power; hence, in the derivation we assume $p_i \gamma_i \neq 0$, for $i = 1, \dots, M$. Note that the expected distortion is monotonically decreasing in the transmit power P , hence the power constraint can be taken as an equality $\sum_{i=1}^M P_i = P$, and the optimization is formulated as

$$\begin{aligned} \mathbb{E}[D]^* &= \min_{P_1, \dots, P_M} \mathbb{E}[D] \\ \text{subject to } P_i &\geq 0, \sum_{i=1}^M P_i = P, \forall i = 1, \dots, M. \end{aligned} \quad (6)$$

We first consider the power allocation between two layers in the next section, then the analysis is extended to consider more than two layers in Section V. The layered source coding broadcast scheme can be straightforwardly extended to MISO/SIMO systems. The equivalent single-antenna channel distribution is found by using isotropic inputs at the transmitter for MISO systems, and performing maximal-ratio combining at the receiver for SIMO systems.

IV. TWO-LAYER OPTIMAL POWER ALLOCATION

Suppose the channel fading distribution has only two states: the channel power gain realization is either α or β , with $\beta > \alpha > 0$. The transmitter then sends two layers ($M = 2$) of codewords as shown in Fig. 3. Let T_1 denote the total transmit power constraint, and T_2 denote the power allocated to layer 2; the remaining power $T_1 - T_2$ is allocated to layer 1. The decodable rates for the virtual receivers are denoted by R_1, R_2 ; with successive decoding, they are given as follows:

$$R_2 = \log(1 + \beta T_2) \quad (7)$$

$$R_1 = \log\left(1 + \frac{\alpha(T_1 - T_2)}{1 + \alpha T_2}\right). \quad (8)$$

Suppose we generalize slightly and consider the weighted distortion

$$D_1 = u 2^{-b R_1} + w 2^{-b(R_1 + R_2)} \quad (9)$$

where the weights $\{u, w\}$ are nonnegative. Note that the weighted distortion D_1 is the expected distortion $\mathbb{E}[D]$ when the weights $\{u, w\}$ are the probabilities of the fading realizations.

Given T_1 , the total power available to the two layers, we optimize over T_2 to minimize the weighted distortion

$$D_1^* = \min_{T_2 \in [0, T_1]} D_1 \quad (10)$$

$$= \min_{T_2 \in [0, T_1]} \left(\frac{1 + \alpha T_1}{1 + \alpha T_2}\right)^{-b} \left[u + (1 + \beta T_2)^{-b} w\right]. \quad (11)$$

The minimization can be solved by the Lagrange method. We form the Lagrangian

$$L(T_2, \lambda_1, \lambda_2) = D_1 + \lambda_1(T_2 - T_1) - \lambda_2 T_2. \quad (12)$$

Applying the Karush–Kuhn–Tucker (KKT) necessary conditions, the gradient of the Lagrangian vanishes at the optimal power allocation T_2^* . Specifically, the KKT conditions stipulate that at T_2^* , either one of the inequality constraints is active, or $dD_1/dT_2 = 0$. Only one solution satisfies the KKT conditions, which leads to the optimal power allocation

$$T_2^* = \min(U_2, T_1) = \begin{cases} U_2, & \text{if } U_2 \leq T_1 \\ T_1, & \text{else} \end{cases} \quad (13a)$$

where

$$U_2 \triangleq \begin{cases} 0, & \text{if } \beta/\alpha \leq 1 + u/w \\ \frac{1}{\beta} \left(\left[\frac{w}{u} \left(\frac{\beta}{\alpha} - 1 \right) \right]^{\frac{1}{1+b}} - 1 \right), & \text{else.} \end{cases} \quad (14)$$

In Section V-D, we show that the distortion minimization can be posed as a convex optimization problem; hence, the KKT conditions are necessary and sufficient for optimality (we assume the given total power is nonzero so Slater's condition holds). Interestingly, U_2 depends only on the layer parameters w, β, u, α (which are derived from the channel fading distribution) and the bandwidth ratio b , but not on the total power T_1 . In other words, the higher layer is allocated a *fixed* amount of power as long as there is sufficient power available. The optimal power allocation, therefore, adopts a simple policy: first assign power to the higher layer up to a ceiling of U_2 , then assign all remaining power, if any, to the lower layer.

Under optimal power allocation T_2^* , the minimum weighted distortion as a function of the total power T_1 is given by

$$D_1^* = \begin{cases} (1 + \alpha T_1)^{-b} W_1, & \text{if } U_2 \leq T_1 \\ u + (1 + \beta T_1)^{-b} w, & \text{else} \end{cases} \quad (15a)$$

where

$$W_1 \triangleq (1 + \alpha U_2)^b [u + (1 + \beta U_2)^{-b} w]. \quad (16)$$

Note that when the total power constraint $T_2 \leq T_1$ is not active (15a), the consequent minimum weighted distortion is analogous to that of a *single* layer with channel power gain α and an equivalent weight W_1 . On the other hand, when the total power constraint is active (15b), it is equivalent to one with channel power gain β and an equivalent weight w (with an additive constant u in the distortion). Hence, under optimal power allocation, with respect to the minimum expected distortion, the two layers

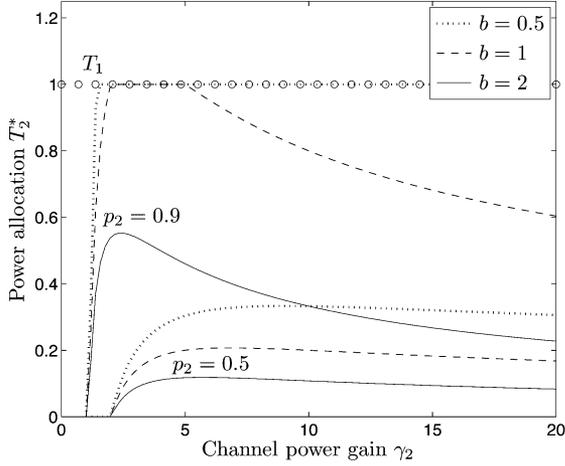


Fig. 4. Optimal power allocation between two layers ($T_1 = 0$ dB)

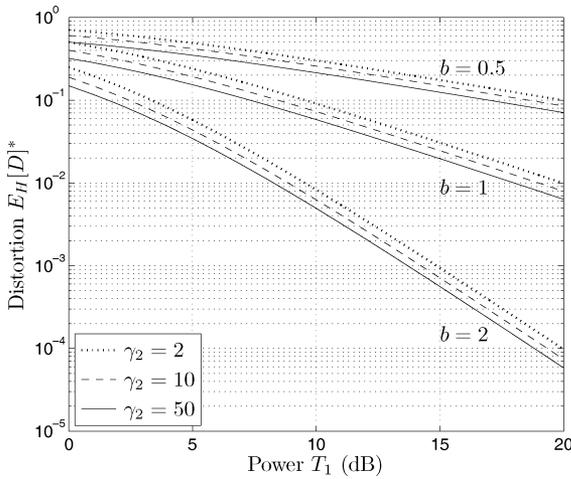


Fig. 5. Two-layer minimum expected distortion ($p_2 = 0.5$).

can be represented by a single *aggregate* layer; this idea is explored further when we consider multiple layers in Section V.

The optimal power allocation and the minimum expected distortion for a channel that has two discrete fading states is shown in Figs. 4 and 5, respectively, where T_2^* , the power assigned to the top layer, is computed under the following parameters:

$$\begin{aligned} w &= p_2, & \beta &= \gamma_2 \\ u &= 1 - p_2, & \alpha &= 1. \end{aligned} \quad (17)$$

Fig. 4 shows that when the channel power gain γ_2 is small, an increase in γ_2 leads to a larger power allocation T_2^* at the top layer, up to the total available power T_1 . However, as γ_2 further increases, T_2^* begins to fall. This is because the transmission power in the higher layer is in effect interference to the lower layer. When the top layer has a strong channel, the overall expected distortion is dominated by the bottom layer; in which case it is more beneficial to distribute the power to minimize the interference. Fig. 5 plots the two-layer minimum expected distortion on a logarithmic scale, and it shows that the relative

power gain of the channels has only a marginal impact on the expected distortion, as the overall distortion is in general dominated by the weaker channel.

