Source Fidelity over Fading Channels: Performance of Erasure and Scalable Codes

Konstantinos E. Zachariadis, *Member, IEEE*, Michael L. Honig, *Fellow, IEEE*, and Aggelos K. Katsaggelos, *Fellow, IEEE*

***Abstract*—We consider the transmission of a Gaussian source through a block fading channel. Assuming each block is decoded independently, the received distortion depends on the trade-off between quantization accuracy and probability of outage. Namely, higher quantization accuracy requires a higher channel code rate, which increases the probability of outage. We first treat an outage as an erasure, and evaluate the received mean distortion with erasure coding across blocks as a function of the code length. We then evaluate the performance of scalable, or multi-resolution coding in which coded layers are superimposed within a coherence block, and the layers are sequentially decoded. Both the rate and power allocated to each layer are optimized. In addition to analyzing the performance with a finite number of layers, we evaluate the mean distortion at high Signal-to-Noise Ratios as the number of layers becomes infinite. As the block length of the erasure code increases to infinity, the received distortion converges to a deterministic limit, which is less than the mean distortion with an infinite-layer scalable coding scheme. However, for the same standard deviation in received distortion, infinite layer scalable coding performs slightly better than erasure coding, and with much less decoding delay.**

***Index Terms*—Source-channel coding, scalable coding, fading channel, broadcast channel, rate distortion.**

I. INTRODUCTION

TRANSMISSION of a continuous source, such as an image or video, through a fading channel must account for distortion due to both quantization and channel-induced errors. If the channel code-word is sufficiently long, i.e., extends over multiple fading cycles, then the ergodic capacity can be used to determine the achievable information rate, which in turn determines the (fixed) source rate. However, a slowly-varying fading channel and/or short channel code-words, relative to the channel variations, can lead to *outages*, which substantially increase the received distortion. The trade-off between distortion due to source quantization and channel-induced distortion in the presence of fading has been studied for different source models in [1]–[5]. Other related work on optimization of source and channel coding parameters in different contexts is presented in [6]–[13].

In this paper we consider the use of erasure and scalable codes, which can be used to compensate for channel fading. Namely, erasure codes can eliminate errors due to outages, whereas scalable codes attempt to match the transmitted rate to the channel conditions. To provide insight, we consider a model in which a continuous Gaussian source is transmitted through a block Rayleigh fading channel. In the absence of delay constraints, the distortion is minimized by coding over coherence blocks. The source rate should then be matched to the channel code rate, which can be no greater than the ergodic capacity. However, with a decoding delay constraint the ergodic capacity cannot always be achieved, making it more difficult to identify an appropriate source rate.[1] Moreover, in some applications it may not be computationally feasible to code optimally over coherence blocks.

The erasure coding scheme contains an inner code, which achieves the capacity for a particular Signal-to-Noise Ratio (SNR). The achievable rate can then be characterized in terms of the outage capacity [15]. Namely, an outage occurs when the actual SNR falls below the SNR associated with the code rate. Increasing the source rate therefore decreases the distortion due to quantization, but increases the probability of outage due to the associated increase in channel code rate. An outer erasure code is concatenated with the inner code to correct the errors due to outages. Although this concatenated coding scheme is suboptimal (i.e., does not achieve the ergodic capacity), it is relatively simple. We optimize the inner channel code rate to maximize the capacity of the resulting erasure channel, assuming infinite-length erasure codewords, and compare the distortion with that associated with the ergodic capacity. Numerical results show that the gap between distortions for the two schemes is relatively small for small SNRs, but increases with SNR. We also characterize the standard deviation of the distortion (over channel realizations) as a function of the length of the erasure code.

We then consider the performance of scalable source coding, where for each coherence block the source is partitioned into layers representing successive refinements of the quantizer. Each layer, corresponding to a particular quantizer, can have a different rate and power, and all layers are transmitted simultaneously as a superposition code [16, Ch. 14]. The

Paper approved by F. Alajaji, the Editor for Source and Source/Channel Coding of the IEEE Communications Society. Manuscript received July 13, 2006; revised April 23, 2007 and August 14, 2007. This work was supported in part by ARO under grant DAAD190310119 and NSF under grant CCR-0073686, and was presented at Milcom 2004 and Globecom 2005. This work was performed when K. E. Zachariadis was with the Electrical and Computer Engineering Department (now EECS), Northwestern University.

K. E. Zachariadis is with the Kellogg School of Management, Northwestern University, Evanston, IL 60201-2001, USA (e-mail: k-zachariadis@kellogg.northwestern.edu).

M. L. Honig and A. K. Katsaggelos are with the Electrical Engineering and Computer Science Department, Northwestern University, Evanston IL 60208-3118, USA (e-mail: {mh, aggk}@ece.northwestern.edu).

Digital Object Identifier 10.1109/TCOMM.2008.060387.

[1]If the source and channel bandwidths are the same, then for our model uncoded (analog) transmission of the source symbols is optimal [10], [14]. However, this is not true for mismatched bandwidths, and for pre-quantized sources.

layers are sequentially decoded, and the loss of a particular layer implies the loss of all succeeding layers. The received distortion for a given channel state then depends on how many layers are decodable. Here the Gaussian source provides additional flexibility since it is inherently *refinable* [17]. The effective code rate and distortion for each coherence block therefore depend upon the random channel state.

We first evaluate the mean distortion of a scalable code with two layers, and present a numerical method for evaluating the mean distortion with more layers. We then evaluate the mean distortion with an infinite number of layers at high SNRs. In that case, we obtain closed-form expressions for the power and rate allocations across layers along with the minimum mean distortion, and the associated standard deviation. Finally, we present numerical results, which compare the distortions for both schemes as a function of SNR. Specifically, the *deterministic*, or channel-independent, distortion with an *infinite-length* erasure code is less than the *mean* distortion, averaged over channel realizations, with infinite-layer (finite-length) scalable coding. This gap increases with SNR.

With a finite-length erasure code the mean distortion depends on the particular sequence of channel realizations. The associated variance, with respect to the channel, decreases to zero as the code length tends to infinity. Numerical results show that if the length of the erasure code is adjusted so that the standard deviation of the distortion is the same as that produced by the scalable code, then scalable coding gives a small improvement in mean distortion relative to the erasure code. Furthermore, since the erasure code still spans multiple coherence blocks, the associated decoding delay is much less for the scalable coding scheme.

Scalable coding has been previously proposed for the transmission of a single source over a Rayleigh fading channel [11], [12], [18]. In that work the transmitter views the communications medium as a degraded Gaussian channel with a continuum of receivers (corresponding to an infinite number of layers), each corresponding to a different fading level. There the problem is to allocate rates across layers to maximize total throughput, whereas here we allocate both rates and powers to minimize the expected distortion.

Related work on the performance of scalable coding for the same model as that considered here is presented in [13], [19]. There the performance of the scalable coding scheme considered here is compared with a *sequential* layered coding scheme in which refinement layers are transmitted sequentially over successive time slots in a single coherence block. (Sequential layering is also considered in [20], where the rates and channel codewords (but not powers) across layers are numerically optimized.) The performance metric in [13], [19] is the *distortion exponent*, introduced in [4], which is the asymptotic (large-SNR) decay rate of the mean distortion with SNR. It is shown in [13], [19] that the infinite-layer scalable coding scheme achieves the optimal distortion exponent. The distortion exponent is used to compare the performance of other related coding schemes in [10], [21], [22].

Based on our finite-SNR results, we find that the distortion exponents for infinite-length erasure coding, optimal coding (corresponding to the ergodic capacity), and infinite-layer scalable coding are the same. Furthermore, this exponent is also obtained when transmitting the Gaussian source over an Additive White Gaussian Noise (AWGN) channel at the channel capacity. Hence this metric cannot be used in general to infer relative performance at finite SNRs. (See also [10].)

Other related work on hybrid digital-analog transmission for parallel fading channels is presented in [10], [21], [23]–[25]. A broadcast (scalable) code is proposed in [24], although only power is optimized across the layers, which are assigned the same source rate. Earlier work on progressive layered multimedia transmission has appeared in [26]–[28], where the focus is mostly on practical implementations of the proposed schemes. Enhancement of scalable video coding to provide fine granular scalability is discussed in [29], [30]. That work serves as practical motivation for considering an infinite-layer scalable code.

The outline of the paper is as follows. In the next section, the system model and preliminary results are presented, assuming single-layer encoding of the source. Section III discusses the performance with erasure codes. The performance of scalable coding with two layers is analyzed in Section IV. This analysis is extended to more than two layers in Section V, and to a continuum of layers in Section VI. A numerical comparison of the performance of erasure and scalable coding is presented in Section VII. Finally, high SNR performance results are derived in the appendix.

II. Single Layer

We consider transmitting an *i.i.d.* Gaussian source sequence $\mathbf{X} = \{X_i\}_{i=1}^K$ over a block Rayleigh fading channel. Each X_i is a real Gaussian random variable with zero mean and variance $\sigma_x^2 = 1$. The source encoding function maps $\mathbf{X}$ to an index i in the set $\mathcal{C} = \{1, ..., e^{KR}\}$, and the decoder maps an index in $\mathcal{C}$ to an estimate $\hat{\mathbf{X}}$. Distortion is measured in the mean squared error (MSE) sense, *i.e.*,

$$MSE = \frac{1}{K}\sum_{i=1}^{K}\left(X_i - \hat{X}_i\right)^2,$$

where $\hat{X}_i$ is the reconstructed value of the ith sample. The rate-distortion function is [16]

$$D(R) = e^{-2R}, \tag{1}$$

where R is the source information rate in nats per source sample and D is the distortion.[2]

The channel capacity, conditioned on the channel gain h, is

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

in nats per second per Hz (per dimension), where P is the transmitted power, and N is the noise power. Since the channel gain h is exponentially distributed, the capacity is a random variable with cumulative distribution function (cdf) [15]

$$\Pr\left[C\left(\frac{hP}{N}\right) < R_{\rm t}\right] = 1 - \exp\left(-\gamma^{-1}\left(e^{2R_{\rm t}} - 1\right)\right), \tag{3}$$

where $\gamma = \frac{HP}{N}$ is the (average) SNR, and $H = E[h]$.

