

PERFORMANCE OF MULTIPLE DESCRIPTION CODERS ON A REAL CHANNEL

Amy R. Reibman, Hamid Jafarkhani
AT&T Labs – Research
{amy,hamid}@research.att.com

Michael T. Orchard * Yao Wang
Princeton Univ. Polytechnic Univ.
orchard@princeton.edu yao@vision.poly.edu

ABSTRACT

In this paper we explore the ability of multiple description (MD) source coders to achieve good performance on channels other than ideal MD channels. We examine both the overall system design and compare the performance of a system with MD source coder to that of a more traditional system using a layered source coder.

For the memoryless channels we consider, MD source coding cannot achieve acceptable performance for a memoryless Gaussian source without appropriate channel coding. Also, in memoryless channels, a system with MD source coding outperforms a layered source coding system only in very poor channels. The introduction of memory in the channel degrades the performance of both systems equally. Using interleaving to reduce the impact of memory in the channel has more influence on performance than the choice of source coder.

1. INTRODUCTION

There has been extensive work in the information theory community, including [1], on finding achievable regions for multiple descriptions, and a recent revival in the signal processing community on building compression systems that can approach these bounds [2], [3], [4], [5], [6], [7], [8]. Also, new tools (the redundancy-rate distortion (RRD) curve) have been presented [5] that allow more flexible analysis and design of MD systems. However, these systems have all been analyzed under the assumption of ideal multiple description (MD) channels, in which 2 identical channels may undergo erasures independently.

In this paper we explore the ability of source coders developed for these ideal MD channels to achieve good performance on more realistic channels. Our goals are first, to examine the overall system design to support an MD source coder on a real channel, and second, to compare the performance of such a system using an MD source coder to that of a more traditional system using a layered source coder.

We begin in Section 2 by examining the ideal environment for which MD coding was developed: two independent channels which may cause complete channel erasures, independently of each other. We explore the degree to which MD coding outperforms more traditional methods like layered coding. This ideal MD environment occurs occasionally in real-life, for example in a military environment in which there are multiple paths, all hostile.

However, most actual physical channels do not conform to the requirements of ideal MD channels, in two possible

ways. First, the data destined for the two idealized MD channels may not be sent on two physically independent channels, either because the two channels do not fail independently, or because there is only a single physical channel. In this paper, we consider only the latter: the case of a single physical channel. Second, erasures may not be pure erasures but instead be caused by the inability of forward error correction (FEC) to correct the bit errors in a block. Thus, a decoder might be designed to extract useful information from partially corrupted data [9].

In what follows, we explore the encoder system design required to support an MD source coder on both a memoryless channel (Section 3) and a channel with memory (Section 4). As a baseline for comparison, we use a similar system using a layered source coder with unequal error protection, designed to be optimal for the average channel conditions. Neither of these systems are able to take into account any knowledge at the encoder of the instantaneous channel conditions, nor do they consider decoding partially corrupt data.

2. IDEAL MD CHANNELS

Figure 1 shows the channel environment for which Multiple Descriptions was designed. There are two independent identical channels and the challenge is to decompose the source into two related descriptions such that if either description is lost a reasonable reconstruction is still possible. The source coder knows of the existence of the two channels with independent losses but does not know which of the channels are currently operational. The decoder, however, knows this information.

The RRD function [7] for a single memoryless Gaussian variable with variance σ^2 can be found from Ozarow's bound [1] to be

$$D_1(\rho; D_0^*) \geq \frac{1}{2} \left(\sigma^2(1 - \sqrt{1 - 2^{-2\rho}}) + D_0^*(1 + \sqrt{1 - 2^{-2\rho}}) \right). \quad (1)$$

ρ is the redundancy, or additional rate, beyond that necessary to achieve a two-channel distortion of D_0^* , required to reduce the one-channel distortion, D_1 . Note that D_1 has slope $-\infty$ at $\rho = 0$, decays faster than exponentially for small ρ , but decays exponentially for large redundancies.