V. MULTIPLE-LAYER POWER ALLOCATION

In this section, we consider the case when the fading distribution has M fading states as depicted in Fig. 2, where M is finite and $M \geq 2$. For notational convenience, we write the power assignment as a cumulative sum starting from the top layer

$$T_j \triangleq \sum_{i=j}^M P_i, \quad \text{for } j = 1, \dots, M. \quad (18)$$

The original power assignments $\{P_1, \dots, P_M\}$ can then be recovered from $\{T_1, \dots, T_M\}$ by taking their differences. By definition, $T_1 = P$ is given; hence the optimization is over the variables T_2, \dots, T_M

$$\begin{aligned} E[D]^* &= \min_{T_2, \dots, T_M} E[D] \\ &\text{subject to } 0 \leq T_M \leq \dots \leq T_2 \leq P. \end{aligned} \quad (19)$$

A. Expected Distortion Recurrence Relations

In terms of the cumulative power variables T_1, \dots, T_M , the expected distortion in (5) can be written as

$$E[D] = \sum_{i=1}^M p_i \left(\prod_{j=1}^i \frac{1 + \gamma_j T_j}{1 + \gamma_j T_{j+1}} \right)^{-b} \quad (20)$$

where $T_{M+1} \triangleq 0$. We factor the sum of cumulative products in (20) and rewrite the expected distortion as a set of recurrence relations

$$D_M \triangleq (1 + \gamma_M T_M)^{-b} p_M \quad (21)$$

$$D_i = \left(\frac{1 + \gamma_i T_i}{1 + \gamma_i T_{i+1}} \right)^{-b} (p_i + D_{i+1}) \quad (22)$$

where i runs from $M-1$ down to 1. We refer to D_i as the cumulative distortion, which represents the cumulative effects on the expected distortion from layers i and above, with $D_1 = E[D]$. Note that given D_{i+1} from the previous recurrence step, the term D_i depends on only two adjacent power allocation variables T_i and T_{i+1} ; therefore, in each recurrence step i we solve for the optimal T_{i+1}^* in terms of T_i

$$D_M^* \triangleq D_M \quad (23)$$

$$D_i^* = \min_{T_{i+1} \in [0, T_i]} \left(\frac{1 + \gamma_i T_i}{1 + \gamma_i T_{i+1}} \right)^{-b} (p_i + D_{i+1}^*). \quad (24)$$

In the last recurrence step ($i = 1$), the minimum expected distortion $E[D]^*$ is then given by D_1^* .

B. Reduction Through Optimal Power Allocation

We consider the layers from top to bottom. In each recurrence step, the minimum distortion D_i^* in (24) can be found by optimally allocating power between two adjacent layers as described in Section IV. In the first recurrence step ($i = M-1$), we consider the power allocation between the topmost two

layers. The minimal distortion D_M^* is found by setting the parameters in (11) to be

$$\begin{aligned} w_{M-1} &= p_M, & \beta_{M-1} &= \gamma_M \\ u_{M-1} &= p_{M-1}, & \alpha_{M-1} &= \gamma_{M-1} \end{aligned} \quad (25)$$

where the subscripts on the layer parameters w, β, u, α designate the recurrence step. In general, in recurrence step i , the power allocation between layer i and layer $i+1$ can be found by the optimization

$$D_i^* = \min_{T_{i+1} \in [0, T_i]} \left(\frac{1 + \alpha_i T_i}{1 + \alpha_i T_{i+1}} \right)^{-b} \left[u_i + (1 + \beta_i T_{i+1})^{-b} w_i \right] \quad (26)$$

the solution of which is given in (15)

$$D_i^* = \begin{cases} (1 + \alpha_i T_i)^{-b} W_i, & \text{if } U_{i+1} \leq T_i \\ u_i + (1 + \beta_i T_i)^{-b} w_i, & \text{else.} \end{cases} \quad \begin{matrix} (27a) \\ (27b) \end{matrix}$$

There are two cases to the solution of D_i^* . In the first case, the power allocation is not constrained by the available power T_i , and we substitute (27a) in the recurrence relation (24) to find the minimum distortion in the next recurrence step $i-1$

$$D_{i-1}^* = \min_{T_i \in [0, T_{i-1}]} \left(\frac{1 + \gamma_{i-1} T_{i-1}}{1 + \gamma_{i-1} T_i} \right)^{-b} \cdot \left[p_{i-1} + (1 + \alpha_i T_i)^{-b} W_i \right]. \quad (28)$$

The minimization in (28) has the same form as the one in (26), but with the following parameters:

$$\begin{aligned} w_{i-1} &= W_i, & \beta_{i-1} &= \alpha_i \\ u_{i-1} &= p_{i-1}, & \alpha_{i-1} &= \gamma_{i-1}. \end{aligned} \quad (29)$$

Hence, the minimization can be solved the same way as in the last recurrence step. In the second case, the power allocation is constrained by the available power T_i , thus we instead substitute (27b) in (24) in the next recurrence step $i-1$

$$D_{i-1}^* = \min_{T_i \in [0, T_{i-1}]} \left(\frac{1 + \gamma_{i-1} T_{i-1}}{1 + \gamma_{i-1} T_i} \right)^{-b} \cdot \left[p_{i-1} + u_i + (1 + \beta_i T_i)^{-b} w_i \right] \quad (30)$$

which again has the same form as in (26), with the following parameters:

$$\begin{aligned} w_{i-1} &= w_i, & \beta_{i-1} &= \beta_i, \\ u_{i-1} &= p_{i-1} + u_i, & \alpha_{i-1} &= \gamma_{i-1}. \end{aligned} \quad (31)$$

Therefore, in each recurrence step, the two-layer optimization procedure described in Section IV can be used to find the minimum distortion and the optimal power allocation between the current layer and the aggregate higher layer.

C. Feasibility of Unconstrained Minimizer

When we proceed to the next recurrence step, however, it is necessary to determine which set of parameters in (29), (31) should be applied. Note that in the optimization in (26), if the available power T_i is unlimited (i.e., $T_i = \infty$), then the optimal power allocation is $T_{i+1}^* = U_{i+1}$ as given in (13a); hence, U_{i+1} is the *unconstrained* minimizer of D_i . Consequently, we can first assume the minimization in (26) is unconstrained by T_i and its solution is given by (27a). If the unconstrained allocation U_{i+1} is found to be feasible, then it is indeed the optimal allocation. On the other hand, if U_{i+1} is subsequently shown to be infeasible, then we backtrack to the minimization in (26) and adopt the *constrained* solution given by (27b). In this case, $T_{i+1}^* = T_i$ as given in (13b), which implies layer i is inactive since $P_i^* = T_i - T_{i+1}^* = 0$.

We ascertain the feasibility of U_{i+1} by verifying that it does not exceed the available power allocation T_i from the lower layer i , which in turn depends on the power allocation T_{i-1} from the next lower layer $i-1$, and so on. The procedure can be accomplished by the recursive algorithm shown in Algorithm 1 in the Appendix. We start by allocating power between the top-most two layers (line 1). In each recursion step, we compute the unconstrained allocation U (line 3). If U does not exceed the total power P , we first assume it is feasible, and proceed in the recursion to find the power allocation T_i^* from the lower layer (line 10). If U turns out to be infeasible, then we repeat the allocation step with the constrained minimization parameters (line 16). The recursion continues until the bottom layer is reached (line 4). In the best case, if the unconstrained allocations for all layers are feasible, the algorithm has complexity $O(M)$. In the worst case, if all unconstrained allocations are infeasible, each recursion step performs two power allocations and the algorithm has complexity $O(2^M)$.