[2]Throughout the paper we are going to use $\exp(x)$ and e^x interchangeably, to denote the value of the exponential function at x.

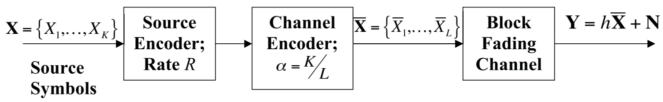


Fig. 1. System Model: K source symbols span a coherence time, h is the channel gain, and $\mathbf{N}$ is AWGN.

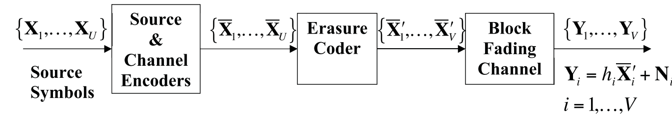


Fig. 2. System Model with Erasure Coding: Each $\mathbf{X}_i$, $i = 1, \ldots, U$, contains K source symbols. The erasure code rate is U/V.

For a desired transmission rate R_{t} we will denote the outage probability as p. Referring to Fig. 1, during each (fixed) coherence time T, the source emits a sequence of K source symbols, represented by the vector $\mathbf{X}$, and the channel encoder maps the output of the source encoder to a sequence of L channel symbols $\bar{\mathbf{X}} = \{\bar{X}_1, ..., \bar{X}_L\}$. Since we use the channel L times to represent KR source nats, the channel code rate, $R_{\mathrm{t}} = \frac{KR}{L} = \alpha R$, where the "processing gain" $\alpha = K/L$ [1], [4]. We will assume optimal channel coding, in the sense of achieving the outage capacity.

Suppose the source is transmitted over the fading channel with (source) rate R. Two events can occur; (i) a bad channel state is encountered with probability p, given by (3), which results in a distortion at the receiver equal to the source variance; (ii) $\alpha R < C(hP/N)$ in which case the receiver correctly decodes the message and the distortion is given by $D(R)$ in (1). Hence the expected distortion at the receiver is

$$\begin{aligned} E[D] &= D(R) \cdot \Pr\left[C\left(\frac{hP}{N}\right) \geq \alpha R\right] \\ &\quad +1 \cdot \Pr\left[C\left(\frac{hP}{N}\right) < \alpha R\right] \\ &= (D(R) - 1) \cdot (1 - p) + 1. \end{aligned} \tag{4}$$

We wish to minimize $E[D]$ over the source rate R. The optimal R satisfies the necessary condition

$$\alpha e^{2R(\alpha+1)} - \alpha e^{2R\alpha} - \gamma = 0.$$

Solving this for the special case $\alpha = 1$ yields

$$R^* = \frac{1}{2}\log\left(\frac{1}{2} + \frac{\sqrt{1+4\gamma}}{2}\right), \tag{5}$$

and

$$E[D(R^*)] = \frac{1-\sqrt{1+4\gamma}}{1+\sqrt{1+4\gamma}} \exp\left(\frac{1-\sqrt{1+4\gamma}}{2\gamma}\right) + 1. \tag{6}$$

In analogy with (4) we can also compute the moment generating function of $D(R^*)$ for $\alpha = 1$, which is given by

$$\begin{aligned} \Phi(s) &= E[e^{sD(R^*)}] \\ &= e^{sD(R^*)}(1-p) + e^s\, p \\ &= \exp\left(\frac{2(s-1)}{1+\sqrt{1+4\gamma}}\right) \\ &\quad + \exp(s)\left[1 - \exp\left(\frac{1-\sqrt{1+4\gamma}}{2\gamma}\right)\right]. \end{aligned}$$

Since $E[D^2(R^*)] = \Phi''(0)$ from the above expression we can compute

$$E[D^2(R^*)] = \frac{2\left(1 - 2\gamma - \sqrt{1+4\gamma}\right)}{\left(1+\sqrt{1+4\gamma}\right)^2} e^{\frac{1-\sqrt{1+4\gamma}}{2\gamma}} + 1.$$

Note that for an Additive White Gaussian Noise (AWGN) channel the capacity is

$$C_{\mathrm{AWGN}} = \frac{1}{2}\log(1+\gamma),$$

the optimal source rate is

$$R_{\mathrm{AWGN}} = C_{\mathrm{AWGN}}/\alpha = \frac{1}{2\alpha}\log(1+\gamma),$$

and the (deterministic) distortion is

$$D_{\mathrm{AWGN}} = (1+\gamma)^{-\frac{1}{\alpha}}. \tag{7}$$

For the block Rayleigh fading channel, if the transmitter codes over many blocks, then the optimal channel rate is the ergodic capacity

$$C_{\mathrm{erg}} = E\left[\frac{1}{2}\log\left(1 + \frac{Ph}{N}\right)\right] = \frac{1}{2}e^{\frac{1}{\gamma}}\ \mathrm{E}_1\left(\frac{1}{\gamma}\right),$$

where $\mathrm{E}_1(x) = \int_1^{+\infty} \frac{e^{-xt}}{t}dt$. The distortion in this case is

$$\begin{aligned} D_{\mathrm{erg}} &= e^{-2C_{\mathrm{erg}}/\alpha} \\ &= \exp\left[-\frac{e^{\frac{1}{\gamma}}}{\alpha}\ \mathrm{E}_1\left(\frac{1}{\gamma}\right)\right]. \end{aligned} \tag{8}$$

III. Erasure Codes

We now view the channel as a symbol erasure channel, where each symbol refers to a channel codeword $\bar{\mathbf{X}}$, and the erasure probability is the outage probability due to fading. For the time being, if we assume that the erasure code spans an infinite number of blocks, then the achievable rate is given by the capacity [16]

$$C_{\mathrm{era}} = \log(|\mathcal{C}|)(1 - p_{\mathrm{era}})/L = \alpha R e^{-\gamma^{-1}\left(e^{2\alpha R}-1\right)}, \tag{9}$$

where $|\mathcal{C}|$ is the cardinality of the symbol set $\mathcal{C}$, *i.e.*, e^{KR} in our case, and $p_{\mathrm{era}} = p = 1 - e^{-\gamma^{-1}\left(e^{2\alpha R}-1\right)}$ is the outage probability. Since the capacity is deterministic in this case, the distortion $D_{\mathrm{era}} = e^{-2C_{\mathrm{era}}/\alpha}$.

The source rate that maximizes C_{era} satisfies $2\alpha R e^{2\alpha R} = \gamma$, and the unique real solution is given by

$$R^*_{\mathrm{era}} = \frac{1}{2\alpha}\mathrm{W}(\gamma), \tag{10}$$

where $\mathrm{W}(\cdot)$ is Lambert's W function (also called Omega function), and is the inverse function of $f(x) = xe^x$. The corresponding received distortion is

$$D^*_{\mathrm{era}} = \exp\left[-\frac{\mathrm{W}(\gamma)}{\alpha}e^{-\frac{e^{\mathrm{W}(\gamma)}-1}{\gamma}}\right]. \tag{11}$$

We now evaluate the mean distortion when a finite-length erasure code is used. Referring to Fig. 2, the erasure code maps U code words generated by the inner code,

$\{\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2, ..., \bar{\mathbf{X}}_U\}$, to V code words $\{\bar{\mathbf{X}}'_1, \bar{\mathbf{X}}'_2, ..., \bar{\mathbf{X}}'_V\}$. The code words $\{\bar{\mathbf{X}}'_1, \bar{\mathbf{X}}'_2, ..., \bar{\mathbf{X}}'_V\}$ traverse the channel sequentially; each spans a channel coherence block of duration T, as before. In this case the information (source) rate is RU/V, whereas the inner code rate remains αR. The source bits are assumed to be interleaved across the V code words, so that the distortion statistics do not change from block to block. In what follows, we assume a linear (V, U, d) erasure channel code [31] with minimum distance d, i.e., the U information codewords can be recovered from $(V - d + 1)$ codewords. Furthermore, we assume that $d = V - U + 1$, corresponding to a Maximum Distance Separable (MDS) code. In that case, the U information codewords can be recovered from *any* subset of U correctly decoded codewords.

The expected distortion corresponding to source rate R and erasure code rate U/V is given by

$$E[D] = D\left(\frac{U}{V}R\right)\Pr[\mathbb{C}] + (1 - \Pr[\mathbb{C}]), \qquad (12)$$

where $\mathbb{C}$ is the event that at least U code words are correctly decoded, i.e.,

$$\begin{aligned}\Pr[\mathbb{C}] &= \sum_{i=U}^{V}\binom{V}{i}(1-p)^i p^{V-i} \\ &= \sum_{i=0}^{V-U}\binom{V}{i}p^i(1-p)^{V-i},\end{aligned}$$

where p is the outage probability for each block, given by (3).