We compare the performance of a generic MD coder in Figure 1 to that of a more traditional non-MD encoder adapted for use on ideal MD channels, shown in Figure 2. In this system, we transmit on both MD channels a coarse, high priority (HP), encoding with distortion

$$D_H = \sigma^2 2^{-2R_H}, \quad (2)$$

This work is conducted in AT&T Labs-Research, where the last two authors are consultants.

and split a finer, low priority (LP) encoding with overall distortion

$$D_{HL} = \sigma^2 2^{-2(R_H + R_L)} \quad (3)$$

between the two channels. This is possible because the rate-distortion function of a memoryless Gaussian source is successively refinable. This system applies redundancy explicitly to the most significant bits, and is a simple alternative to using an MD coder. The total rate sent on both channels is $2R_H + R_L$ with redundancy $\rho = R_H$. The one-channel distortion is $D_1 = (D_H + D_{HL})/2$, or,

$$D_1(\rho; D_0^*) = (\sigma^2 2^{-2\rho} + D_0^*)/2, \quad (4)$$

which decays exponentially in redundancy.

The super-exponential decay in (1) provides a clear advantage to be gained by MD coding at small ρ on ideal MD channels. However, for large ρ , MD coding can at best perform a factor of two, or one bit better than a simple alternative.

3. ONE MEMORYLESS CHANNEL

First, we examine the scenario where there is one memoryless channel and the MD coder must be adapted to send its information across the single channel, as shown in Figure 3. For comparison, we consider a layered coding system (see Figure 4) using unequal error protection that is designed to be optimal for the given channel.

We examine two models for a memoryless channel in this section: a binary symmetric channel (BSC) with probability of bit error P_e and a random erasure channel (REC) with probability of erasure P_l .

3.1. Channel coding

For either channel model, the MD system and the layered coding system must both packetize their bits. We require that for both systems, the number of data samples N included in each packet of length P bits be fixed, and that each sample use R_0^* bits for both source and channel coding.

For the BSC, a packet of length P consists of N data samples followed by an N_{CRC} -bit CRC, all protected by FEC. The CRC allows packets to be discarded if they contain any residual errors after the FEC. Thus,

$$P = NR_0^* = (NR + N_{CRC})/r^c,$$

where R bits are used for source coding each sample. We use a family of RCPC codes for the FEC, with possible code rates $r^c \in \{1, 8/9, 8/10, 8/12, \dots, 8/24\}$ [10]. The probability of block loss after channel coding, P_l^{cc} , is controlled by the BSC error probability P_e , the packet length P , the CRC, and the RCPC code rate selected.

For the REC, a packet of length P again consists of N data samples. However, now channel coding consists of applying an (n, k) Hamming code across $k = n - p$ packets, so that in a block of n packets, p missing packets can be reconstructed. The resulting probability of packet loss after channel coding, P_l^{cc} , is a function of the channel erasure probability P_l and the Hamming parameter p , where $1 - P_l^{cc} = \sum_{i=0}^p \binom{n}{i} (1 - P_l)^{n-i} P_l^i$. With the addition of channel coding, the rate available for source coding decreases, with $R_0^* = Rn/k$.

3.2. MD coding system

Figure 3 shows the system for MD coding. The two outputs of the MD source coder feed the two identical channel coders, which operate as described above. The minimum channel coding for the BSC is a CRC, so that packets with errors can be discarded. The multiplexer in Figure 3 alternates the packets or blocks intended for each ideal MD channel onto the memoryless channel.

The overall bit-rate per variable must be held constant, at R_0^* . Any channel coding takes rate away from that available for source coding. For MD source coding, we have $R = R^* + \rho$ bits per sample, with channel coding rate r_{MD}^c . The average distortion per variable can be written as

$$D(r_{MD}^c, \rho) = (1 - P_l^{cc})^2 D_0 + 2P_l^{cc}(1 - P_l^{cc}) D_1 + (P_l^{cc})^2 \sigma^2, \quad (5)$$

and we can choose the redundancy ρ and the channel coding rate r_{MD}^c to optimize D for a given channel.

3.3. Layered coding system

Figure 4 shows the layered coding system for a single physical channel. Both the HP and LP layers are channel coded using their own channel coder. A CRC is again the minimal channel coding for the BSC, to discard packets with errors. (Thus, we require that the layered decoder be restricted from extracting any information from packets with errors.)