D. Convex Distortion Cost Function

In the previous sections, we consider minimizing a linear objective function of the possible distortion realizations $D_{\text{rlz}}^{(k)}$'s; in particular, we consider the expected distortion $E[D] = \sum_{k=1}^M p_k D_{\text{rlz}}^{(k)}$. Analytical solutions are presented that characterize the optimal power allocation that minimizes the expected distortion. However, the expected distortion $E[D]$ does not capture a user's sensitivity regarding the uncertainty in the range of possible outcomes of the distortion realizations $D_{\text{rlz}}^{(k)}$'s. In this subsection, we present a numerical optimization framework in which a wider class of objective functions is permissible. Specifically, we consider the minimization of a distortion cost function $J(D_{\text{rlz}}^{(1)}, \dots, D_{\text{rlz}}^{(M)})$, where $J(\cdot)$ is convex in $D_{\text{rlz}}^{(1)}, \dots, D_{\text{rlz}}^{(M)}$. We show that the distortion minimization can be formulated as a convex problem; hence, its solution can be computed efficiently by standard numerical methods in convex optimization [24].

In terms of the cumulative power variables T_j 's defined in (18), the distortion realization $D_{\text{rlz}}^{(k)}$ given in (4) can be written as

$$D_{\text{rlz}}^{(k)} = \prod_{j=1}^k \left(\frac{1 + \gamma_j T_j}{1 + \gamma_j T_{j+1}} \right)^{-b}, \quad k = 1, \dots, M \quad (32)$$

where $T_{M+1} \triangleq 0$ as previously defined. Note that in (32), $D_{\text{rlz}}^{(k)}$ can be written in terms of $D_{\text{rlz}}^{(k-1)}$ as follows:

$$D_{\text{rlz}}^{(k)} = D_{\text{rlz}}^{(k-1)} \left(\frac{1 + \gamma_k T_k}{1 + \gamma_k T_{k+1}} \right)^{-b}, \quad k = 1, \dots, M \quad (33)$$

where again recall $D_{\text{rlz}}^{(0)} \triangleq 1$. Next, we rearrange (33) and write

$$T_j = \left(T_{j+1} + \frac{1}{\gamma_j} \right) \left(\frac{D_{\text{rlz}}^{(j)}}{D_{\text{rlz}}^{(j-1)}} \right)^{-1/b} - \frac{1}{\gamma_j}, \quad j = 1, \dots, M \quad (34)$$

which we expand recursively initializing from $j = 1$ to arrive at an expression for the total power $T_1 \triangleq \sum_{i=1}^M P_i$, which is given by

$$T_1 = -\frac{1}{\gamma_1} + \sum_{i=1}^M \left(\frac{1}{\gamma_i} - \frac{1}{\gamma_{i+1}} \right) (D_{\text{rlz}}^{(i)})^{-1/b} \quad (35)$$

where $\gamma_{M+1} \triangleq \infty$. Under the power constraint $\sum_{i=1}^M P_i \leq P$, the distortion cost function minimization problem can be formulated as

$$\begin{aligned} & \text{minimize} \quad J(D_{\text{rlz}}^{(1)}, \dots, D_{\text{rlz}}^{(M)}) \\ & \text{over} \quad D_{\text{rlz}}^{(1)}, \dots, D_{\text{rlz}}^{(M)} \\ & \text{subject to} \end{aligned} \quad (36)$$

$$-\frac{1}{\gamma_1} + \sum_{i=1}^M \left(\frac{1}{\gamma_i} - \frac{1}{\gamma_{i+1}} \right) (D_{\text{rlz}}^{(i)})^{-1/b} \leq P \quad (37)$$

$$0 \leq D_{\text{rlz}}^{(M)} \leq \dots \leq D_{\text{rlz}}^{(1)} \leq 1 \quad (38)$$

where (37) corresponds to the power constraint from (34), and (38) corresponds to the realized rates $R_{\text{rlz}}^{(k)}$'s being nonnegative in (4). Note that the constraints (37), (38) are convex: in (37), $(D_{\text{rlz}}^{(i)})^{-1/b}$ is convex, and $1/\gamma_i - 1/\gamma_{i+1} > 0$, which follows from the system model assumptions $\gamma_M > \dots > \gamma_1 > 0$. Therefore, the feasible distortion region $\{D_{\text{rlz}}^{(1)}, \dots, D_{\text{rlz}}^{(M)}\}$, as characterized by (37), (38), is convex. The objective function $J(D_{\text{rlz}}^{(1)}, \dots, D_{\text{rlz}}^{(M)})$ in (36) is convex by assumption. It follows that minimizing $J(D_{\text{rlz}}^{(1)}, \dots, D_{\text{rlz}}^{(M)})$ over a convex region is a convex optimization problem: it can be efficiently solved by standard convex optimization numerical methods. For instance, the optimization problem above can be solved using the CVX software package [25], [26].

For example, to characterize the user's sensitivity to the uncertainty in the realized distortion, we may consider a risk-sensitive distortion cost function

$$J_\varphi(D_{\text{rlz}}^{(1)}, \dots, D_{\text{rlz}}^{(M)}) \triangleq \text{E}[D] + \varphi \text{VAR}[D] \quad (39)$$

where $\text{VAR}[D]$ denotes the distortion variance

$$\text{VAR}[D] = \text{E}[(D - \text{E}[D])^2] \quad (40)$$

$$= \sum_{k=0}^M p_k \left(D_{\text{rlz}}^{(k)} - \sum_{i=0}^M p_i D_{\text{rlz}}^{(i)} \right)^2. \quad (41)$$

Note that $J_\varphi(\cdot)$ is a convex function of $D_{\text{rlz}}^{(1)}, \dots, D_{\text{rlz}}^{(M)}$. In (39), $\varphi \geq 0$ is a given scalar constant, which represents the risk-aver-

sion parameter [24]. Accordingly, we may specify a suitable value of φ to model the user's willingness to trade off an increase in the expected distortion $\text{E}[D]$ in return for a reduction in the distortion variance $\text{VAR}[D]$. Under the convex optimization formulation in (36)–(38), we may additionally consider convex constraints on $D_{\text{rlz}}^{(1)}, \dots, D_{\text{rlz}}^{(M)}$. For example, to guarantee worst case performance under unfavorable fading states, we may consider the following maximum distortion or variance constraints:

$$\text{E}[D] \leq D_{\text{max}} \quad (42)$$

$$\text{VAR}[D] \leq V_{\text{max}} \quad (43)$$

$$D_{\text{rlz}}^{(k)} \leq D_{\text{max}}^{(k)}, \quad k = 1, \dots, M \quad (44)$$

where D_{max} , V_{max} , and $D_{\text{max}}^{(k)}$'s are given constants. In Section VI, numerical examples are presented where we minimize the risk-sensitive distortion cost function $J_\varphi(\cdot)$ under discretized Rayleigh fading.

VI. DISCRETIZED RAYLEIGH FADING DISTRIBUTION

In this section, we present numerical results produced by the multiple-layer power allocation algorithms described in Section V. In the examples, we assume the channel pmf is taken from a discretized Rayleigh fading distribution. Specifically, for a channel under Rayleigh fading with unit power, the channel power gain γ is exponentially distributed with unit mean, and its pdf is given by

$$f(\gamma) = e^{-\gamma}, \quad \text{for } \gamma \geq 0. \quad (45)$$

We truncate the pdf at $\gamma = \Gamma$, quantize γ into M evenly spaced levels

$$\gamma_i \triangleq i\Gamma/M, \quad \text{for } i = 1, \dots, M \quad (46)$$

with $\gamma_0 \triangleq 0$, and discretize the probability distribution of γ to the closest lower level γ_i

$$p_i \triangleq \int_{i\Gamma/M}^{(i+1)\Gamma/M} f(\gamma) d\gamma, \quad \text{for } i = 0, \dots, M-1 \quad (47)$$

$$p_M \triangleq \int_{\Gamma}^{\infty} f(\gamma) d\gamma. \quad (48)$$

While it is possible to consider the optimal discretization of a fading distribution that minimizes expected distortion [16], [17], [27], in this paper we assume the channel pmf is given and do not consider such a step.