Given U and V, we select R to minimize $E[D]$ in (12). As $V \to \infty$ with fixed U/V, $R \to R^*_{\text{era}}$ in (10) and the corresponding outage, or erasure probability $p_{\text{era}} = 1 - U/V$. For $\alpha = 1$, and $\gamma = 10$, 15, 20 dB, the corresponding optimal rate from (10) is $R^*_{\text{era}} = 0.87, 1.26, 1.69$ nats per source sample, respectively, and the corresponding optimal erasure code rates can be approximated as $U/V = 3/4, 7/10, 4/5$. Fig. 3 shows the minimum $E[D]$ and the corresponding standard deviation versus V for fixed U/V. We see that $E[D]$ decreases with V (and tends to a finite limit, designated by the dotted straight line in the figure), and as expected, the standard deviation tends to zero. If $U = V$, corresponding to no erasure code, then the optimal rate $R^* = 0.65, 0.90, 1.20$ nats per source sample and $E[D]^* = -3.54, -5.40, -7.51\,dB$ for $\gamma = 10, 15, 20$ dB respectively, which is significantly larger than that achievable with an erasure code.

From the expression (11) we can evaluate the distortion exponent of the erasure coding scheme, given by [4]

$$\Delta = -\lim_{\gamma\to\infty}\frac{\log E[D]}{\log\gamma}. \qquad (13)$$

That is, the mean distortion is assumed to decrease as $\gamma^{-\Delta}$, and Δ reflects the diversity achieved by the coding scheme. For the erasure coding scheme with infinite-length codewords, $E[D]$ is replaced by D^*_{era}, and evaluating (13) gives $\Delta = 1/\alpha$. Furthermore, replacing $E[D]$ by either D_{erg} in (8) or D_{AWGN} in (7) gives the same distortion exponent. Hence the erasure coding scheme achieves the optimal distortion exponent (i.e., achieved with an optimal block code). We defer a numerical comparison of distortion with erasure and optimal coding for finite SNRs to Section VII.

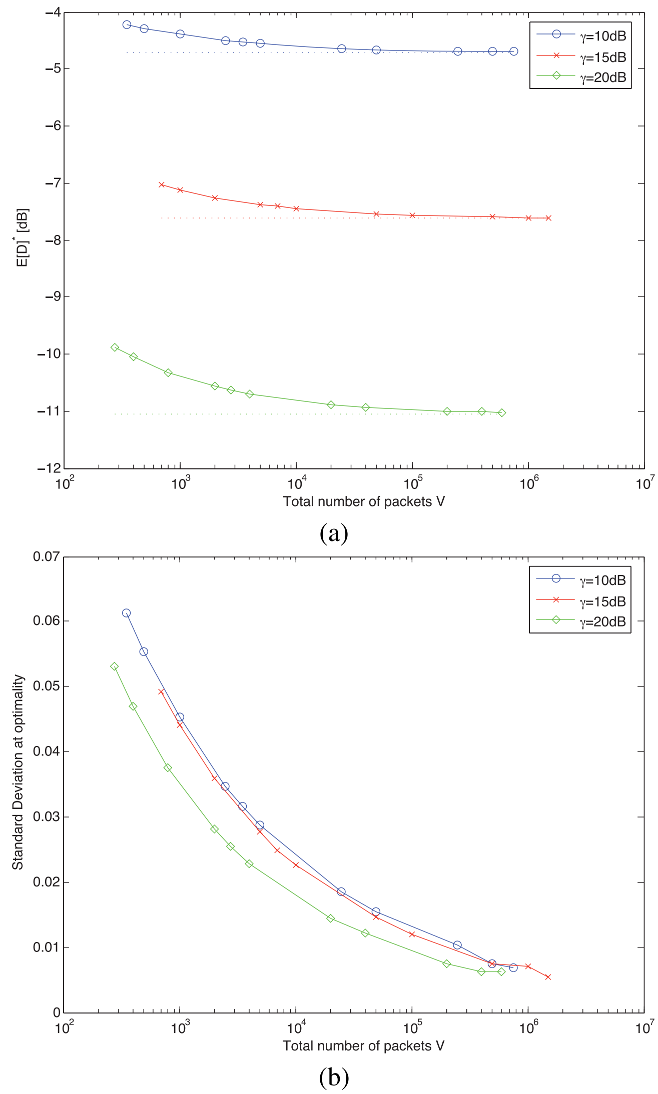


Fig. 3. (a) Minimum expected distortion and (b) the corresponding standard deviation vs V for various values of γ (the dotted lines in (a) are the D^* corresponding to γ).

IV. Scalable Coding

In this section we apply scalable or broadcast coding [32] to achieve performance gains relative to single-layer transmission with a single block decoding delay. Namely, a two-layer code for the broadcast channel can be used for the single-user channel considered here, where the channel is modeled as being either in a "bad" state or a "good" state with some probability. For a target transmission rate R_{t}, if the channel is in a "bad" state we have an outage and if the channel is in a "good" state, the packet is decoded. Our goal is to achieve some minimum information rate when the channel is bad, while being able to get a higher rate when the channel is good.

We therefore consider a scalable code in which the source vector $\mathbf{X}$ is mapped to two indices; the first index i_1 is chosen from the set $\mathcal{C}_c = \{1, ..., e^{KR_1}\}$ and the second index i_2 is taken from the set $\mathcal{C}_r = \{1, ..., e^{K(R-R_1)}\}$. Correct decoding of $i_1 \in \mathcal{C}_c$ (the coarse layer) allows reconstruction of the source with asymptotic distortion $D_1 = D(R_1) = e^{-2R_1}$.

Correct decoding of *only* index $i_2 \in \mathcal{C}_r$ (the refinement layer) results in a much greater distortion than the rate-distortion bound, *i.e.*, $D_2 > e^{-2(R-R_1)}$. In what follows, we assume that the index i_2 alone provides no information about the source. A Gaussian source is "successively refinable" [17], meaning that reception of both indices i_1 and i_2 gives asymptotic distortion $D_0 = D(R) = e^{-2R}$, which achieves the rate-distortion bound.

In the two-layer scalable coding scheme the source is coded into coarse and refinement indices. Each index is the input to an optimal channel coder with code rate α, as in the previous section. Let $\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2$ denote the channel codewords corresponding to indices i_1, and i_2, respectively. We assume that $\bar{\mathbf{X}}_1$ has power $\beta_1 P$, $0 \le \beta_1 \le 1$, and $\bar{\mathbf{X}}_2$ has power $(1-\beta_1)P$. The sum of the codewords $\bar{\mathbf{X}} = \bar{\mathbf{X}}_1 + \bar{\mathbf{X}}_2$, which has power P, is transmitted through the channel. This is an example of a superposition code [16]. In this scheme the coarse layer has rate R_1 and the refinement layer has rate $R - R_1$. If the channel realization does not allow reliable transmission at rate R_1, then the received distortion is the source variance. If the channel does support rate R_1, then the coarse index is decoded first, treating the refinement codeword as noise. The codeword $\bar{\mathbf{X}}_1$ is then subtracted from the received codeword to obtain codeword $\bar{\mathbf{X}}_2$. If the channel can support the additional rate $R - R_1$, then $\bar{\mathbf{X}}_2$ can be correctly decoded to give overall distortion $D(R)$. Otherwise, successful decoding of $\bar{\mathbf{X}}_1$ alone gives distortion $D(R_1)$.

Let $\mathbb{A}$ denote the event that $\bar{\mathbf{X}}_1$ is successfully decoded and $\mathbb{B}$ denote the event that $\bar{\mathbf{X}}_2$ is successfully decoded *given* that $\bar{\mathbf{X}}_1$ is successfully decoded. Then

$$\begin{aligned}\Pr[\mathbb{A}] &= \Pr\left[C\left(\frac{h\beta_1 P}{h(1-\beta_1)P+N}\right) \ge \alpha R_1\right] \\ &= \Pr\left[h \ge h^1_{\text{thr}}\right],\end{aligned}$$

where

$$h^1_{\text{thr}} = \frac{e^{2\alpha R_1}-1}{\frac{P}{N}[1-(1-\beta_1)e^{2\alpha R_1}]},$$

and N is the noise power as in Section II. Note that $h^1_{\text{thr}} > 0$ requires

$$1-(1-\beta_1)e^{2\alpha R_1} > 0 \Rightarrow R_1 > \frac{1}{2\alpha}\log\frac{1}{1-\beta_1}.$$

We also have

$$\begin{aligned}\Pr[\mathbb{B}] &= \Pr\left[C\left(\frac{h(1-\beta_1)P}{N}\right) \ge \alpha(R-R_1)|\mathbb{A}\right] \\ &= \Pr\left[h \ge h^2_{\text{thr}}|h \ge h^1_{\text{thr}}\right],\end{aligned}$$

where

$$h^2_{\text{thr}} = \frac{e^{2\alpha(R-R_1)}-1}{(1-\beta_1)\frac{P}{N}}.$$

We can assume $h^1_{\text{thr}} \le h^2_{\text{thr}}$ without loss of generality. Given h^1_{thr} and h^2_{thr} the mean distortion is

$$\begin{aligned}E[D] &= \Pr[h > h^2_{\text{thr}}]D(R) \\ &\quad + \Pr[h^1_{\text{thr}} < h < h^2_{\text{thr}}]D(R_1) + \Pr[h < h^1_{\text{thr}}] \\ &= D(R)\exp\left(\frac{e^{2\alpha(R-R_1)}-1}{\gamma(1-\beta_1)}\right) \\ &\quad + D(R_1)\exp\left(\frac{e^{2\alpha R_1}-1}{\gamma[1-(1-\beta_1)e^{2\alpha R_1}]}\right) \\ &\quad - D(R_1)\exp\left(\frac{e^{2\alpha(R-R_1)}-1}{\gamma(1-\beta_1)}\right) \\ &\quad + \left(1-\exp\left(\frac{e^{2\alpha R_1}-1}{\gamma[1-(1-\beta_1)e^{2\alpha R_1}]}\right)\right), \quad (14)\end{aligned}$$

where γ is the channel SNR. In Table I we compare the mean distortion of the scalable coder (14), minimized over β_1, R_1, and R, with the single layer coder (4) for several values of γ and $\alpha = 1$. As γ increases, ${\beta_1}^*$, ${R_1}^*$ and R^* increase, and the reduction in expected distortion achieved by the scalable coder, relative to the single-layer coder (last column in Table I) at optimality, increases with γ.