For the layered coder, the overall bit-rate per variable, R_0^* , is constant,

$$R_0^* = (R_H + N_{CRC}/N)/r_H^c + (R_L + N_{CRC}/N)/r_L^c,$$

where r_H^c and r_L^c are the rates of the HP and LP channel coders, respectively. These are chosen to optimize the average distortion for a given probability of lost packets after channel coding, which is given by

$$D = (1 - p_l^L)(1 - p_l^H) D_{HL} + p_l^L(1 - p_l^H) D_H + p_l^H \sigma^2.$$

Here, p_l^L and p_l^H are the probabilities of block loss after channel coding for the LP and HP channels, respectively.

3.4. Comparison

We compare the performance of the two systems above numerically, assuming ideal source coders that achieve the bounds on performance for a Gaussian source. The RCPC channel coders are simulated with the BSC channel, assuming a packet length of $P = 47$ bytes, and the two-byte CRC with generator polynomial $G(x) = x^{16} + x^{15} + x^2 + 1$. Figure 5 shows the average distortion as a function of the BSC P_e , while Figure 6 shows the same as a function of the REC P_l , each with $R_0^* = 8$ bits per sample. In each case, we show the optimal system performance after jointly selecting the best channel coder and source coder rates. Also shown for reference is the performance of a one-layer coder ($R_L = 0$) with optimal bit allocation between source and channel coders. We use exhaustive search over a finely discretized set of rates for the optimal bit allocation in all cases.

In a BSC, Figure 5, we see that MD coding with simple error detection provides unsatisfactory system performance, compared even to a one-layer source coder with optimal channel coding. The three systems with error control

coding (the one-layer coder, the two-layer coder, and the MD coder) all perform essentially the same, although the MD coder performs marginally better for very high bit error rates.

Similarly, in a REC, Figure 6, MD coding without channel coding performs worse than even a one-layer coder with optimal selection of a packet-level Hamming code. The MDC system with optimized channel coding performs better than the optimized layered system for high channel erasure probabilities, greater than $10^{-1.4}$. (Note that the figure shows erasure rates as high as 100%.) At these very high erasure probabilities, the channel coder is allocated such a large rate that almost no rate is allocated to the LP layer of the two-layer coding system. Finally, because we search only over Hamming codes with $2 \leq p \leq 6$, layered coding outperforms both one-layer and MD coders for low erasure probabilities. Allowing stronger Hamming codes would eliminate this difference.

Thus, we draw the following conclusions for a memoryless channel. First, MD source coding is most effective for very high erasure and bit error probabilities. However, error control coding is more important than sophisticated source coding techniques, as its performance gains are greater. Therefore, a system using an MD source coder should also use channel coding to obtain acceptable performance on memoryless channels. A well-designed system optimizes the allocation of rate between both source and channel coding.

4. ONE CHANNEL WITH MEMORY

In this section, we examine the influence of memory on the performance of each system design. Due to space constraints, we describe results only for a finite state erasure channel. (Similar results are achieved for a Gilbert-Elliott channel.) We explore here only those system design issues concerned with the memory in the channel. In particular, we do not consider channel coding for the MD source coder, and only consider simple duplication or triplication of HP packets for the two-layer coder (with no channel coding of LP packets), although we know from above that channel coding is essential.

We consider a two-state model with a good state, G, and a bad state, B, having steady-state probabilities π_G and π_B respectively. When in state G, no erasures occur, while in state B the probability of loss is $P(\text{loss}|B)$. The probability of transitioning from B to G is $0 < g \leq \pi_G$ and the probability of transitioning from G to B is $0 < b \leq \pi_B$. When $b = \pi_B$ and $g = \pi_G$, the finite state erasure channel is memoryless.

The MD system and the comparison layered coding system are identical to those for the memoryless channel, with the exception of the multiplexer. Rather than alternating packets from each channel the multiplexer in Figure 3 generates packets each of length P bits and alternately sends M consecutive packets from each description onto the channel. Thus, the correlation between the errors seen by each description can be reduced as M is increased, at the expense of increased delay. In particular, the probability that both outputs of the MD system experience erasure is $\pi_B(\pi_B + \pi_G(1 - b - g)^M)P(\text{loss}|B)^2$. Similarly, the multiplexer in Figure 4 sends M HP packets on the channel

before sending the M duplicate (or triplicate) HP packets, to reduce the correlation between losses of repeated information.