The optimal power allocation that minimizes the expected distortion $\text{E}[D]$ for the discretized Rayleigh-fading pmf is shown in Figs. 6 and 7. The Rayleigh-fading pdf is truncated at $\Gamma = 2$, and discretized into $M = 24$ levels. The truncation is justified by the observation that in the output the highest layers near Γ are not assigned any power. Fig. 6 plots the optimal power allocation P_i^* 's for different layers (indexed by the channel power gain γ_i) at SNRs $P = 0, 5$, and 10 dB, with the bandwidth ratio $b = 1$. We observe that the highest layers are inactive ($P_i^* = 0$), and within the range of active layers a lower layer is in general allocated more power than a higher layer, except at the lowest active layer which is assigned the remaining power. As SNR increases, the power allocations of the higher layers are unaltered, but the range of active layers extends

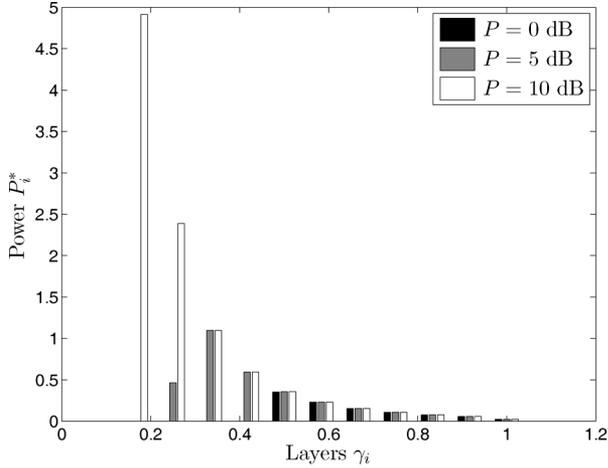


Fig. 6. Optimal power allocation that minimizes the expected distortion $E[D]$ ($b = 1$).

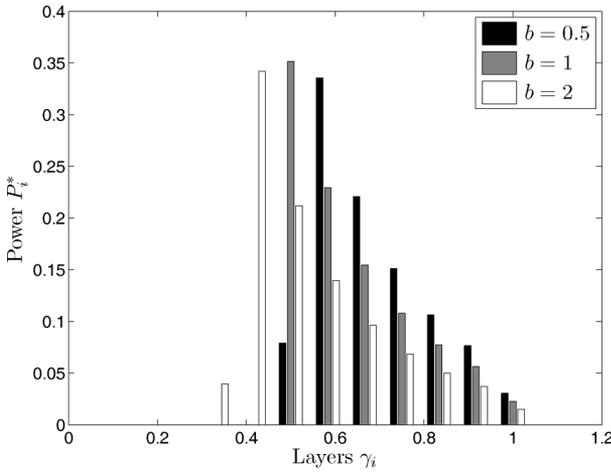


Fig. 7. Optimal power allocation that minimizes the expected distortion $E[D]$ ($P = 0$ dB).

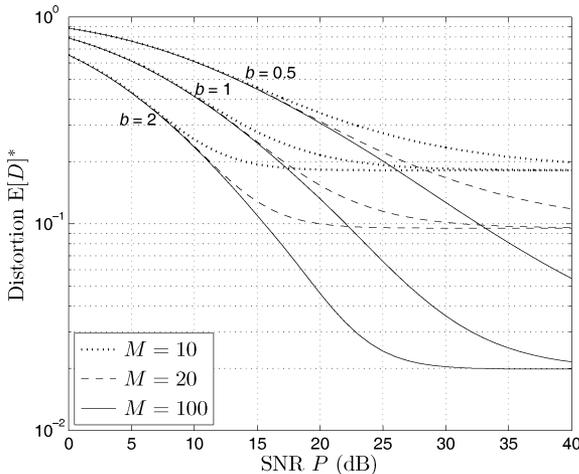


Fig. 8. Minimum expected distortion under optimal power allocation.

further into the lower layers. On the other hand, Fig. 7 plots the allocation P_i^* 's for different bandwidth ratios $b = 0.5, 1, 2$ at the SNR of 0 dB. It can be observed that a higher b (i.e., more

channel uses per source symbol) has the effect of spreading the power allocation further across into the lower layers.

Intuitively, the higher layers have stronger channels but suffer from larger risks of being in outage, while the lower layers provide higher reliability but at the expense of having to cope with less power-efficient channels. Accordingly, the optimal power allocation is concentrated around the middle layers. Furthermore, as SNR increases, the numerical results suggest that, to minimize the expected distortion in a Rayleigh-fading channel, it is more favorable to utilize the weaker channels with the extra power rather than accepting the larger outage risks from the higher layers.

The minimum expected distortion $E[D]^*$ under optimal power allocation is shown in Fig. 8 on a logarithmic scale. When the bandwidth ratio b is higher, $E[D]^*$ decreases as expected. However, the improvement in $E[D]^*$ from refining the resolution M of the discretization is almost negligible at low SNRs. At high SNRs, on the other hand, the distortion is dominated by the outage probability

$$P_{\text{out}} \triangleq \Pr\{\gamma_0 = 0 \text{ is realized}\} \quad (49)$$

$$= \int_0^{\Gamma/M} f(\gamma) d\gamma \quad (50)$$

which is decreasing in M . Therefore, when the SNR is sufficiently high, the expected distortion $E[D]^*$ reaches a floor that is dictated by P_{out} . This behavior is due to having evenly spaced γ_i 's; the performance could be improved by optimizing the quantization level γ_i 's for the given M layers.

As a comparison, we consider the expected distortion lower bounds when the system has CSI at the transmitter (CSIT). Under the discretized Rayleigh-fading pmf, suppose the realized channel power gain is known to be γ_k , then it is optimal for the transmitter to concentrate all power on layer k to achieve the instantaneous distortion $D_{q\text{-CSIT}} = (1 + \gamma_k P)^{-b}$. Thus, with the quantized CSIT, the expected distortion is given by

$$E[D_{q\text{-CSIT}}] = \sum_{k=0}^M p_k (1 + \gamma_k P)^{-b}. \quad (51)$$

In terms of the original Rayleigh-fading pdf $f(\gamma)$, with perfect CSIT, the expected distortion is similarly given by

$$E[D_{\text{CSIT}}] = \int_0^\infty e^{-\gamma} (1 + \gamma P)^{-b} d\gamma \quad (52)$$

where the definite integral can be evaluated numerically. The expected distortions are plotted in Fig. 9 for the cases of no CSIT, quantized, and perfect CSIT. It can be observed that at low SNRs, quantized CSIT is nearly as good as perfect CSIT, whereas at high SNRs, quantized CSIT provides only marginal improvement over no CSIT, as the expected distortion is dominated by the probability of outage.

Numerical examples that illustrate the minimization of a convex cost function of the possible distortion realizations $D_{\text{rlz}}^{(k)}$'s are shown in Figs. 10 and 11. We consider the risk-sensitive distortion cost function: $J_\varphi(D_{\text{rlz}}^{(1)}, \dots, D_{\text{rlz}}^{(M)}) \triangleq E[D] + \varphi \text{VAR}[D]$. Fig. 10 shows the optimal power allocation that minimizes $J_\varphi(\cdot)$ for different values of the risk-aversion parameter φ . It is observed that a large φ , which represents the user's aversion to large variations in the realized distortion,

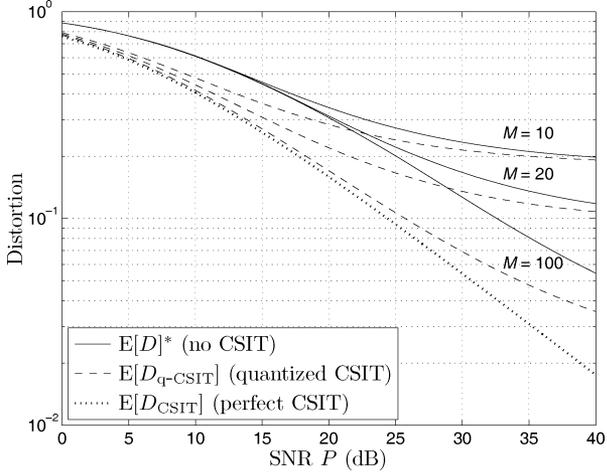


Fig. 9. Comparison of the achievable expected distortion with lower bounds assuming quantized and perfect CSIT ($b = 0.5$).

