Asymptotical Analysis of Several Multiple Description Scenarios with $L \geq 3$ Descriptions

Sorina Dumitrescu *Senior Member, IEEE,* and Ting Zheng

***Abstract*—This work compares theoretically the performance of several representative practical multiple description (MD) frameworks with $L \geq 3$ symmetric descriptions. The first scenario is the classic unequal erasure protection (UEP) scheme using a successively refinable code (SRC) and Reed-Solomon codes. The second scenario is an improvement upon UEP by applying domain partitioning and permuted Reed-Solomon codes. The third scenario uses a finer partitioning and erasure correction via repetition codes. Additionally, the MD lattice vector quantizer and another recent MD scheme are considered in the comparison.**

The aforementioned MD schemes are compared in terms of the expected squared error asymptotically achievable as the rate R of a description approaches ∞, assuming independent description losses. Our analysis reveals that the improvement of the second scenario upon the first one when $R \to \infty$ can reach up to 1.68 dB, but it approaches 0 as the description loss rate p goes to 0 and L approaches ∞. Additionally, we find that the first two schemes outperform the third one with an unbounded gain when $R \to \infty$, as $p \to 0$ or $L \to \infty$. Further, we show that the first three scenarios achieve unbounded improvements over the other two as $R \to \infty$ and $p \to 0$.

On the other hand, we point out that some of the results of our asymptotic analysis rely on strong assumptions and therefore an experimental validation is needed before applying them to practical situations.

***Index Terms*—Multiple description coding, asymptotical analysis, high resolution assumption.**

I. Introduction

The central idea in multiple description (MD) coding is to produce several descriptions of a source such that from each description the source can be reconstructed to a certain fidelity, while any subset of descriptions can be jointly decoded improving the quality of the reconstruction. As the number of jointly decoded descriptions gradually increases the quality of the reconstruction improves consistently.

Research on MD code design has received consistent attention over the past two decades due to increasing potential for applications in modern communication systems. Some of the practical designs developed over the years for two descriptions were based on scalar [1]–[5], vector [6], or lattice vector quantizers (MDLVQ) [7], on correlated transforms [8], [9], on domain partitioning [10]–[13], on low density generator matrix codes (LDGM) [14], etc. We would like to point out that by practical designs we understand constructive and computationally tractable schemes as opposed to nonconstructive approaches that are routinely used to derive information theoretical limits.

In this work we are interested in MD schemes with $L \geq 3$ symmetric descriptions, where ”symmetric” means that all descriptions have the same rate and the quality of the reconstruction is a function of the number of jointly decoded descriptions. Extensions of the aforementioned approaches from $L = 2$ to $L \geq 3$ exist, but they are generally marked by restrictions. For instance, in the case of MD scalar quantizers, the generalization to more than two descriptions requires developing good index assignments for the general case. This problem was tackled in [15] for the special case when only the side distortions (achieved for individual descriptions) and the central distortion (achieved when all descriptions are jointly decoded) are of interest. On the other hand, in [16] the optimal index assignment problem was solved for another particular case, namely when all component quantizers have convex bins. Finally, a multi-stage index assignment is proposed in [17], but with an explicit construction only for the case of three descriptions.

The existing MDLVQ framework for $L \geq 3$ was introduced in [18] and further investigated in [19]–[21]. The MDLVQ consists of a lattice vector quantizer (termed central quantizer) and an index assignment which maps each central lattice point to an L-tuple of points from a sublattice. When all descriptions are available at the decoder the central lattice point is used as the reconstruction. When only a subset of descriptions is available, the reconstruction is the arithmetic average of the sublattice points corresponding to the received descriptions. Notice that this framework has only one degree of freedom, while $L-1$ degrees of freedom are desired as pointed out in [17].

Very recently the use of LDGM codebooks in conjunction with message passing algorithms was proposed for approaching in practice known achievable points of the MD rate-distortion region. The MD scheme of [14] is designed for $L = 2$ and finite alphabet sources, targeting the Zhang-Berger region of [22], while the MD scheme of [23] is developed for a general L, but for the special case when the decoder can receive either only one or all descriptions. Additionally, the framework is designed for

The first author is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada. The second author was with the same department at the time the paper was written. Emails: sorina@mail.ece.mcmaster.ca, zhengt@grads.ece.mcmaster.ca.

the case of a Gaussian source with squared error distortion, relying on specifics of the solution to the MD problem in this situation. The extension of this approach to more general cases is currently under research.