TABLE I
SCALABLE/SINGLE LAYER CODING RESULTS FOR SEVERAL SIGNAL TO NOISE RATIOS γ

γ [dB]	Two-layer				Single-layer	
	${\beta_1}^*$	${R_1}^*$	R^*	$E[D]^*$ [dB]	R^*	$E[D]^*$ [dB]
-5	0.56	0.05	0.10	-0.41	0.10	-0.41
5	0.75	0.25	0.45	-2.11	0.45	-2.04
15	0.85	0.55	1.05	-5.87	0.90	-5.40
25	0.91	0.85	1.75	-11.11	1.45	-9.79
35	0.96	1.25	2.50	-17.12	2.00	-14.58

The probabilities of correct reception in the two-layer scheme for $\gamma = 15$ dB are 0.70, 0.19, and 0.11 for both layers, one layer and neither layer, respectively. In contrast, for single-layer transmission the probability of successful decoding is 0.86. Hence two-layer scalable coding reduces the probability of correctly receiving the entire message, as well as the probability of a decoding failure, with respect to single-layer coding. This allows a shift in probability mass for the two-layer scheme to the event in which only the first layer is decoded.

V. $Q \ge 2$ LAYERS

Now consider $Q \ge 2$ layers, where the first i layers, $1 \le i \le Q$, are assigned power $\beta_i P$, and give total rate R_i if successfully decoded (i.e., $R_i \ge R_{i-1}$ and $\beta_i \ge \beta_{i-1}$). Let $\Delta R_i \triangleq R_i - R_{i-1}$, which is the incremental rate associated with layer i. The probability that the received distortion is $D(R_i)$ is then the probability of successfully decoding the ith layer, but not the $\{i+1\}$st layer. Since the ith layer has power $\Delta\beta_i \triangleq \beta_i - \beta_{i-1}$, and receives interference power $(1-\beta_i)P$ from layers $i+1$ through Q, the probability of successful decoding is $\pi_i - \pi_{i+1}$, where

$$\begin{aligned}\pi_i &= \Pr\left[C\left(\frac{h\Delta\beta_i P}{h(1-\beta_i)P+N}\right) \ge \alpha\Delta R_i\right] \\ &= \exp\left\{-\frac{1}{\gamma}\frac{e^{2\alpha\Delta R_i}-1}{\Delta\beta_i-(1-\beta_i)\left[e^{2\alpha\Delta R_i}-1\right]}\right\}.\end{aligned}$$

Our problem is to

$$\min_{\beta_1,\ldots,\beta_Q,R_1,\ldots,R_Q} E[D] = \sum_{i=0}^{Q} D(R_i)\cdot(\pi_i - \pi_{i+1}),$$

subject to the boundary conditions

$$\pi_0 = 1,\ \pi_{Q+1} = 0,\ \beta_Q = 1,\ \beta_0 = 0, R_0 = 0,$$

and inequality constraints

$$\beta_{i+1} \geq \beta_i,\ R_{i+1} \geq R_i,\ 1 \geq \pi_i \geq \pi_{i+1} \geq 0,\ i = 1,\ldots,Q.$$

The objective function can be written as

$$E[D] = \sum_{i=0}^{Q} D(R_i)\cdot(\pi_i - \pi_{i+1}) = 1 + \sum_{i=1}^{Q} L(\beta_i, R_i, \Delta\beta_i, \Delta R_i), \tag{15}$$

where

$$L(\beta_i, R_i, \Delta\beta_i, \Delta R_i) = \exp(-2R_i)\,(1-\exp(2\Delta R_i)) \times \\ \times \exp\left\{-\frac{1}{\gamma}\frac{e^{2\alpha\Delta R_i}-1}{\Delta\beta_i - (1-\beta_i)\,[e^{2\alpha\Delta R_i}-1]}\right\}.$$

Necessary conditions for this optimization problem can be derived by applying standard methods from the discrete calculus of variations [33]. However, those conditions do not appear to be straightforward to solve, so that instead we present a numerical solution based on dynamic programming. For the time being, we assume a fixed terminal rate $R_Q = \bar{R}$. Let

$$V_Q(\beta_Q, R_Q) = V_Q(1,\bar{R}) \\ = \min_{\substack{\beta_1,\ldots,\beta_Q \\ R_1,\ldots,R_Q}} \left\{\sum_{i=1}^{Q} L(\beta_i, R_i, \Delta\beta_i, \Delta R_i)\right\},$$

which can be written in recursive form as

$$V_Q(\beta_Q, R_Q) = \min_{\beta_{Q-1},R_{Q-1}} \{L(\beta_Q, R_Q, \Delta\beta_Q, \Delta R_Q) \\ + V_{Q-1}(\beta_{Q-1}, R_{Q-1})\}, \tag{16}$$

which is the dynamic programming Optimality Equation (OE). Using the OE and the boundary condition for $Q = 1$,

$$V_1(\beta_1, R_1) = L(\beta_1, R_1, \beta_1, R_1),$$

the optimization problem can be solved recursively in Q. Inputs to the following algorithm are the number of layers Q, γ, and the terminal rate $\bar{R}$. We quantize the power and rate allocation, and let $\delta\beta$ and δR denote the increments between quantized values of β_i and R_i, respectively. For every pair (β_1, R_1) where $\beta_1 = 0 : \delta\beta : 1$ and $R_1 = 0 : \delta R : \bar{R}$,[3] compute the terminal condition $V_1(\beta, R)$, and keep those values in a table. From the OE we have for the second layer

$$V_2(\beta_2, R_2) = \min_{\beta_1,R_1} \{L(\beta_2, R_2, \Delta\beta_2, \Delta R_2) + V_1(\beta_1, R_1)\}.$$

For every pair (β_2, R_2) where $\beta_2 = 0 : \delta\beta : 1$ and $R_2 = 0 : \delta R : \bar{R}$, minimize the expression $L(\beta_2, R_2, \Delta\beta_2, \Delta R_2) + V_1(\beta_1, R_1)$ over (β_1, R_1). For every (β_2, R_2) we only have to consider pairs (β_1, R_1) such that $\beta_2 \geq \beta_1$ and $R_2 \geq R_1$. This generates a table of values for $V_2(\beta_2, R_2)$. We then increment the layer index and repeat.

[3]The notation $x = 0 : s : b$ means $x = 0, s, 2s, \ldots, b$.

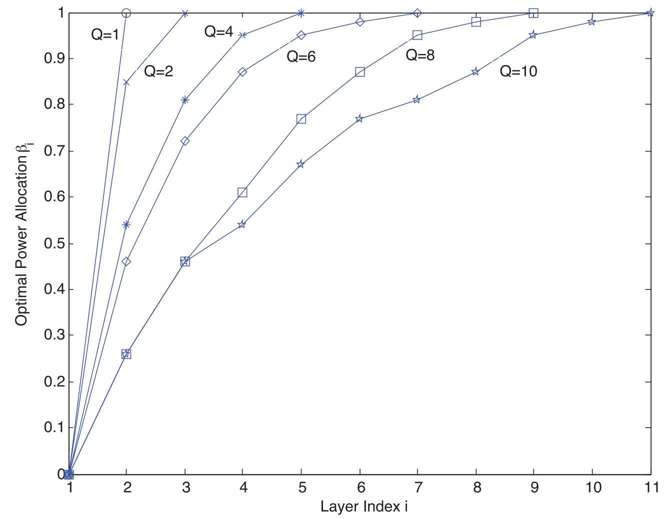


Fig. 4. Optimized power allocations for different (total) number of layers Q ($\gamma = 15dB$, $\alpha = 1$).

Note that for each layer $i = 2,\ldots,Q$, given a particular pair (β_i, R_i), there is a (unique) corresponding pair (β_{i-1}, R_{i-1}) at layer $i-1$, which minimizes the OE. The optimal sequence of powers and rates (β_i, R_i), $i = 1,\ldots,Q$, is then the sequence corresponding to the minimum value $V_Q(\beta_Q, R_Q)$. Using the table entries, which were computed to solve the optimization problem for Q layers, gives the solution for any number of layers from 1 to $Q-1$. However, the optimal sequence (β_i, R_i), $i = 1,\ldots,Q$, for Q layers does not necessarily contain the optimal sequence for $q < Q$ layers. This is because of the terminal constraint on the total power, $\beta_Q = 1$. We also note that the optimized terminal rate may exceed the initial choice of $\bar{R}$. In that case, the algorithm should be rerun with larger $\bar{R}$ so that the optimized terminal rate $R_Q < \bar{R}$.

Let $M = [1/\delta\beta]$ and $N = [\bar{R}/\delta R]$ ($[\cdot]$ is the integer part) be the number of quantized values of β_i and R_i ($i = 1,\ldots,Q$), respectively. Determining the power and rate allocations that minimize the expected distortion with Q-layer scalable coding by exhaustive search has complexity $O(M^Q N^Q)$, whereas the dynamic programming algorithm has complexity $O(MNQ/2)$.

Power and rate allocations given by the preceding algorithm for $\gamma = 15$ dB and $\alpha = 1$ are shown in Figs. 4 and 5. The optimal terminal rate converges to a finite value as the number of layers increases. Fig. 6 shows that the minimum expected distortion decreases and tends to a finite limit as the number of layers increases.