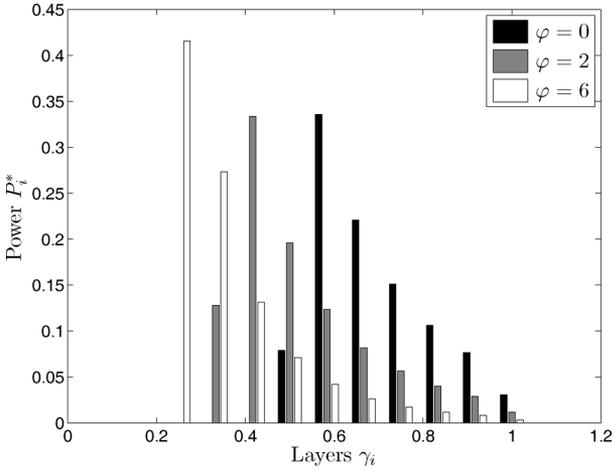


Fig. 10. Optimal power allocation that minimizes $E[D] + \varphi \text{VAR}[D]$ ($P = 0$ dB, $b = 0.5$).

shifts and concentrates the power allocation towards the lower layers. Fig. 11 plots the corresponding expected distortion $E[D]$ and distortion variance $\text{VAR}[D]$ that minimizes $J_\varphi(\cdot)$. As φ increases, it shows the tradeoff of accepting a higher $E[D]$ for the reduction in $\text{VAR}[D]$.

VII. CONTINUOUS FADING DISTRIBUTION

In this section, we consider continuous fading distributions, and we focus on the minimization of a linear distortion cost function, the expected distortion, by extending on the optimal power allocation analytical solutions derived in Section V-B. We study the limiting process as the discretization resolution of the fading distribution tends to infinity, and consider the optimal power distribution that minimizes the expected distortion when the fading distribution of the channel is given by a continuous

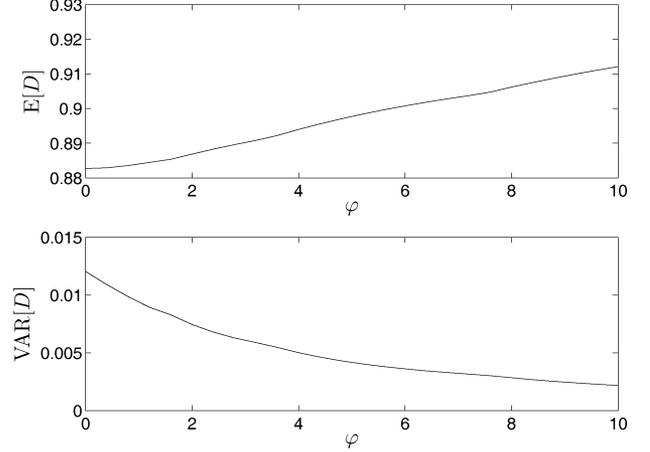


Fig. 11. Expected distortion and variance corresponding to $\min\{E[D] + \varphi \text{VAR}[D]\}$ ($P = 0$ dB, $b = 0.5$).

pdf. Specifically, we assume the layers are evenly spaced, with $\gamma_{i+1} - \gamma_i = \Delta\gamma$, and we consider the limiting process as $\Delta\gamma \rightarrow 0$ to obtain the power distribution

$$\rho(\gamma) \triangleq \lim_{\Delta\gamma \rightarrow 0} \frac{1}{\Delta\gamma} P_{\lceil \gamma / \Delta\gamma \rceil} \quad (53)$$

where for discrete layers the power allocation P_i is referenced by the integer layer index i , while the continuous power distribution $\rho(\gamma)$ is indexed by the channel power gain γ .

Consider the optimal power allocation between layer γ and its next lower layer $\gamma - \Delta\gamma$. Let $T(\gamma - \Delta\gamma)$ denote the available transmit power for layers $\gamma - \Delta\gamma$ and above, of which $T(\gamma)$ is allocated to layers γ and above; the remaining power $T(\gamma - \Delta\gamma) - T(\gamma)$ is allocated to layer $\gamma - \Delta\gamma$. The optimal power allocation $T^*(\gamma)$ is given by the solution to the two-layer optimization problem in Section IV, with the parameters in Fig. 3 correspondingly set to be

$$\begin{aligned} w &= W(\gamma), & \beta &= \gamma \\ u &= f(\gamma)\Delta\gamma, & \alpha &= \gamma - \Delta\gamma \end{aligned} \quad (54)$$

where $f(\gamma)$ is the pdf of the channel power gain with $f(\gamma)\Delta\gamma$ representing the probability that layer $\gamma - \Delta\gamma$ is realized, and $W(\gamma)$ is interpreted as an equivalent probability weight summarizing the aggregate effect of the layers γ and above.

From (13), (14), the optimal power allocation is given by

$$T^*(\gamma) = \begin{cases} U(\gamma), & \text{if } U(\gamma) \leq T(\gamma - \Delta\gamma) \\ T(\gamma - \Delta\gamma), & \text{else} \end{cases} \quad (55a)$$

where $U(\gamma)$ is defined in (56a) and (56b) at the bottom of the page. We assume there is a region of γ where the cumulative power allocation is not constrained by the power available from

$$U(\gamma) \triangleq \begin{cases} 0, & \text{if } \gamma \geq W(\gamma)/f(\gamma) + \Delta\gamma \\ \frac{1}{\gamma} \left(\left[\frac{W(\gamma)}{f(\gamma)(\gamma - \Delta\gamma)} \right]^{\frac{1}{1+b}} - 1 \right), & \text{else.} \end{cases} \quad (56a)$$

$$(56b)$$

the lower layers, i.e., $U(\gamma) \leq U(\gamma - \Delta\gamma)$ and $U(\gamma) \leq P$. In this region the optimal power allocation $T^*(\gamma)$ is given by the unconstrained minimizer $U(\gamma)$ in (55a). In the solution to $U(\gamma)$ we need to verify that $U(\gamma)$ is nonincreasing in this region, which corresponds to the power distribution $\rho^*(\gamma)$ being nonnegative. Following (15a), we write the cumulative distortion from layers γ and above in the form

$$D^*(\gamma) = (1 + \gamma T(\gamma))^{-b} W(\gamma). \quad (57)$$

Substituting in the unconstrained cumulative power allocation $U(\gamma)$, the cumulative distortion at layer $\gamma - \Delta\gamma$ becomes

$$D^*(\gamma - \Delta\gamma) = \left(\frac{1 + (\gamma - \Delta\gamma)T(\gamma - \Delta\gamma)}{1 + (\gamma - \Delta\gamma)U(\gamma)} \right)^{-b} \cdot [f(\gamma)\Delta\gamma + (1 + \gamma U(\gamma))^{-b} W(\gamma)] \quad (58)$$

which is of the form in (57) if we define $W(\gamma - \Delta\gamma)$ by the recurrence equation

$$W(\gamma - \Delta\gamma) = (1 + (\gamma - \Delta\gamma)U(\gamma))^b \cdot [f(\gamma)\Delta\gamma + (1 + \gamma U(\gamma))^{-b} W(\gamma)]. \quad (59)$$

As the spacing between the layers condenses, in the limit of $\Delta\gamma$ approaching zero, the recurrence (58), (59) become differential equations. The optimal power distribution $\rho^*(\gamma)$ is given by the derivative of the cumulative power allocation

$$\rho^*(\gamma) = -T'^*(\gamma) \quad (60)$$

where $T^*(\gamma)$ is described by solutions in three regions

$$T^*(\gamma) = \begin{cases} 0, & \gamma > \gamma_0 \\ U(\gamma), & \gamma_P \leq \gamma \leq \gamma_0 \\ P, & \gamma < \gamma_P. \end{cases} \quad (61a)$$

$$(61b)$$

$$(61c)$$

In region (61a) when $\gamma > \gamma_0$, corresponding to cases (55a) and (56a), no power is allocated to the layers and (59) simplifies to $W(\gamma) = 1 - F(\gamma)$, where $F(\gamma) \triangleq \int_0^\gamma f(s) ds$ is the cumulative distribution function (cdf) of the channel power gain. The boundary γ_0 is defined by the condition in (56a) which satisfies

$$\gamma_0 f(\gamma_0) + F(\gamma_0) - 1 = 0. \quad (62)$$

Under Rayleigh fading, when $f(\gamma) = \bar{\gamma}^{-1} e^{-\gamma/\bar{\gamma}}$, where $\bar{\gamma}$ is the expected channel power gain, (62) evaluates to $\gamma_0 = \bar{\gamma}$. For other fading distributions, γ_0 may be computed numerically.