Another MD system for L symmetric descriptions, which was extensively investigated in the context of robust image/video transmission over packet lossy channels/networks[1], is based on forward error correction (FEC) [24]–[26]. It uses a a successively refinable source code (SRC) and unequal erasure protection via maximum separable distance codes, such as Reed-Solomon (RS) codes. The output of the SRC is first divided into consecutive segments (layers) of non-decreasing lengths. Then each segment is protected using an RS code of fixed total length L. Finally, the L descriptions are formed across the channel codewords. This framework is also known as priority encoded transmission (PET) [27], or simply as the unequal error protection (UEP) scheme. This technique has attracted a lot of attention in multimedia streaming research where scalable image and video coders are widely available, such as SPIHT [28], EBCOT [29], 3D SPIHT [30], or H.264 SVC [31]. The problem of optimal redundancy allocation in rate-distortion (RD) sense has been thoroughly researched as well [24]–[26], [32]–[37]. Furthermore, it was shown in [38] that the individual description rate needed to achieve an L-tuple of distortions, where each distortion corresponds to a number of decoded descriptions, is only higher by a constant amount than the theoretical optimum, result which, along with the simplicity of the scheme, encourages its widespread application.

An improvement to the UEP scheme, termed MUEP (which stands for Multi-stream UEP), was proposed in [39] in the context of wavelet-based image coders. In MUEP the set of wavelet coefficients is first partitioned into L subsets and the SRC is applied separately to each subset. Further, permuted RS codes of length L and non-increasing strengths are applied across the descriptions. As argued in [39], when the RD curves of the L SRC streams are identical and, additionally, coincide with the RD curve of the SRC applied to the whole set, the MUEP scheme strictly outperforms UEP. On the other hand, the magnitude of this improvement was only determined empirically for several images at small bit rates (up to 0.5 bpp), using SPIHT as the SRC.

Another trend in MD coding relies on partitioning in the transform domain. Such approaches have been widely investigated for applications in image and video transmission as well. In [10]–[13] schemes for $L = 2$ were designed, whereas [40]–[42] address the case of $L \geq 3$. The approach can be described as follows. In order to produce L descriptions the set of transform coefficients is partitioned into N subsets $\mathcal{P}_1, \cdots, \mathcal{P}_N$, for some $N \geq L$. Next, each subset is encoded at I different rates $R_1 < R_2 < \cdots < R_I$, for some $I \geq 2$, thus producing I different bitstreams for each subset. Finally, each description is formed by selecting one bitstream for each subset and concatenating them. From the received descriptions the decoder uses the highest resolution bitstream available for each subset $\mathcal{P}_n$, $1 \leq n \leq N$. Notice that when the source coder used to encode each subset $\mathcal{P}_n$ is an SRC, the bitstream encoding $\mathcal{P}_n$ at some rate R_i is actually a prefix of the bitstream encoding $\mathcal{P}_n$ at any higher rate R_j, $j > i$. Thus, the i-th layer output by the source coder will appear in any description which contains the R_j-rate bitstream, for any $j \geq i$. Consequently, the aforementioned MD scenarios can be regarded as FEC-based MD schemes, where the FEC is performed using repetition codes. Furthermore, for the MD frameworks of [10]–[13], [40], [41][2], since $I = 2$, there are only two layers. The base layer is protected via a length L repetition code, which is also an $(L, 1)$ RS code, while the refinement layer has no channel protection. It follows that when $L = 2$ such a scheme is equivalent to MUEP, while when $L \geq 3$ it is a particular case of MUEP with $R_2 = R_3 = \cdots = R_L$. The approach of [42] however has $L - 1$ degrees of freedom since $I = L$, but it is not a particular case of MUEP because the number of subsets is $N = L!$ and the unequal erasure protection is achieved by varying the length of the repetition codes. Finally, a related framework is [43], which was developed as an improvement to [40] using staggered scalar quantizers instead of repetition codes.