Figure 7 shows the performance of the MD system and the layered coding system for the values of $M \in \{1, 2, 5, 10\}$. The system with layered coding is shown with both duplication and triplication of the HP layer. The right vertical axis corresponds to the memoryless channel. Thus, one could translate the results for $P_t = 10^{-1.3}$ from Figure 6 onto this axis to examine the performance gain of using better channel coders.

As the memory in the channel increases (and g decreases), the performance of each system degrades because the correlation among packet losses. Using interleaving (increasing M) to separate packets containing information about a sample improves the performance of each system by making erasures less correlated. In fact, for a given channel, the choice of the appropriate value of M to make the channel appear nearly memoryless is more important than choosing between the system with the MD source coder and the layered coder.

It is interesting to note that the optimal design of the source coders in Figure 7 are nearly independent of the channel transition probability g . However, since the layered coder with triplication outperforms duplication for small g but is worse for large g (for the given π_G), the design of the optimal channel coder may depend heavily on the transition probability g . Different values of π_G have somewhat different behavior in this regard.

5. DISCUSSION

In this paper, we considered only the case of one physical channel. If there are two physical channels that do not have independent errors or erasures, we expect MDC to outperform a layered coder only if there is a distinction between the two physical transport channels, in that what happens on one channel should be different than what happens on the other channel.

We also did not consider the ability of a decoder to extract useful information from partially corrupt data. Because MD source coding assumes a complete erasure and is not able to distinguish between the case where a single bit is in error and that where many bits are in error, it would not perform as well as a decoder that can process partially corrupt information.

Finally, we only considered a memoryless Gaussian source. We expect source coding will play a more important role relative to channel coding for sources with memory.

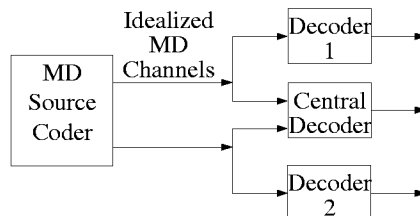


Figure 1: MD source coder for ideal MD channels.

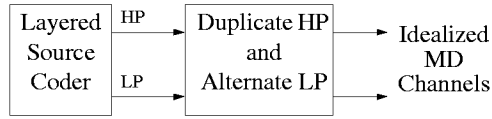


Figure 2: Layered source coder for ideal MD channels.

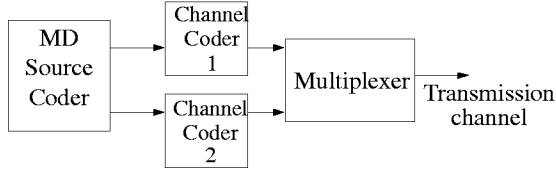


Figure 3: MD source coder for one real channel.

6. REFERENCES

- [1] L. Ozarow, "On a source coding problem with two channels and three receivers," BSTJ, vol. 59, p. 1921, Dec. 1980.
- [2] V. A. Vaishampayan, "Design of multiple description scalar quantizer," IEEE Trans. Inform. Theory, vol. 39, pp. 821–834, May 1993.
- [3] A. Ingle and V. A. Vaishampayan, "DPCM system design for diversity systems with applications to packetized speech," IEEE Trans. Speech and Audio Processing, vol. 3, pp. 48–57, Jan. 1995.
- [4] Y. Wang, M. Orchard, and A. Reibman, "Multiple Description Image Coding for Noisy Channels by Pairing Transform Coefficients," in Proc. IEEE 1997 First Workshop on Multimedia Signal Processing, (Princeton, NJ), Jun., 1997.
- [5] M. Orchard, Y. Wang, V. Vaishampayan, and A. Reibman, "Redundancy Rate Distortion Analysis of Multiple Description Image Coding Using Pairwise Correlating Transforms," in Proc. ICIP97, (Santa Barbara, CA), Oct., 1997.
- [6] V. K. Goyal and J. Kovacevic, "Optimal multiple description transform coding of Gaussian Vectors", Data Compression Conference, pp. 388–397, 1998.
- [7] Y. Wang, M. Orchard, and A. R. Reibman, "Optimal Pairwise Correlating Transforms for Multiple Description Coding", in Proc. ICIP98, (Chicago, IL), Oct., 1998.
- [8] H. Jafarkhani and V. Tarokh, "Multiple Description Trellis Coded Quantization", in Proc. ICIP98, (Chicago, IL), Oct., 1998.
- [9] H. Jafarkhani and N. Farvardin, "Design of Channel Optimized Vector Quantizers in the Presence of Channel Mismatch," in Proc. ICASSP-98, (Seattle, WA), May 1998.
- [10] J. Hagenauer, "Rate-Compatible Punctured Convolutional Codes (RCPC Codes) and their Applications," IEEE Trans. on Com., vol. 36, No. 4, pp. 389–400, Apr. 1988.