In region (61b), when $\gamma_P \leq \gamma \leq \gamma_0$, corresponding to cases (55a) and (56b), the optimal power distribution is described by a set of differential equations. We apply the first-order binomial expansion $(1 + \Delta\gamma)^b \cong 1 + b\Delta\gamma$, and (59) becomes

$$W'(\gamma) = \lim_{\Delta\gamma \rightarrow 0} \frac{W(\gamma) - W(\gamma - \Delta\gamma)}{\Delta\gamma} \quad (63)$$

$$= b \frac{W(\gamma)}{\gamma} - (1 + b) \left[f(\gamma) \left(\frac{W(\gamma)}{\gamma} \right)^b \right]^{\frac{1}{1+b}} \quad (64)$$

which we substitute in (56b) to obtain

$$U'(\gamma) = - \left(\frac{2/\gamma + f'(\gamma)/f(\gamma)}{1 + b} \right) [U(\gamma) + 1/\gamma]. \quad (65)$$

Hence, $U(\gamma)$ is described by a first-order linear differential equation. With the initial condition $U(\gamma_0) = 0$, its solution is given by

$$U(\gamma) = \frac{- \int_{\gamma_0}^{\gamma} \frac{1}{s} \left(\frac{2}{s} + \frac{f'(s)}{f(s)} \right) [s^2 f(s)]^{\frac{1}{1+b}} ds}{(1 + b) [\gamma^2 f(\gamma)]^{\frac{1}{1+b}}} \quad (66)$$

and condition (55b) in the lowest active layer becomes the boundary condition $U(\gamma_P) = P$. In [20], the power distribution in (66) is derived using the calculus of variations method.

Similarly, as $\Delta\gamma \rightarrow 0$, the evolution of the expected distortion in (58) becomes

$$D'(\gamma) = - \frac{b\gamma U'(\gamma)}{1 + \gamma U(\gamma)} D(\gamma) - f(\gamma) \quad (67)$$

$$= \left[\frac{b}{1 + b} \left(\frac{2}{\gamma} + \frac{f'(\gamma)}{f(\gamma)} \right) \right] D(\gamma) - f(\gamma) \quad (68)$$

which is again a first-order linear differential equation. With the initial condition $D(\gamma_0) = W(\gamma_0) = \gamma_0 f(\gamma_0)$, its solution is given by

$$D(\gamma) = \frac{- \int_{\gamma_0}^{\gamma} f(s) \left[\left(\frac{s}{\gamma_0} \right)^2 \frac{f(s)}{f(\gamma_0)} \right]^{\frac{-b}{1+b}} ds + \gamma_0 f(\gamma_0)}{\left[\left(\frac{\gamma}{\gamma_0} \right)^2 \frac{f(\gamma)}{f(\gamma_0)} \right]^{\frac{-b}{1+b}}}. \quad (69)$$

Under Rayleigh fading, for instance, the solutions to $U(\gamma)$ and $D(\gamma)$ are given by

$$U(\gamma) = \frac{\int_{\bar{\gamma}}^{\gamma} \left(\frac{1}{\bar{\gamma}} - \frac{2}{s} \right) [s^{1-b} e^{-s/\bar{\gamma}}]^{\frac{1}{1+b}} ds}{(1 + b) [\gamma^2 e^{-\gamma/\bar{\gamma}}]^{\frac{1}{1+b}}} \quad (70)$$

$$D(\gamma) = \frac{-\frac{1}{\bar{\gamma}} \int_{\bar{\gamma}}^{\gamma} e^{-s/\bar{\gamma}} \left[\left(\frac{s}{\bar{\gamma}} \right)^2 e^{-(s-\bar{\gamma})/\bar{\gamma}} \right]^{\frac{-b}{1+b}} ds + e^{-1}}{\left[\left(\frac{\gamma}{\bar{\gamma}} \right)^2 e^{-(\gamma-\bar{\gamma})/\bar{\gamma}} \right]^{\frac{-b}{1+b}}}. \quad (71)$$

The integrals in (70), (71) can be computed numerically by evaluating the incomplete gamma function.

Finally, in region (61c), when $\gamma < \gamma_P$ corresponding to case (55b), the transmit power P has been exhausted, and no power is allocated to the remaining layers. Hence, the minimum expected distortion is

$$E[D]^* = D(0) = F(\gamma_P) + D(\gamma_P) \quad (72)$$

where the last equality follows from when $\gamma < \gamma_P$ in region (61c), $\rho^*(\gamma) = 0$ and $D(\gamma) = \int_{\gamma}^{\gamma_P} f(s) ds + D(\gamma_P)$.

VIII. RAYLEIGH FADING WITH DIVERSITY

In this section, we consider the optimal power distribution and the minimum expected distortion when the wireless channel undergoes Rayleigh fading with a diversity order of L from the realization of independent fading paths. Specifically, we assume the fading channel is characterized by the Erlang distribution

$$f_L(\gamma) = \frac{(L/\bar{\gamma})^L \gamma^{L-1} e^{-L\gamma/\bar{\gamma}}}{(L-1)!}, \quad \gamma \geq 0 \quad (73)$$

which corresponds to the average of L i.i.d. channel power gains, each under Rayleigh fading with an expected value of $\bar{\gamma}$.

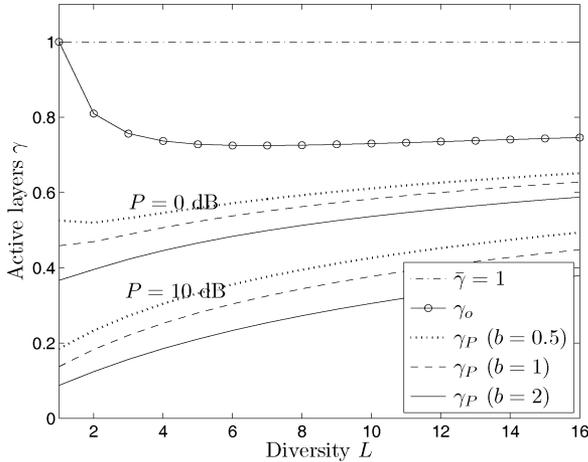


Fig. 12. Span of active layers under optimal power distribution.

The L -diversity system may be realized by having L transmit antennas using isotropic inputs, by relaxing the decoding delay constraint over L fading blocks, or by having L receive antennas under maximal-ratio combining when the power gain of each antenna is normalized by $1/L$.

The optimal power distribution (61) concentrates the transmit power over a range of active layers; the upper and lower boundaries γ_o, γ_P of the span of the active layers are plotted in Fig. 12. A higher SNR P or a larger bandwidth ratio b extends the span further into the lower layers but the upper boundary γ_o remains unperturbed. As L increases, the fluctuation in the channel realization is diminished by the diversity of combining multiple independent fading paths, and the power distribution becomes more concentrated, albeit slowly. By the law of large numbers, at asymptotically large L , we expect all power to concentrate at $\bar{\gamma}$.