II. Contribution

While a multitude of practical MD schemes for $L \geq 3$ symmetric descriptions have been proposed so far, a thorough performance comparison between them is lacking. As a notable example, the UEP scheme, which is one of the first practical MD frameworks, is not directly compared with most of the newer approaches. The aim of this work is to address this shortcoming by performing an analytical comparison of several representative MD schemes for $L \geq 3$ symmetric descriptions. For this purpose we select UEP, MUEP, the MD system of [42], MDLVQ and the scheme of [43]. We will alternatively refer to these five MD frameworks as scenario 1, 2, 3, 4 and 5, respectively. We point out that among these five schemes a direct comparison was performed only between MUEP and UEP (for several practical cases with description rate up to 0.5 bits per sample), and between scenario 5 and MDLVQ (for $L = 4$ and description rate of 5 bits per sample).

In our analysis we use as a measure of performance the expected distortion at the receiver, with the squared error as the distortion measure, assuming that the description

[1]Notice that in the context of data transmission over packet lossy channels, each packet can be regarded as a description.

[2]Notice that MD scheme in [41] has both symmetric and asymmetric variants. In this work we refer exclusively to the symmetric (or balanced) variant.

losses are independent. To make the analysis theoretically tractable we assume that the rate R of individual description grows to ∞ and, further, use the high resolution quantization assumption to derive the expression of the distortion. Additionally, we will assume that the SRC used in UEP, MUEP and the MD system of [42] is based on scalar quantization, assumption which is in accordance with the practice in image and video coding. Finally, the expected distortion for each scenario is minimized by solving the related redundancy allocation problem with a constraint on the rate of the individual description.

For scenarios 4 and 5 we will use the high rate analysis available in the literature and will complement it with the derivation of the analytical expression of the optimal expected distortion. For scenarios 1, 2 and 3, deriving the optimal expected distortion is more challenging since the associated optimization problem has $L-1$ variables versus only one variable in scenarios 4 and 5. The problem was addressed in the context of UEP without assuming a particular expression of the operational RD function of the SRC and without deriving an analytical expression for the solution. In this work we are interested in the analytical expression of the solution in order to facilitate the analysis, to which aim we will exploit the convenient exponential form of the operational RD function of the SRC at high resolution. Inspired by previous work we convert the optimization problem into a simpler problem with fewer constraints, conversion which is based on the construction of the lower convex hull of a certain curve, which we will refer to as the *alpha-beta* curve. Based on the convex hull of the alpha-beta curve we next derive the analytical expression of the optimal expected distortion. To have a complete analytical solution we need to have the explicit expression of the convex hull as well. Therefore, we proceed to investigate the convexity properties of the alpha-beta curve for each scenario. For scenario 3 it turns out that the curve is convex all the time, while for the other two schemes we show that the curve is convex only when the description loss rate p is sufficiently small. For the remaining values of p (up to $\frac{1}{2}$) we identify properties of the curve which considerably simplify the convex hull construction. These results will simplify the expression of the optimal expected distortion for certain cases of interest facilitating the theoretical analysis. The significance of these results transcends this work because they establish a basis for the theoretical comparison with other existing MD schemes not covered here, or to be developed in the future.

Next we proceed to the asymptotical comparison of the five scenarios. Numerical evaluation shows that the gain in performance of the second scenario upon the first one as $R \to \infty$ may attain 1.68 dB (when $L = 3$ and $p \to 0$), and remains higher than 0.9 dB for all $L \leq 10$ and $p \leq 0.1$. Using a result from [38] it follows that this gain remains bounded over all L, while our theoretical analysis shows that it approaches 0 as $p \to 0$ and $L \to \infty$. We point out that some preliminary results on the asymptotical comparison between UEP and MUEP for a Gaussian memoryless source were first mentioned in [44].

Regarding the relation between UEP/MUEP and the third scheme, our analysis reveals that each of the first two schemes outperforms the third one in the limit as $R \to \infty$, when p is sufficiently small or L is sufficiently large, and that the gain goes to ∞ as $p \to 0$ while L is fixed, or as $L \to \infty$ while p is fixed. The third scenario may also be superior to each of the other two in certain cases, for instance, when $L = 3$, $R = 1$ and $p \leq 0.35$. However, we show that if the performance of scenario 3 as $R \to \infty$ is higher than that of the other two schemes for some $p \in (0, \frac{1}{2}]$, then this could happen only for a finite range of values of L and with a gain in performance that remains upper bounded by a constant.