VI. Infinite Number of Layers

To obtain additional insight, we consider the limiting performance in which the number of layers tends to infinity. Let $t = i/Q$ denote the normalized layer index. As $Q \to \infty$, t becomes a continous variable in $(0, 1)$, and we assume that the set of fractional power allocations $\{\beta_i\}$ converges uniformly to the function $\beta(t)$, and the set of rates $\{R_i\}$ converges uniformly to the function $r(t)$, where $\beta(t)$ and $r(t)$ are twice-differentiable. This assumption is illustrated in Fig. 7, which shows plots of optimized power and rate allocations across

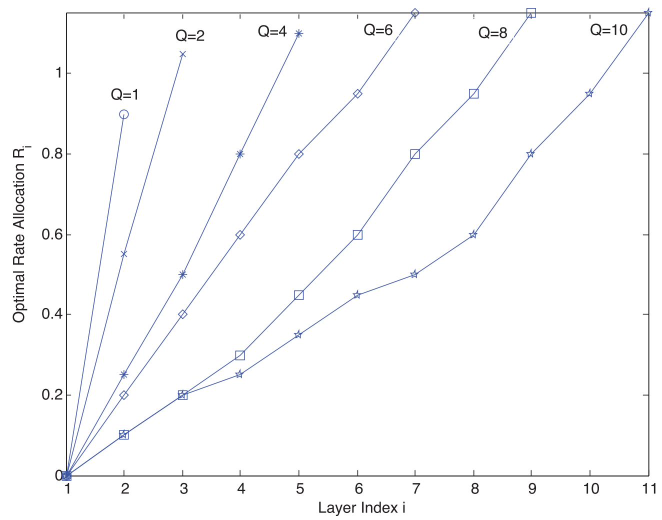


Fig. 5. Optimized rate allocations for different (total) number of layers Q ($\gamma = 15dB$, $\alpha = 1$).

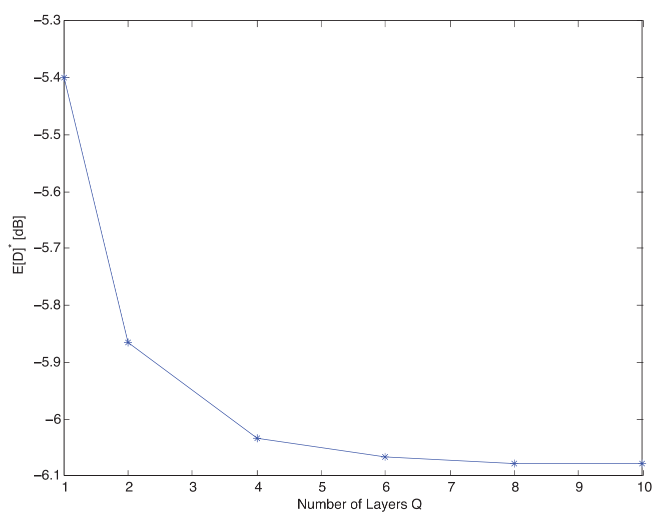


Fig. 6. Minimum expected distortion vs (total) number of layers Q ($\gamma = 15dB$, $\alpha = 1$).

layers for different numbers of layers Q with $\alpha = 1$ and γ=15 dB. It is then straightforward to show that the expected distortion in (15) converges to

$$E[D] = 1 - \int_0^1 \hat{L}(\beta(t), r(t), \beta'(t), r'(t))dt, \tag{17}$$

where

$$\hat{L}(\beta(t), r(t), \beta'(t), r'(t)) = 2r'(t)\exp[-2r(t)] \times \\ \times \exp\left\{-\frac{1}{\gamma}\frac{1}{\beta'(t)\frac{1}{2\alpha r'(t)} - [1-\beta(t)]}\right\}, \tag{18}$$

and prime denotes derivative.

We seek to minimize the preceding expression with respect to $\beta(t), r(t), \beta'(t)$, and $r'(t)$. The functional $\hat{L}(\beta, r, \beta', r')$ does not depend on the layer index t, so that our problem is *autonomous* [34]. Consequently, the Euler-Lagrange necessary

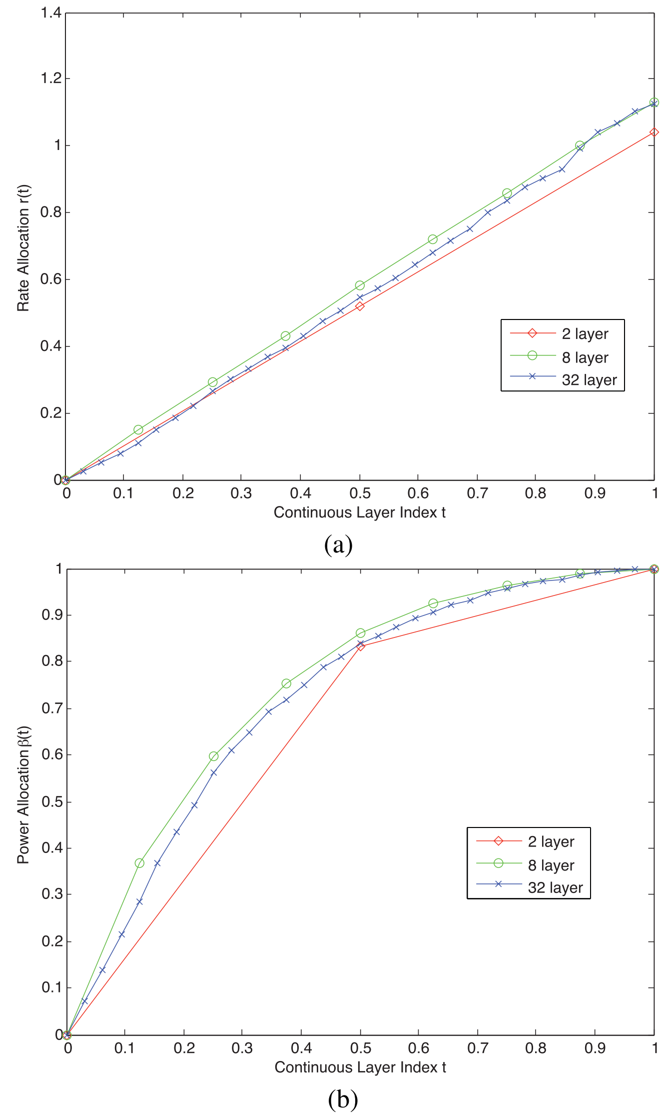


Fig. 7. Optimized rate and power allocations across layers for different number of layers Q with $\alpha = 1$, $\gamma = 15dB$.

conditions for optimality are [34]

$$\hat{L}_\beta = \hat{L}_{\beta'\beta}\beta' + \hat{L}_{\beta'\beta'}\beta'', \tag{19}$$

$$\hat{L}_r = \hat{L}_{r'r}r' + \hat{L}_{r'r'}r'', \tag{20}$$

where $\hat{L}_x = \partial\hat{L}/\partial x$, $\hat{L}_{xy} = \partial^2\hat{L}/\partial x\partial y$, and we have the following transversality condition since $r(1)$ is free

$$\hat{L}_{r'}(\beta(1), r(1), \beta'(1), r(1)) = 0. \tag{21}$$

Solving these equations appears to be difficult. However, we can obtain a closed-form solution for $\alpha = 1$ in the high SNR regime. Namely, in that case the necessary conditions can be simplified, and we derive the approximate solution in the appendix,

$$\beta(t) = d\exp\left[-\frac{3(at+b)^{2/3}}{a}\right]\left[3(at+b)^{2/3} + a\right] \\ +1, \tag{22}$$

$$r(t) = \frac{3}{2a}(at+b)^{2/3} + c, \tag{23}$$

where

$$c = \frac{1+\mathrm{W}(\gamma)}{2}, \tag{24}$$

$$b = \left[\frac{1}{3}\left(2c - \frac{1}{\sqrt{(2c)}}\right)\right]^{-3}, \tag{25}$$

$$a = -\frac{1.5b^{2/3}}{c}, \tag{26}$$

$$d = \frac{2}{\gamma e a}, \tag{27}$$

and again $\mathrm{W}(\cdot)$ denotes Lambert's W-Function (*cf* Section III). The corresponding minimum expected distortion

$$E[D]^* \doteq \exp\left(-\mathrm{W}(\gamma/2)\right)\left(1+\frac{1}{2}\mathrm{W}(\gamma/2)\right), \tag{28}$$

where $\doteq$ denotes asymptotic equality for large γ, and is defined in the appendix. Replacing $D(R_i)$ by $e^{sD(R_i)}$ in (15), and letting $Q \to \infty$ gives the moment generating function for D,

$$\begin{aligned}\Phi(s) &= E[e^{sD(r(t))}] \\ &= e^s - \int_0^1 s e^{se^{-2r(t)}} 2r'(t) e^{2r(t)} dt \\ &\quad - \int_0^1 \frac{1}{\gamma} \frac{s e^{se^{-2r(t)}} 2r'(t) e^{2r(t)}}{\beta'(t)\frac{1}{2r'(t)} - (1-\beta(t))} dt.\end{aligned}$$

Substituting the large SNR expressions for $\beta(t)$, $r(t)$ gives

$$\Phi(s) \doteq \exp\left(\frac{s}{\exp(2c-1)}\right) + s\frac{\mathrm{E}_1\left(-\frac{s}{e^{\mathrm{W}(\gamma/2)}}\right) - \mathrm{E}_1(-s)}{2e^{\mathrm{W}(\gamma/2)}},$$