Fig. 13 shows the optimal power allocation $\rho^*(\gamma)$. It can be observed that a smaller bandwidth ratio b reduces the spread of the power distribution. In fact, as b approaches zero, the optimal power distribution that minimizes expected distortion converges to the power distribution that maximizes expected capacity. To show the connection, we take the limit in the distortion-minimizing cumulative power distribution in (66)

$$\lim_{b \rightarrow 0} U(\gamma) = \frac{1 - F(\gamma) - \gamma f(\gamma)}{\gamma^2 f(\gamma)} \quad (74)$$

which is equal to the capacity-maximizing cumulative power distribution as derived in [4]. Essentially, from the first-order expansion $e^b \cong 1 + b$ for small b , $E[D] \cong 1 - bE[C]$ when the bandwidth ratio is small, where $E[C]$ is the expected capacity in nats per second, and hence minimizing expected distortion becomes equivalent to maximizing expected capacity. For comparison, the capacity-maximizing power distribution is also plotted in Fig. 13. Note that the distortion-minimizing power distribution is more conservative, and it is more so as b increases, as the allocation favors lower layers in contrast to the capacity-maximizing power distribution.

Fig. 14 shows the minimum expected distortion $E[D]^*$ versus SNR for different diversity orders. With infinite diversity, the

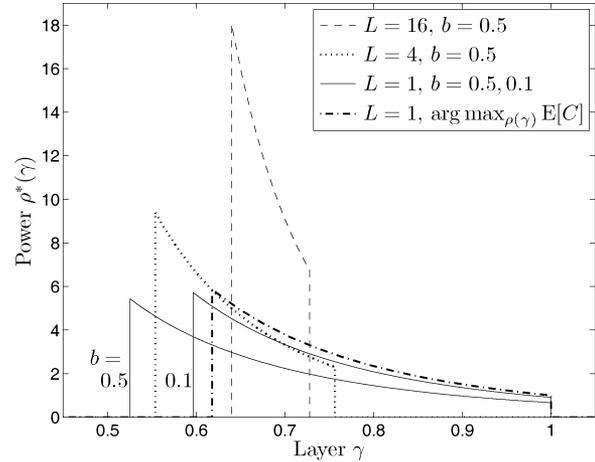


Fig. 13. Optimal power distribution ($P = 0$ dB).

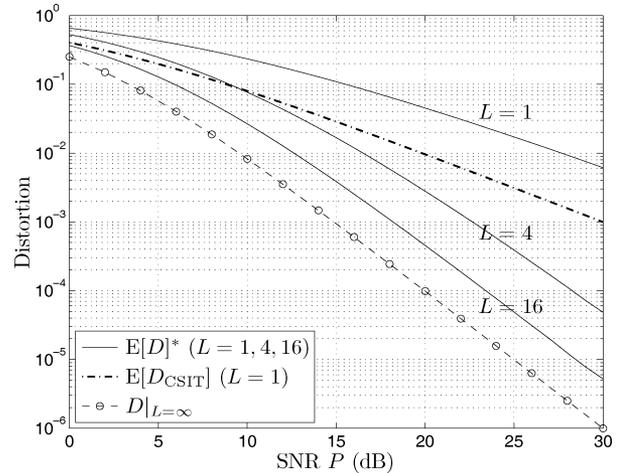


Fig. 14. Minimum expected distortion ($b = 2$).

channel power gain becomes constant at $\bar{\gamma}$, and the distortion is given by

$$D|_{L=\infty} = (1 + \bar{\gamma}P)^{-b}. \quad (75)$$

In the case when there is no diversity ($L = 1$), a lower bound to the expected distortion is also plotted. The lower bound assumes the system has CSIT, which allows the transmitter to concentrate all power at the realized layer to achieve the expected distortion

$$E[D_{CSIT}] = \int_0^\infty e^{-\gamma} (1 + \gamma P)^{-b} d\gamma. \quad (76)$$

Note that at high SNR, the performance benefit from diversity exceeds that from CSIT, especially when the bandwidth ratio b is large. In particular, in terms of the distortion exponent Δ [11], it is shown in [13] that in a MISO or SIMO channel, layered broadcast coding achieves

$$\Delta \triangleq - \lim_{P \rightarrow \infty} \frac{\log E[D]}{\log P} = \min(b, L) \quad (77)$$

where L is the total diversity order from independent fading blocks and antennas. Moreover, the layered broadcast coding distortion exponent is shown to be optimal and CSIT does not

improve Δ , whereas diversity increases Δ up to a maximum as limited by the bandwidth ratio b .

IX. CONCLUSION

We considered the problem of source–channel coding over a delay-limited fading channel without CSI at the transmitter, and derived the optimal power allocation that minimizes the end-to-end distortion in the layered broadcast coding transmission scheme with successive refinement. In the allocation of transmission power between two layers of codewords in a two-state fading channel, the optimal allocation that minimizes the expected distortion has a particular structure that lends itself to be generalized to the cases when the channel has multiple discrete fading states or a continuous fading distribution. Specifically, the optimal two-layer allocation assigns power first to the higher layer, up to a power ceiling that depends only on the channel fading distribution but independent of the total available power; any surplus over the power ceiling is allocated to the lower layer. When the channel has multiple discrete fading states, we write the minimum expected distortion as a set of recurrence relations, and in each recurrence step the two-layer optimization procedure solves the power allocation between the current layer and the aggregate higher layer. The optimization framework is extended to consider convex distortion cost functions with convex constraints by posing the minimization as a convex optimization problem.

We applied the power allocation algorithms to the pmf of a discretized Rayleigh fading distribution. We observed that the optimal power allocation is concentrated around the middle layers, and within this range the lower layers are assigned more power than the higher ones. As the SNR increases, the allocations of the higher layers remain unchanged, and the extra power is allocated to the idle lower layers. The distortion-minimizing power distribution, therefore, is conservative: it is more beneficial to utilize the lower layers, despite their weaker channel gains, than the higher layers as the latter have larger risks of being in outage.

We also derived the optimal power distribution that minimizes the expected distortion when the fading distribution of the channel is given by a continuous pdf. We computed the optimal power distribution for Rayleigh-fading channels with diversity order L , and showed that increasing the diversity L concentrates the power distribution towards the expected channel power gain $\bar{\gamma}$, while a larger bandwidth ratio b spreads the power distribution further into the lower layers. On the other hand, in the limit as b tends to zero, the optimal power distribution that minimizes expected distortion converges to the power distribution that maximizes expected capacity. While the expected distortion can be improved by acquiring CSIT or increasing the diversity order, it is shown that at high SNR, the performance benefit from diversity exceeds that from CSIT, especially when the bandwidth ratio b is large. Under continuous channel fading, we focused on minimizing the expected distortion, which is a linear cost function of the distortion realizations. Future research may include considering the minimization of a general convex distortion cost function under continuous channel fading distributions.

APPENDIX

Algorithm 1 Multiple-Layer Power Allocation

```

1: ALLOC( $M - 1, p_M, \gamma_M, p_{M-1}, \gamma_{M-1}$ )  ▷ Start from top
2: procedure ALLOC( $i, w, \beta, u, \alpha$ )
3:   Compute  $U$  from  $w, \beta, u, \alpha$ 
4:   if  $i = 1$  then  ▷ Bottom layer
5:      $T_2^* \leftarrow \min(U, P)$ 
6:     return
7:   end if
8:   if  $U < P$  then  ▷ Within total power  $P$ 
9:     Compute  $W$  from  $U, w, \beta, u, \alpha$ 
10:    ALLOC( $i - 1, W, \alpha, p_{i-1}, \gamma_{i-1}$ ) ▷ Unconstrained
11:    if  $T_i^* \geq U$  then
12:       $T_{i+1}^* \leftarrow U$   ▷  $U$  is feasible
13:    return
14:    end if
15:  end if
16:  ALLOC( $i - 1, w, \beta, p_{i-1} + u, \gamma_{i-1}$ )  ▷ Constrained
17:   $T_{i+1}^* \leftarrow T_i^*$ 
18: end procedure