Finally, the analytical comparison of the three scenarios based on SRC with the MDLVQ framework and with the MD scheme of [43] shows that the former three schemes outperform the latter two when $R \to \infty$ and p is sufficiently small, and that the difference in performance becomes unbounded as p approaches 0. This result could be attributed to the fact that each of the former three scenarios has $L-1$ degrees of freedom, while each of the latter two schemes has only one degree of freedom. Additionally, we prove that for $R \to \infty$ and p small enough scenario 5 is always better than MDLVQ for all $L \geq 4$, conclusion which agrees with the numerical results reported in [43] for $L = 4$ and $R = 5$.

The paper is structured as follows. The next section describes in detail the three MD scenarios based on SRC. Section III discusses the general formulation of the RD optimization problem covering the first three scenarios and its analytical solution given in terms of the lower convex hull of the so-called *alpha-beta* curve associated to the problem. The following section addresses the RD optimization problem for the case of independent description losses. The convexity properties of the alpha-beta curve are investigated for each scheme. The results are used in Section V, respectively VI, to compare the asymptotical performance between the three scenarios, respectively versus MDLVQ and the scheme of [43]. Finally, Section VII concludes the paper.

III. The Three MD Scenarios Based on SRC

In this section we describe the first three MD scenarios analyzed in this work. We start by specifying the common aspects of the three schemes and clarifying the notations.

We assume that a linear orthonormal transform was first applied to the signal, for instance an orthonormal wavelet transform in the case of images, with the purpose of decorrelating the signal[3]. Therefore, the signal distortion equals the distortion of the transform coefficients. We consider an SRC which operates on the transform coefficients and denote by $D_o(R)$ the corresponding operational RD function, where R is measured in bits per sample and the distortion in squared difference per sample. In scenarios 2 and 3 the set of transform coefficients is divided into subsets of equal size and the same SRC is separately applied on each subset. We assume that the operational RD function of the SRC on each subset coincides with the operational RD function on the whole set[4].

For all three scenarios the bitstream output by the SRC is divided into L consecutive segments called layers. We denote by R_k the rate of the prefix formed of the first k layers, $1 \leq k \leq L$, and by $\mathbf{R}$ the L-tuple $(R_1, \cdots, R_L)$. By convention, $R_0 = 0$. A distinctive feature of the SRC is that if the decoder has all layers 1 through k, but does not have layer $k+1$, then the first k layers can be decoded while any available layers beyond the $(k+1)$-th are useless.

Additionally, let $D_i(\mathbf{R}, k)$ denote the average distortion when only k descriptions are received in the case of scenario $i, i = 1, 2, 3$, and rate L-tuple $\mathbf{R}$, where the average is computed assuming that each subset of k descriptions is equally likely. Finally, $R_i(\mathbf{R})$ denotes the rate of a single description for scenario i and rate tuple $\mathbf{R}$.

A. Scenario 1 (UEP)

In scenario 1 the whole set of transform coefficients is subjected to the SRC, and the output bitstream is divided into L layers. Next, for each $k, 1 \leq k \leq L$, the k-th layer is partitioned into groups of k consecutive symbols (a symbol is a group of a fixed number of bits) which are further encoded by a strict systematic (L, k) Reed Solomon (RS) code. We use the term "strict systematic code" in order to enforce the fact that the information symbols are placed at the beginning of the channel codeword. Such a code ensures that the k information symbols can be recovered from any k channel symbols available at the decoder. The L descriptions are formed across the channel codewords, such that for each j, the j-th description contains the j-th symbol of each channel codeword. Figure 1(a) illustrates this MD scheme for $L = 3$.

UEP ensures that when only $k, 1 \leq k < L$, descriptions are available at the decoder, the first k layers of the source bitstream can be completely recovered. The missing source symbols from layers $k+1$ to L will prevent the decoding of any available symbols from these layers. Therefore, only the first k layers can be decoded leading to $D_1(\mathbf{R}, k) = D_o(R_k)$, for $1 \leq k \leq L$. Additionally, note that $R_1(\mathbf{R}) = \sum_{i=1}^{L} \frac{R_i - R_{i-1}}{i} = \sum_{i=1}^{L-1} \frac{1}{i(i+1)} R_i + \frac{R_L}{L}$.