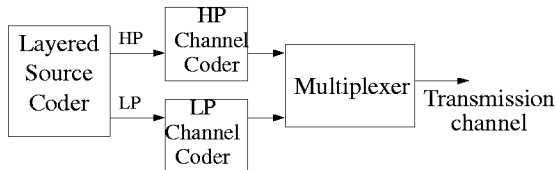


Figure 4: Layered source coder for one real channel.

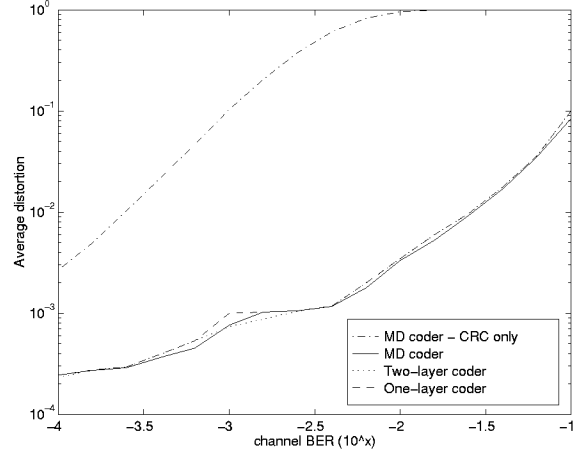


Figure 5: Performance on a BSC.

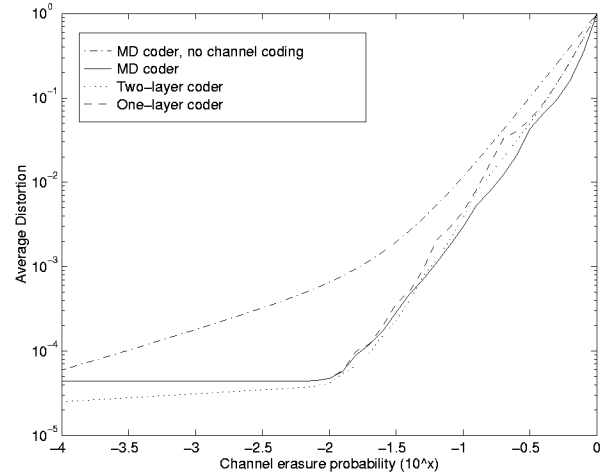


Figure 6: Performance on a random erasure channel.

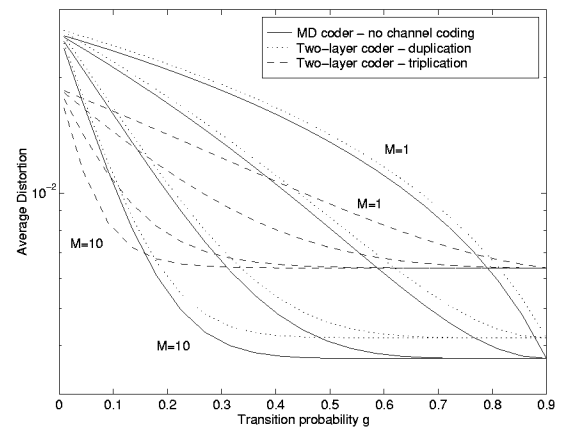


Figure 7: Performance on a 2-state erasure channel, $\pi_G = .9$, $P(\text{loss}|B) = 1/2$, $P(\text{erasure})=10^{-1.3}$.