```

REFERENCES

- [1] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *IRE Nat. Conv. Rec.*, 1959, pp. 142–163, pt 4.
- [2] T. M. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 1, pp. 2–14, Jan. 1972.
- [3] S. Shamai (Shitz), "A broadcast strategy for the Gaussian slowly fading channel," in *Proc. IEEE Int. Symp. Information Theory*, Ulm, Germany, Jun. 1997, p. 150.
- [4] 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.
- [5] P. A. Whiting and E. M. Yeh, "Broadcasting over uncertain channels with decoding delay constraints," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 904–921, Mar. 2006.
- [6] T. J. Goblick, Jr., "Theoretical limitations on the transmission of data from analog sources," *IEEE Trans. Inf. Theory*, vol. IT-11, no. 4, pp. 558–567, Oct. 1965.
- [7] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.
- [8] 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.
- [9] U. Mittal and N. Phamdo, "Hybrid digital-analog (HDA) joint source-channel codes for broadcasting and robust communications," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1082–1102, May 2002.
- [10] Z. Reznic, M. Feder, and R. Zamir, "Distortion bounds for broadcasting with bandwidth expansion," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3778–3788, Aug. 2006.
- [11] 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.
- [12] D. Gündüz and E. Erkip, "Source and channel coding for quasi-static fading channels," in *Proc. Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, Oct./Nov. 2005, pp. 18–22.
- [13] D. Gündüz and E. Erkip, "Joint source-channel codes for MIMO block fading channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 116–134, Jan. 2008.
- [14] G. Caire and K. Narayanan, "On the distortion SNR exponent of hybrid digital–analog space–time coding," *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2867–2878, Aug. 2007.
- [15] F. Etemadi and H. Jafarkhani, "Optimal layered transmission over quasi-static fading channels," in *Proc. IEEE Int. Symp. Information Theory*, Seattle, WA, Jul. 2006, pp. 1051–1055.
- [16] S. Sesia, G. Caire, and G. Vivier, "Lossy transmission over slow-fading AWGN channels: A comparison of progressive, superposition and hybrid approaches," in *Proc. IEEE Int. Symp. Information Theory*, Adelaide, Australia, Sep. 2005, pp. 224–228.

- [17] K. E. Zachariadis, M. L. Honig, and A. K. Katsaggelos, "Source fidelity over fading channels: Erasure codes versus scalable codes," in *Proc. IEEE Globecom Conf.*, St. Louis, MO, Nov. 2005, vol. 5, pp. 2558–2562.
- [18] C. T. K. Ng, D. Gündüz, A. J. Goldsmith, and E. Erkip, "Recursive power allocation in Gaussian layered broadcast coding with successive refinement," in *Proc. IEEE Int. Conf. Communications*, Glasgow, Scotland, Jun. 2007, pp. 889–896.
- [19] C. T. K. Ng, D. Gündüz, A. J. Goldsmith, and E. Erkip, "Minimum expected distortion in Gaussian layered broadcast coding with successive refinement," in *Proc. IEEE Int. Symp. Information Theory*, Nice, France, Jun. 2007, pp. 2226–2230.
- [20] 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.
- [21] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley-Interscience, 1991.
- [22] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 269–275, Mar. 1991.
- [23] B. Rimoldi, "Successive refinement of information: Characterization of the achievable rates," *IEEE Trans. Inf. Theory*, vol. 40, no. 1, pp. 253–259, Jan. 1994.
- [24] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [25] M. Grant and S. Boyd, "CVX: Matlab Software for Disciplined Convex Programming" (Web Page and Software) 2008 [Online]. Available: <http://stanford.edu/~boyd/cvx>
- [26] M. C. Grant and S. P. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control (A Tribute to M. Vidyasagar)*, V. Blondel, S. Boyd, and H. Kimura, Eds. New York: Springer, 2008, Lecture Notes in Control and Information Sciences, pp. 95–110.
- [27] F. Etemadi and H. Jafarkhani, "Optimal rate and power allocation for layered transmission with superposition coding," in *Proc. IEEE Data Compression Conf.*, Snowbird, UT, Mar. 2007, p. 380.

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.

Since 2009, he has been a Member of Technical Staff at Bell Laboratories, 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.

Deniz Gündüz (S'03–M'08) received the B.S. degree in electrical and electronics engineering from the Middle East Technical University, Ankara, Turkey, in 2002 and the M.S. and Ph.D. degrees in electrical engineering from Polytechnic Institute of New York University (formerly Polytechnic University), Brooklyn, NY in 2004 and 2007, respectively.

He is currently a Consulting Assistant Professor at the Department of Electrical Engineering, Stanford University, Stanford, CA, and a Postdoctoral Research Associate in the Department of Electrical Engineering, Princeton University, Princeton, NJ. In 2004, he was a summer researcher in the laboratory of information theory (LTH) at EPFL in Lausanne, Switzerland. His research interests lie in the areas of communication theory and information theory with special emphasis on joint source–channel coding, cooperative communications, network security and cross-layer design.

Dr. Gündüz is the recipient of the 2008 Alexander Hessel Award of Polytechnic University given to the best Ph.D. dissertation, and a coauthor of the

paper that received the Best Student Paper Award at the 2007 IEEE International Symposium on Information Theory.

Andrea J. Goldsmith (S'90–M'93–SM'99–F'05) received the B.S., M.S., and Ph.D. degrees in electrical engineering from University of California, Berkeley.

She is a Professor of Electrical Engineering at Stanford University, Stanford, CA, and was previously an assistant professor of Electrical Engineering at the California Institute of Technology (Caltech), Pasadena. She is also founder and CTO of Quantenna Communications, Inc., and has previously held industry positions at Maxim Technologies, Memorylink Corporation, and AT&T Bell Laboratories. Her research includes work on wireless information and communication theory, MIMO systems and multihop networks, sensor networks, cross-layer wireless system design, and wireless communications for distributed control. She is author of the book *Wireless Communications* and coauthor of the book *MIMO Wireless Communications*, both published by Cambridge University Press.

Dr. Goldsmith is a Fellow of Stanford. She has received several awards for her research, including the National Academy of Engineering Gilbreth Lectureship, the Alfred P. Sloan Fellowship, the Stanford Terman Fellowship, the National Science Foundation CAREER Development Award, and the Office of Naval Research Young Investigator Award. In addition, she was a corecipient of the 2005 IEEE Communications Society and Information Theory Society joint paper award. She currently serves 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. She 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 the IEEE Communications Society, the president of the IEEE Information Theory Society, and was the Technical Program Cochair for the 2007 IEEE International Symposium on Information Theory. She also founded the student committee of the IEEE Information Theory Society, is an inaugural recipient of Stanford's postdoc mentoring award, and will serve as Stanford's faculty senate chair in 2009/2010.

Elza Erkip (S'93–M'96–SM'05) received the B.S. degree in electrical and electronic engineering from the Middle East Technical University, Ankara, Turkey, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA.

Currently, she is an Associate Professor of Electrical and Computer Engineering at the Polytechnic Institute of New York University, Brooklyn, NY. In the past, she has held positions at Rice University and at Princeton University. Her research interests are in information theory, communication theory, and wireless communications.

Dr. Erkip received the NSF CAREER Award in 2001, the IEEE Communications Society Rice Paper Prize in 2004, and the ICC Communication Theory Symposium Best Paper Award in 2007. She coauthored a paper that received the IEEE International Symposium on Information Theory (ISIT) Student Paper Award in 2007. Currently, she is an Associate Editor of IEEE TRANSACTIONS ON INFORMATION THEORY, an Associate Editor of IEEE TRANSACTIONS ON COMMUNICATIONS, the Co-Chair of the GLOBECOM 2009 Communication Theory Symposium, and the Publications Chair of ITW 2009, Taormina. She was a Publications Editor of IEEE TRANSACTIONS ON INFORMATION THEORY during 2006–2009, a Guest Editor of IEEE SIGNAL PROCESSING MAGAZINE in 2007, the MIMO Communications and Signal Processing Technical Area Chair of the Asilomar Conference on Signals, Systems, and Computers in 2007, and the Technical Program Co-Chair of Communication Theory Workshop in 2006.