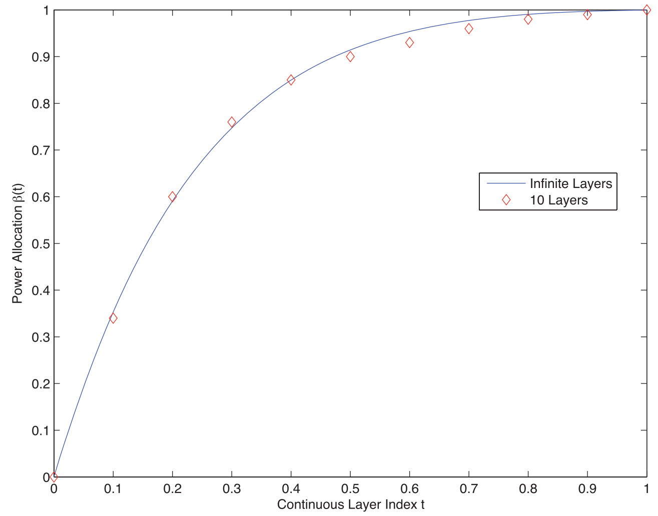
from which we can compute

$$E[D^2]^* = \Phi''(0) = e^{-\mathrm{W}(\gamma/2)}.$$

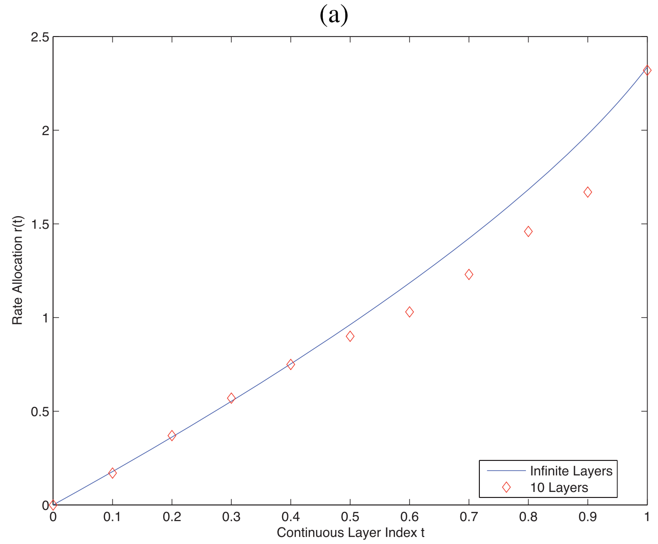
Also of interest is the maximum achievable rate, assuming all layers are successfully decoded, given by

$$r(1) = \bar{R} = \frac{1}{2}\mathrm{W}(\gamma/2).$$

Fig. 8 shows plots of the power and rate allocations $\beta(t)$ and $r(t)$ for $\gamma = 30$ dB and $\alpha = 1$. Also shown are the corresponding power and rate allocations with 10 layers obtained by solving the discrete optimization numerically. Those results are nearly the same as the infinite-layer results. The expected distortion for this example is approximately 15 dB (with both 10 and infinite layers).



(a)



(b)

Fig. 8. High SNR ($\gamma = 30dB$) power and rate allocations for infinite-layer scalable coding. Results with $Q = 10$ layers are also shown.

VII. Comparison

We now illustrate the preceding results by comparing the performance of erasure and scalable codes for specific cases. Fig. 9 shows the minimum (expected) distortion with $\alpha = 1$ versus SNR γ for infinite-length erasure (11) and scalable codes. For $\gamma > 25$ dB, the analytical results for an infinite-layer scalable code (given by (28)) are shown, and for $\gamma < 25$ dB, numerical results for 10 layers are shown. Also shown for comparison are the analogous curves for an AWGN channel (7), codes which achieve ergodic capacity (8), and a single-layer channel code with no additional outer code (6).

These results show that at high SNRs, the (deterministic) distortion for erasure coding is approximately 7 dB less than the expected distortion for scalable coding. Of course, the erasure codes also incur infinite decoding delay. The minimum distortion for the AWGN channel is similar to that shown for the fading channel with codes that achieve the ergodic channel. In both cases, the distortion is much less than that for the block fading channel with erasure codes, and the gap (in dB) increases with SNR (although as discussed in Section III, the distortion exponents are the same). Similarly, the gap between single-layer coding and infinite-layer scalable coding increases substantially at high SNRs.

Fig. 10 shows the standard deviation of the minimum distortion versus SNR for single- and infinite-layer scalable coding ($\alpha = 1$). At low SNRs the standard deviation is small since the received distortion is high with high probability. As the SNR increases, the standard deviation increases until it reaches a maximum, then decreases at high SNRs. Also note that the standard deviation associated with infinite-layers decreases faster at large SNRs than with a single layer.

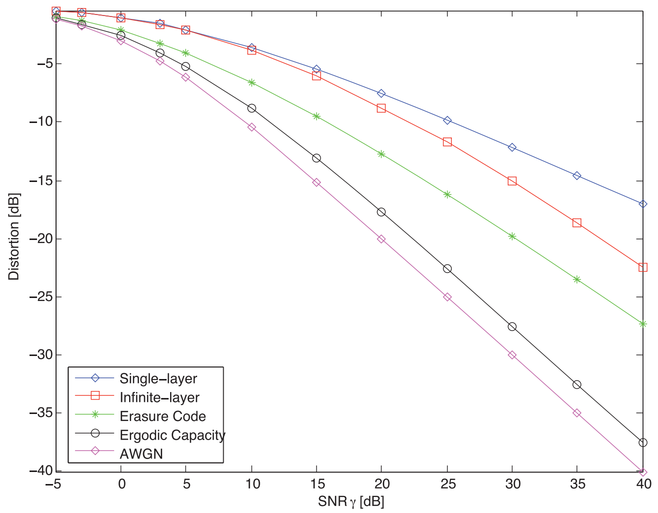


Fig. 9. Comparison of distortion versus SNR with five different coding schemes.

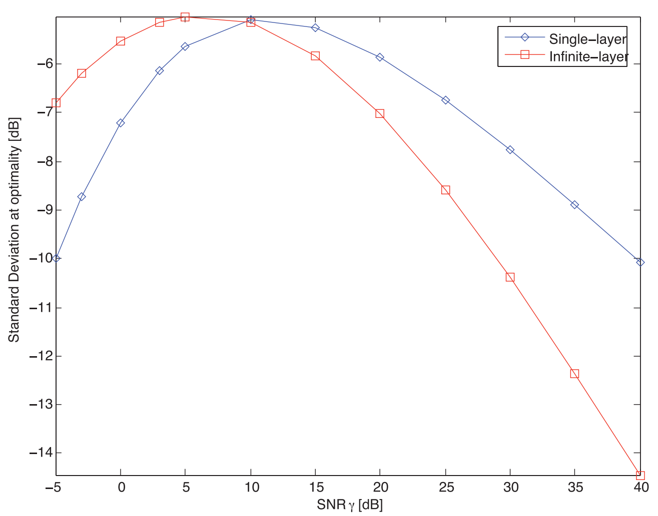


Fig. 10. Standard deviation of the minimum distortion versus SNR with single- and infinite-layer scalable coding.

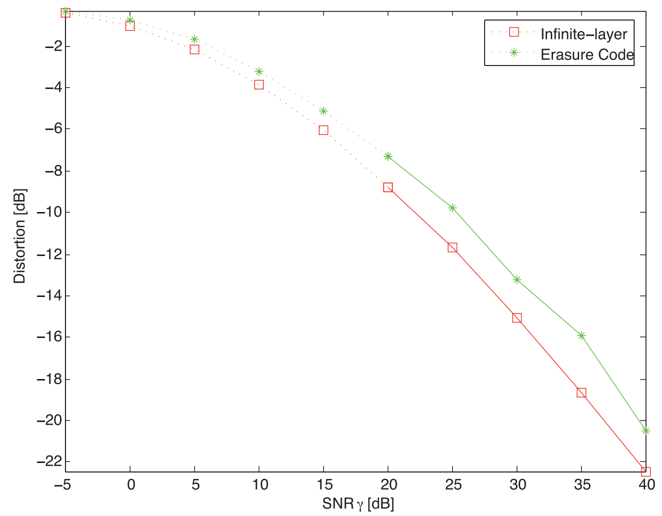


Fig. 11. Expected distortion versus SNR for erasure codes and infinite-layer scalable codes with fixed standard deviation. The dotted lines indicate that the standard deviation for the erasure code is less than that with scalable coding.

Fig. 11 compares the received distortion for scalable coding with infinite layers with the distortion corresponding to *finite-length* erasure codes (as opposed to the infinite-length performance shown in Fig. 9) for the case where $\alpha = 1$. Namely, the length of the erasure code is chosen so that the standard deviation of the distortion is the same as that for the scalable code. (The curve for the scalable code is the same as that shown in Fig. 9.) Here the erasure code performs slightly worse than the scalable code. As the SNR decreases, the optimal rate U/V for the erasure code increases. Hence, the *minimum* number of transmitted codewords V needed to generate the results in Fig. 10 increases and the corresponding standard deviation decreases. Because of this, for $\gamma < 15$ dB the standard deviation for the erasure code is lower than that for the scalable code. Still, the mean distortion for scalable coding is less than that for erasure coding in this region.

We now compare the distortion exponents for the coding schemes considered. The distortion exponent for scalable coding with Q layers is given in [13], namely, $\Delta = \sum_{i=1}^{Q} \frac{1}{\alpha^i}/(1+\sum_{i=1}^{Q} \frac{1}{\alpha^i})$. Letting $Q \to \infty$ gives $\Delta = 1/\alpha$, which is consistent with the large SNR result (28) with $\alpha = 1$. This exponent is shown in [13] to be optimal, and is consistent with the discussion in Section III, which states that the same optimal exponent is achieved with erasure coding. For single-layer transmission ($Q = 1$) the preceding expression gives $\Delta = 1/2$, which is consistent with (6). The numerical results in Figure 9 illustrate that the finite-SNR performance of coding schemes with the same distortion exponents can be significantly different.