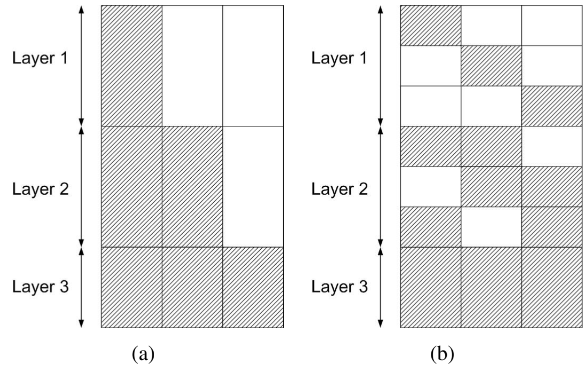

Fig. 1. Illustration of UEP (a) and MUEP (b) schemes for $L = 3$. Each column forms a description. The shaded rectangles represent source symbols and the white rectangles redundancy symbols. The portion marked as "Layer i" represents the $(3, i)$ RS codewords, $i = 1, 2, 3$.

B. Scenario 2 (MUEP)

In scenario 2 [39] the set of transform coefficients is split into L subsets of equal size and each subset is separately encoded using the SRC. This way L independently decodable and progressively refinable source streams are produced. Each such stream is divided into L layers, layer k having bit rate $R_k - R_{k-1}$, $1 \leq k \leq L$. Further, for each k, the source symbols belonging to the k-th layer (from all streams) are partitioned into groups of k, such that any two symbols in a group come from different streams. An algorithm to generate such a grouping is given in [39][5]. Each such group is encoded by an (L, k) permuted systematic RS code. We use the term "permuted systematic code" in order to emphasize that the information symbols are not necessarily placed at the beginning of the codeword, but can be positioned anywhere in the codeword. More specifically, the RS channel codeword encoding a group of source symbols $s_{i_1}, s_{i_2}, \cdots s_{i_k}$, where s_{i_j} belongs to sub-stream i_j, will have its i_j-th channel symbol equal to s_{i_j}, $1 \leq j \leq k$. The descriptions are next formed across the channel codewords, such that for each ℓ, the ℓ-th description contains the ℓ-th symbol of each channel codeword, $1 \leq \ell \leq L$. Figure 1(b) depicts this MD scenario for $L = 3$.

To summarize, each description $i, 1 \leq i \leq L$, contains the source symbols of stream i from all L layers and some additional RS redundancy symbols. When only $k < L$ descriptions are received at the decoder, the first k layers from the missing descriptions will also be recovered. Then the subsets encoded by the received descriptions will be

[3]In the case of a memoryless source the transform is the identity mapping.

[4]If the transform coefficients are the outputs of a memoryless source, this assumption holds asymptotically as the number of samples goes to infinity. However, in practical situations, encoding separately the subsets might incur some loss in performance versus the case when they are encoded together.

[5]This algorithm uses the assumption that the total number of source symbols in layer k is a multiple of k.

reconstructed with distortion $D_o(R_L)$, while the subsets encoded by the missing descriptions will be reconstructed with distortion $D_o(R_k)$. In conclusion,

$$D_2(\mathbf{R},k) = \frac{L-k}{L} D_o(R_k) + \frac{k}{L} D_o(R_L), \qquad (1)$$

for all $1 \le k \le L$. Finally, it is easy to see that $R_2(\mathbf{R}) = R_1(\mathbf{R})$, while $D_2(\mathbf{R},k) \le D_1(\mathbf{R},k)$ for all $1 \le k \le L$, with strict inequality if $D_o(R_k) > D_o(R_L)$.

C. Scenario 3

In scenario 3 [42] the set of transform coefficients is partitioned into $N = L!$ subsets $\mathcal{P}_1, \cdots, \mathcal{P}_N$, of equal cardinality. Each subset is separately encoded using the SRC and the output is divided into L layers such that the bit rate of layer k to be $R_k - R_{k-1}$, $1 \le k \le L$. To each subset $\mathcal{P}_n$ a a permutation $\pi_n : \{1, 2, \cdots, L\} \to \{1, 2, \cdots, L\}$, is assigned in an one-to-one manner. Then description m is formed by concatenating the bitstream of rate $R_{\pi_n(m)}$ encoding subset $\mathcal{P}_n$, for all $1 \le n \le L!$, $1 \le m \le L$.

According to [42], one has

$$D_3(\mathbf{R},k) = \sum_{j=k}^{L} \frac{N(k,j)}{L!} D_o(R_j), \qquad (2)$$

where

$$N(k,j) = \begin{cases} 0, & \text{if } 1 \le j < k \\ \frac{k(j-1)!(L-k)!}{(j-k)!}, & \text{if } k \le j \le L \end{cases}. \qquad (3)$$

Finally, $R_3(\mathbf{R}) = \frac{1}{L}\sum_{k=1}^{L} R_k$.