VIII. Conclusions

We have considered the use of erasure and scalable codes for the transmission of a continuous source through a block fading channel. For both coding schemes the mean and standard deviation of the distortion has been computed, where the average is over the Rayleigh channel distribution. (For the erasure coding scheme, this average applies to finite-length codewords.) A numerical comparison has shown that both coding schemes offer a substantial performance improvement relative to a basic single-layer coding scheme. Both infinite-length erasure and infinite-layer scalable coding have optimal distortion exponents, although finite-SNR performance can be significantly different. For a fixed distortion variance, both coding schemes give similar minimum expected distortion. Scalable coding appears to be more complex than erasure coding, but it incurs a delay of only a single block.

Although we have considered a Gaussian source to simplify the analysis, it is known that the distortion of such a source is an upper bound on the distortion of any continuous source with the same second moment. Hence, our results can serve as a benchmark for the application of erasure and/or scalable codes to any continuous source. We also remark that if instead of a flat Rayleigh fading channel we have a frequency-selective channel, then the ergodic performance corresponding to infinite-length codewords can be obtained with less delay by coding across frequencies. If spatial dimensions are added to the model (i.e., the channel is MIMO), then another possibility, which remains to be studied, is to combine scalable codes with diversity schemes such as multiple description coding [35] and the hybrid digital-analog schemes in [25]. Finally, in a multi-user environment scalable codes offer a natural way of transmitting information with different Quality-of-Service requirements, whereas erasure codes do not directly address different types of service requests.

IX. Acknowledgements

The authors thank J. Nicholas Laneman for suggesting the use of scalable coding in fading channels, Yiftach Eisenberg for helpful discussions concerning this work, and the reviewers for their helpful comments.

Appendix

We approximate a function $f(x,\gamma)$ by a function $g(x,\gamma)$ for *sufficiently high SNR* γ, and write

$$f(x,\gamma) \doteq g(x,\gamma)$$

to denote that

$$\lim_{\gamma\to\infty} \sup_{x\in\mathcal{A}\setminus\{x:g(x,\gamma)=0\}} \left|\frac{f(x,\gamma)}{g(x,\gamma)}\right| = 1, \tag{29}$$

and

$$\lim_{\gamma\to\infty} \sup_{x\in\mathcal{A}} |f(x,\gamma) - g(x,\gamma)| = 0, \tag{30}$$

where $\mathcal{A}$ is a compact subset of the positive reals. As an example, $f(x,\gamma)$ could be the distortion measure, where x is the source rate. In that case, the second condition is clearly satisfied for any f and g, which tend to zero as $\gamma \to \infty$. The first condition then guarantees that f and g converge to zero at the same rate.

A. Single Layer

We first illustrate the high-SNR approximation by considering single-layer transmission, discussed in Section II. In that case, the optimized rate and associated distortion are given by (5) and (6), respectively, for $\alpha = 1$. The high-SNR approximations are obtained by approximating the outage probability

$$\begin{aligned}\Pr\left[C\left(\frac{hP}{N}\right) < \alpha R\right] &= 1 - \exp\left\{-\frac{e^{2\alpha R}-1}{\gamma}\right\} \\ &\doteq \frac{e^{2\alpha R}-1}{\gamma}.\end{aligned}$$

For this approximation to hold in the sup norm we must have $e^{2\alpha R} - 1 < \gamma$, or $R < R^*_{\text{AWGN}} = \log(1+\gamma)/(2\alpha)$, where R^*_{AWGN} is the optimized source rate in the absence of fading. This constraint is clearly satisfied, since fading lowers the optimized rate. For purposes of defining the high-SNR approximation, we can therefore take $\mathcal{A} = [0, (1-\epsilon)R^*_{\text{AWGN}}]$ for some sufficiently small, positive ϵ.

With this approximation the corresponding distortion

$$E[D]_{\text{approx}} = e^{-2R} \cdot \left(1 - \frac{e^{2\alpha R}-1}{\gamma}\right) + \frac{e^{2\alpha R}-1}{\gamma}.$$

Optimizing over R gives for $\alpha = 1$

$$R^*_{\text{approx}} = \frac{1}{4}\log(1+\gamma),$$

and

$$E[D]^*_{\text{approx}} = \frac{1}{\sqrt{1+\gamma}}\left(1 - \frac{\sqrt{1+\gamma}-1}{\gamma}\right) + \frac{\sqrt{1+\gamma}-1}{\gamma}.$$

It is easily verified that for $\alpha = 1$

$$E[D(R^*)] = \min_{R\in\mathcal{A}} E[D] \doteq \min_{R\in\mathcal{A}} E[D]_{\text{approx}} = E[D]^*_{\text{approx}},$$

and furthermore,

$$R^* = \arg\min_{R\in\mathcal{A}} E[D] \doteq \arg\min_{R\in\mathcal{A}} E[D]_{\text{approx}} = R^*_{\text{approx}}.$$

B. Infinite Layers

In this case the expected distortion is given by the functional (17)-(18), where the third term of the product in (18) is the probability of successful decoding up to layer t. We again use the high-SNR approximation $\exp(-x/\gamma) \doteq 1 - x/\gamma$, where x satisfies $(1/2\alpha)\log x \in \mathcal{A}$, to obtain for $\alpha = 1$

$$\begin{aligned}E[D] \;&\doteq\; 1 - \int_0^1 2r'(t)e^{-2r(t)}dt \\ &\quad + \int_0^1 2r'(t)e^{-2r(t)}\frac{1}{\gamma}\frac{1}{\beta'(t)\frac{1}{2\alpha r'(t)} - (1-\beta(t))}dt.\end{aligned}$$

The corresponding Euler-Lagrange equations are again given by (19)-(21), and substituting the high-SNR approximation for the integrand gives, after some algebra, the following system of ordinary differential equations

$$\begin{aligned}3\beta'(t)\rho'(t) + (\beta(t)-1)2(\rho'(t))^2 + \beta''(t) &= 0,\\ (\rho'(t))^2\beta'(t) + (\beta(t)-1)2(\rho'(t))^3 - \alpha\beta'(t)\rho''(t) &= 0,\end{aligned}$$

with initial and terminal conditions

$$\beta(0) = 0,\ \rho(0) = 0,\ \beta(1) = 1,\ 4\rho'(1) = \gamma\beta'(1),$$

where $\rho(t) = \alpha r(t)$, for all t. This system of equations can be equivalently written as a system of first-order equations

$$\begin{aligned}\rho' = w, &\qquad \beta' = x,\\ x' = (1-\beta)2w^2 - 3\beta' w, &\qquad w' = \frac{w^2\beta' + (\beta-1)2w^3}{\alpha\beta'},\end{aligned}$$

with the same initial and terminal conditions. This system yields a closed form solution only for $\alpha = 1$ which is given by (22)-(27).

Since the distortion measure is a smooth, bounded functional of γ and the power and rate allocations across layers within the constraint set $\mathcal{A}$, we expect that the approximate high-SNR solution has the same properties as those observed for the single-layer case. Namely, the associated minimum distortion is a high-SNR approximation of the actual minimum distortion according to the definition (29)-(30). Also, the corresponding power and rate allocations should converge in distribution to the optimal distributions.

REFERENCES

[1] H. Coward, R. Knopp, and S. Servetto, "On the performane of a natural class of joint source/channel codes based upon multiple descriptions," *IEEE Trans. Inform. Theory*, submitted for publication, 2002.

[2] Z. Reznic, R. Zamir, and M. Feder, "Joint source-channel coding for a gaussian mixture source over the gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 776–781, Mar. 2002.

[3] N. Gogate, D. Chung, S. S. Panwar, and Y. Wang, "Supporting image and video applications in a multihop radio environment using path diversity and multiple description coding," *IEEE Trans. Circuits Systems for Video Technol.*, vol. 12, no. 9, pp. 777–791, Sept. 2002.

[4] J. N. Laneman, E. Martinian, G. W. Wornell, and J. G. Apostolopoulos, "Source-channel diversity approaches for multimedia communication," *IEEE Trans. Inform. Theory*, accepted for publication, 2004.

[5] T. Holliday and A. Goldsmith, "Joint source and channel coding for mimo systems," in *Proc. Allerton Conference on Communications, Control, and Computing*, Oct. 2004.

[6] Q. Zhao, P. Cosman, and L. B. Milstein, "Optimal allocation of bandwidth for source coding, channel coding, and spreading in cdma systems," *IEEE Trans. Commun.*, vol. 52, no. 10, pp. 1797–1808, Oct. 2004.

[7] Y. Eisenberg, C. E. Luna, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Joint source coding and transmission power management for energy efficient wireless video communications," *IEEE Trans. Circuits Systems for Video Technol.*, vol. 12, no. 6, pp. 411–424, June 2002.

[8] V. K. Goyal and J. Kovacevic, "Generalized multiple description coding with correlated transforms," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2199–2224, Sept. 2001.

[9] J. G. Apostolopoulos and M. D. Trott, "Path diversity for enchanced media streaming," *IEEE Commun. Mag.*, vol. 42, no. 8, pp. 80–94, Aug. 2004.

[10] B. P. Dunn and J. N. Laneman, "Characterizing source-channel diversity approaches beyond the distortion exponent." in *Proc. Allerton Conference Communication, Control, and Computing*, Monticello, IL, Sept. 2005.