D. Relation between scenario 3 and scenarios 1 and 2

It can be easily seen that the description rate and the expected distortion are identical for scenarios 2 and 3 when $L = 2$ or when $L \ge 3$ and only layer L is nonempty, i.e., $R_1 = \cdots = R_{L-1} = 0$. On the other hand, the relation between the performances of scenario 3 and the other two for $L \ge 3$ and general $\mathbf{R}$ is not trivial. In particular, the fact that scenario 2 uses RS codes, while scenario 3 uses repetition codes, suggests that the second scenario has better performance when the number k of received descriptions is sufficiently high. A natural question is then: can scenario 2 ever beat scenario 3? As we will see in section VI the answer is positive. An example when this happens is when $L = 3$, the rate of each description is 1, and the probability of description loss is higher than 0.35. We believe that this is due to the fact that scenario 3 may exhibit an advantage over scenario 2 when the number of received descriptions k is small. This possible advantage may come from the fact that $D_3(\mathbf{R},k)$ is a weighted sum of distortions $D_o(R_j)$ at all resolutions R_j higher than R_k, rather than at only two resolutions R_k and R_L as in the case of scenario 2.

To be more specific, let $\mathbf{R}$ and $\mathbf{R}'$ denote the L-tuple of rates for scenario 2, respectively 3. Let us also assume that $R_2(\mathbf{R}) = R_3(\mathbf{R}')$ and that $R_L = R'_L$ so that $D_2(\mathbf{R},L) = D_3(\mathbf{R}',L)$. Then one has

$$\sum_{i=1}^{L-1} \frac{1}{i(i+1)} R_i = \frac{1}{L}\sum_{i=1}^{L-1} R'_i. \qquad (4)$$

Let k_0 denote the integer such that $k_0(k_0+1) \le L < (k_0+1)(k_0+2)$. Then $\frac{1}{i(i+1)} > \frac{1}{L}$ for $1 \le i < k_0$, while $\frac{1}{i(i+1)} < \frac{1}{L}$ for $k_0+1 \le i \le L-1$. These relations together with (4) imply that there are tuples $\mathbf{R}$ and $\mathbf{R}'$ such that $R_i > R'_i$ for $k_0+1 \le i \le L$, and $R_i < R'_i$ for $1 \le i < k_0$. The fact that the rates R_i for high resolutions are larger for scenario 2 may lead to $D_2(\mathbf{R},k) < D_3(\mathbf{R}',k)$ for large k. On the other hand, as k becomes smaller, and, in particular, smaller than k_0, the term $\frac{L-k}{L} D_o(R_k)$ from the expression of $D_2(\mathbf{R},k)$ in (1) may become larger than the sum $\sum_{j=k}^{L-1} \frac{N(k,j)}{L!} D_o(R'_j)$ from the expression of $D_3(\mathbf{R}',k)$ from (2). Then there is the possibility that $D_3(\mathbf{R}',k) < D_2(\mathbf{R},k)$.

IV. Derivation of Optimal Expected Distortion for Scenarios 1, 2 and 3

This section addresses the problem of optimizing the expected distortion for given channel loss statistics for scenarios 1-3. We will assume that the transform coefficients can be modeled as the output of a memoryless source with a smooth probability density function (pdf) $\rho(x)$. We will, further, assume that the SRC applied to the transform coefficients is a successively refinable (or multiresolution) scalar quantizer followed by an entropy coder applied to each layer [5], [45]. Additionally, we will use the high resolution quantization assumption in order to make the analysis tractable. According to this assumption, since the pdf is smooth, when the quantizer rate is sufficiently high the pdf is approximately uniform over each quantization bin. Based on this assumption it was shown in [46] that the optimal entropy constrained single resolution scalar quantizer at sufficiently high rate R, is a uniform quantizer with step size $\delta(R) = 2^{h-R}$, where $h = -\int_{\mathbb{R}} \rho(x)\log_2 \rho(x)dx$, and its distortion is approximately equal to $\frac{1}{12}2^{2(h-R)}$. Let us denote by $Q_u(R)$ this quantizer. Notice that, given an increasing sequence of rates $R_1 \le R_2 \le \cdots \le R_L$, such that $2^{R_{k+1}-R_k}$ is an integer for every $1 \le k \le L-1$, the encoder partitions of quantizers $Q_u(R_1), \cdots, Q_u(R_L)$ are embedded, therefore these quantizers form a successively refinable scalar quantizer. Based on the aforementioned considerations we make the assumption that

$$D_o(R_k) = \tau 2^{-2R_k}, \text{ for } 1 \le k \le L, \text{ where } \tau = \tfrac{1}{12}2^{2h}.$$

Additionally, for scenarios 2 and 3, which involve partitioning of the set of transform coefficients into N subsets $\mathcal{P}_1, \cdots, \mathcal{P}_N$, we assume the following simple partitioning rule. For every $i \ge 1$, the i-th transform coefficient is included in subset $\mathcal{P}_{i(modN)}$, where $i(modN)$ denotes the remainder of the integer division of i by N. Then, as the number of samples grows to ∞, the operational RD function of the SRC on each subset becomes identical to $D_o(\cdot)$, therefore we will assume their equality.

To measure the performance of each MD code we employ the expected distortion of the source reconstruction at the decoder conditioned on the fact that at least one description is received[6] and we will refer to this quantity simply as the expected distortion in the sequel. Finally, we will denote by $U(k)$ the conditional probability that only k descriptions are available at the decoder given that at least one is received, for $1 \leq k \leq L$. Then the expected distortion for scenario i and rate tuple $\mathbf{R}$ is $ED_i(\mathbf{R}) = \sum_{k=1}^{L} U(k) D_i(\mathbf{R}, k),\ i = 1, 2, 3$. The associated optimization problem of minimizing $ED_i(\mathbf{R})$ under the constraint $R_i(\mathbf{R}) \leq R$, for some fixed $R > 0$, has the following general formulation covering all three scenarios.

Problem P.

$$\begin{aligned} \text{minimize}_{\mathbf{R}} \quad & \tau \sum_{k=1}^{L} \beta(k) 2^{-2R_k} \\ \text{subject to} \quad & 0 \leq R_1 \leq R_2 \leq \cdots \leq R_L \\ & \sum_{k=1}^{L} \alpha(k) R_k = R, \end{aligned} \tag{5}$$

where the coefficients $\alpha(k), \beta(k)$ are strictly positive for all $k, 1 \leq k \leq L$.

For scenarios 1 and 2 one has $\alpha(k) = \frac{1}{k(k+1)}$, for $1 \leq k \leq L-1$, and $\alpha(L) = \frac{1}{L}$, while for scenario 3, one has $\alpha(k) = \frac{1}{L}$ for $1 \leq k \leq L$. Furthermore, $\beta(k) = U(k)$ for scenario 1, $\beta(k) = V(k)$ for scenario 2 and $\beta(k) = W(k)$ for scenario 3, for all $1 \leq k \leq L$, where $V(k) \triangleq U(k)\frac{L-k}{L}$, for $1 \leq k \leq L-1$, $V(L) \triangleq \sum_{k=1}^{L} \frac{k}{L} U(k)$, and $W(j) \triangleq \frac{1}{L!} \sum_{k=1}^{j} N(k, j) U(k)$, for $1 \leq j \leq L$. We will refer to Problem P specialized to scenario i as problem P(i).

We would like to mention that in practical situations the RD optimization problem of interest additionally contains combinatorial constraints since the symbols used in the MD scheme cannot be broken down into fractional portions. The problem of RD optimization for scenario 1 (including combinatorial constraints) was considered in prior work [24]–[26], [32]–[37], mostly assuming a general function $D_o(R)$. A globally optimal solution algorithm for the most general formulation is given in [32] and a significantly faster solution in [34]. The remaining works mainly use the convexity assumption of the RD curve to obtain efficient globally optimal [33], [34], [37], or sub-optimal solution algorithms. For the latter case, some of these works use the continuous relaxation of the problem by removing the combinatorial constraints [24], [25], [36].

In this work we discard the combinatorial constraints and assume that each R_k may take any positive value. The special form of the operational RD function along with the specifics of the packet loss scenario considered will allow us to derive the analytic problem solution for certain cases, which is one of the contributions of this work. Our approach to solve problem P is inspired by [25], which relies on converting the problem into a simpler one without the constraints (5).