[11] S. Shamai and A. Steiner, "A broadcast approach for a single-user slowly fading MIMO channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2617–2635, Oct. 2003.

[12] K. Lau, O. Takeshita, and M. Fitz, "Optimal rate allocation for superposition coding in quasi-static fading channels," in *Proc. IEEE ISIT*, Lausanne, Switzerland, June 2002.

[13] D. Gunduz and E. Erkip, "Source and channel coding for quasi-static fading channels," in *Proc. Asilomar*, Pacific Grove, CA, Nov. 2005.

[14] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code or not to code," in *Proc. IEEE ISIT*, Sorrento, Italy, June 2000.

[15] L. Ozarow, S. Shamai, and A. D. Wyner, "Infomation theoretic considerations for cellural mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, no. 2, pp. 359–378, May 1994.

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley and Sons: Wiley Series in Telecommunications, 1991.

[17] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inform. Theory*, vol. 37, no. 2, pp. 269–275, Mar. 1991.

[18] S. Shamai, "A broacast strategy for the Gaussian slowly fading channel," in *Proc. IEEE ISIT*, Ulm, Germany, June 1997.

[19] D. Gunduz and E. Erkip, "Source and channel coding for cooperative relaying," in *Proc. SPAWC*, New York, NY, June 2005.

[20] F. Etemadi and H. Jafarkhani, "Optimal layered transmission over quasi-static fading channels," in *Proc. IEEE ISIT*, Seattle, WA, July 2006.

[21] G. Caire and K. Narayanan, "On the SNR exponent of hybrid digital-analog space time coding," in *Proc. Allerton Conference*, Allerton, IL, Sept. 2005.

[22] D. Gunduz and E. Erkip, "Distortion exponent of MIMO fading channels," in *Proc. IEEE Information Theory Workshop*, Punta del Este, Uruguay, Mar. 2006.

[23] M. Skoglund, N. Phamdo, and F. Alajaji, "Design and performance of VQ-baed hybrid digital-analog joint source-channel codes," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 708–720, Mar. 2002.

[24] S. Sessia, G. Caire, and G. Vivier, "Lossy transmission over slow-fading AWGN channels: a comparison of progressive and superposition and hybrid approaches," in *Proc. IEEE ISIT*, Adelaide, Australia, Sept. 2005.

[25] M. Skoglund, N. Phamdo, and F. Alajaji, "Hybrid digital-analog source-channel coding for bandwidth compression/expansion," *IEEE Trans. Inform. Theory*, vol. 52, no. 8, pp. 3757–3763, Aug. 2006.

[26] J. Lu, A. Nosratinia, and B. Aazhang, "Progressive source-channel coding of images over bursty error channels," in *Proc. IEEE Int. Conference on Image Processing*, Chicago, IL, Oct. 1998.

[27] G. Sherwood and K. Zeger, "Error protection for progressive image transmission over memoryless and fading channels," *IEEE Trans. Commun.*, vol. 46, no. 12, pp. 1555–1559, Dec. 1998.

[28] A. Nosratinia, J. Lu, and B. Aazhang, "Source-channel rate allocation for progressive transmission of images," *IEEE Trans. Commun.*, vol. 51, no. 2, pp. 186–196, Feb. 2003.

[29] W. Li, "Overview of fine granularity scalability in mpeg-4 video standard," *IEEE Trans. Circuits Systems for Video Technol.*, vol. 11, no. 3, pp. 301–317, Mar. 2001.

[30] F. Wu, S. Li, and Y.-Q. Zhang, "A framework for efficient progressive fine granularity scalable video goding," *IEEE Trans. Circuits Systems for Video Technol.*, vol. 11, no. 3, pp. 332–344, Mar. 2001.

[31] S. Lin and D. J. Costello, *Error Control Coding: Fundamentals and Applications*. Prentice-Hall, 1982.

[32] T. M. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 2–14, Jan. 1972.

[33] G. Guseinov, "Discrete calculus of variations," in *Proc. AIP Conference Global Analysis and Applied Mathematics: International Workshop on Global Analysis*, vol. 729, Oct. 2004, pp. 170–176.

[34] M. I. Kamien and N. L. Schwartz, *Dynamic Optimization. The Calculus of Variations and Optimal Control in Economics and Management*. North-Holland (Advanced Textbooks in Economics), 1991.

[35] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Processing Mag.*, vol. 18, pp. 74–93, Sept. 2001.

Konstantinos E. Zachariadis is a Ph.D. candidate of Managerial Economics and Strategy in Northwestern University's Kellogg School of Management, and a Fulbright Scholar. His research focuses on the mechanisms behind financial markets and in particular on how traders' information and traders' strategies can affect the price of an asset. Konstantinos started his academic career as an Electrical and Computer Engineer. He holds an MSc in Electrical and Computer Engineering from Northwestern University and a Diploma (MSc equivalent) from the Aristotle University of Thessaloniki in Greece. His research there focused on the tradeoffs associated with the transmission of multimedia in cellular networks.

For the whole duration of his graduate studies Konstantinos was awarded a Fulbright Fellowship and a Northwestern Graduate School Scholarship. In addition, for his MSc he received the Institute of Electrical and Electronic Engineers (IEEE) Life Members Graduate Study Fellowship, and for his PhD the Kellogg School of Management Scholarship, and the Gerondelis Foundation Inc. Scholarship. Konstantinos has worked on research projects funded by the European Union (EU), the National Science Foundation (NSF) and Motorola. He also interned for a summer at Hewlett Packard's Information Technology (IT) consulting branch in Tokyo, Japan. During his studies he has given many talks and attended numerous conferences. He is the author/co-author of several papers in economic theory and engineering, some of which have already been published in relevant journals.

Apart from his studies he has also been intensively involved in school and university student bodies. He re-established and was the chairperson of the student branch of IEEE at the Aristotle University of Thessaloniki, which grew to be the biggest in Greece. As a chairperson he organized many workshops and talks and helped the branch win global IEEE awards. He also served in the corresponding position at Northwestern University. He is a member of IEEE, the Institute of Operations Research and Management Science (INFORMS), and the Technical Chamber of Greece.

Michael L. Honig (S'80-M'81-SM'92-F'97) received a B.S. degree in electrical engineering from Stanford University in 1977, and M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1978 and 1981, respectively. He subsequently joined Bell Laboratories in Holmdel, NJ, where he worked on local area networks and voiceband data transmission. In 1983 he joined the Systems Principles Research Division at Bellcore, where he worked on Digital Subscriber Lines and wireless communications. Since the fall of 1994, he has been with Northwestern University where he is a Professor in the Electrical Engineering and Computer Science Department. He has held visiting positions at Princeton University, the University of Sydney, and the University of California, Berkeley. He has also worked as a free-lance trombonist.

He has served as an editor for the *IEEE Transactions on Information Theory* and the *IEEE Transactions on Communications*. He has also served as a member of the digital signal processing technical committee for the IEEE Signal Processing Society, and as a member of the Board of Governors for the Information Theory Society. He is the recipient of a Humboldt research award for senior U.S. scientists, and the co-recipient of the 2002 IEEE Communications Society and Information Theory Society Joint Paper Award.

Aggelos K. Katsaggelos (S'80-M'85-SM'92-F'98) received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of Thessaloniki, Greece, in 1979 and the M.S. and Ph.D. degrees both in electrical engineering from the Georgia Institute of Technology, in 1981 and 1985, respectively. In 1985 he joined the Department of Electrical and Computer Engineering at Northwestern University, where he is currently professor. He was the holder of the Ameritech Chair of Information Technology (1997-2003). He is also the Director of the Motorola Center for Seamless Communications (since 1998) and a member of the Academic Affiliate Staff, Department of Medicine, at Evanston Hospital (since 1989).

Dr. Katsaggelos is a member of the Publication Board of the IEEE Proceedings, the IEEE Technical Committees on Visual Signal Processing and Communications, and Multimedia Signal Processing, the Editorial Board of Academic Press, Marcel Dekker: Signal Processing Series, Applied Signal Processing, and Computer Journal. He has served as editor-in-chief of the IEEE Signal Processing Magazine (1997-2002), a member of the Publication Boards of the IEEE Signal Processing Society, the IEEE TAB Magazine Committee, an Associate editor for the IEEE Transactions on Signal Processing (1990-1992), an area editor for the journal Graphical Models and Image Processing (1992-1995), a member of the Steering Committees of the *IEEE Transactions on Image Processing* (1992-1997) and the *IEEE Transactions on Medical Imaging* (1990-1999), a member of the IEEE Technical Committee on Image and Multi-Dimensional Signal Processing (1992-1998), and a member of the Board of Governors of the IEEE Signal Processing Society (1999-2001). He has published extensively in the areas of signal processing, multimedia transmission, and computer vision. He is the editor of *Digital Image Restoration* (Springer-Verlag 1991), co-author of *Rate-Distortion Based Video Compression* (Kluwer 1997), co-editor of *Recovery Techniques for Image and Video Compression and Transmission* (Kluwer 1998), and co-author of *Super-Resolution for Images and Video* (Morgan & Claypool, 2007) and *Joint Source-Channel Video Transmission* (Morgan & Claypool, 2007). He is the co-inventor of twelve international patents, a Fellow of the IEEE, and the recipient of the IEEE Third Millennium Medal (2000), the IEEE Signal Processing Society Meritorious Service Award (2001), an IEEE Signal Processing Society Best Paper Award (2001), an IEEE International Conference on Multimedia and Expo Paper Award (2006), and an IEEE International Conference on Image Processing Paper Award (2007). He is a Distinguished Lecturer of the IEEE Signal Processing Society (2007-08).