The crucial step for simplifying the problem is the construction of the lower convex hull of the alpha-beta curve which we define next. Consider the set $\mathcal{P}$ of planar points P_k, $0 \leq k \leq L$, with coordinates (x_k, y_k), where $x_0 = y_0 = 0$, $x_k = x_{k-1} + \alpha(k)$, and $y_k = y_{k-1} + \beta(k)$ for $1 \leq k \leq L$. The *alpha-beta curve* associated to problem P is define as the union of line segments $P_{k-1}P_k$ for $1 \leq k \leq L$. We are particularly interested in the lower convex hull of this curve. Therefore, let P_{j_i} with $0 \leq i \leq M \leq L$, and $0 = j_0 < j_1 < j_2 < \cdots < j_M = L$, be the extreme points of the set $\mathcal{P}$ situated on its lower convex hull. In other words, the lower convex hull of the alpha-beta curve consists of the union of line segments $P_{j_{i-1}}P_{j_i}$ for $1 \leq i \leq M$.

Now consider the following simpler problem derived from problem P. We will refer to it as problem DP, and we will use the notation DP(i) for its specialization to scenario i.

Problem DP.

$$\begin{aligned} \text{minimize}_{\mathbf{R}'} \quad & \tau \sum_{k=1}^{M} \beta'(k) 2^{-2R'_k} \\ \text{subject to} \quad & R'_k \geq 0, \quad 1 \leq k \leq M \\ & \sum_{k=1}^{M} \alpha'(k) R'_k = R, \end{aligned}$$

where

$$\alpha'(k) = \sum_{i=j_{k-1}+1}^{j_k} \alpha(i), \quad \beta'(k) = \sum_{i=j_{k-1}+1}^{j_k} \beta(i),$$

for $1 \leq k \leq M$, and $\mathbf{R}' = (R'_1, \cdots, R'_M)$. Notice that the alpha-beta curve associated to problem DP is actually the lower convex hull of the alpha-beta curve of problem P, therefore it is convex, i.e.,

$$\frac{\beta'(k)}{\alpha'(k)} \leq \frac{\beta'(k+1)}{\alpha'(k+1)}, \quad 1 \leq k \leq M-1. \tag{6}$$

The following result, which follows directly from [25], establishes the connection between the solutions of problems P and DP.

Proposition 1. *Let $\mathbf{R}'_{opt} = (R'_{k,opt})_{1 \leq k \leq M}$ denote an optimal solution to problem DP. Then an optimal solution $\mathbf{R}_{opt} = (R_{i,opt})_{1 \leq i \leq L}$ to problem P can be obtained as*

[6]We use this conditional expected distortion instead of the expected distortion over all possibilities, in order to simplify the performance analysis.

follows

$$R_{j_{k-1}+1,opt} = R_{j_{k-1}+2,opt} = \cdots = R_{j_k,opt} = R'_{k,opt}.$$

for $1 \le k \le M$. *Furthermore, Problems P and DP have the same value of the objective function at optimality.*

In order to present the solution to problem DP we need the following notations

$$S(j) \triangleq \sum_{k=j}^{M} \alpha'(k),$$

$$R(j) \triangleq \frac{1}{2} \sum_{k=j+1}^{M} S(k) \log_2 \frac{\beta'(k)\alpha'(k-1)}{\alpha'(k)\beta'(k-1)}, \qquad (7)$$

for any $j, 1 \le j \le M$, with the convention that $\sum_{k=b}^{a} \gamma(k) = 0$, for any $b > a$ and $\gamma(k)$. We also make the convention that $S(M+1) = 0$ and $R(0) = \infty$. As we will see shortly, the values $R(j)$ are thresholds for the parameter R in problem DP, which determine the form of the solution. Notice that $R(M) = 0$. Therefore, the inequalities $R(M) \le R < R(0)$ are always satisfied. Furthermore, (6) implies that $R(j) \le R(j-1)$ for $2 \le j \le M$. Finally, the next result, whose proof is deferred to Appendix A, presents an explicit expression of the solution of problem DP.

Proposition 2. *Consider problem DP and let* k_0 *be the unique integer in the set* $\{1, 2, \cdots, M\}$, *satisfying*

$$R(k_0) \le R < R(k_0 - 1).$$

Then the unique solution of problem DP is

$$R'_{k,opt} = \begin{cases} 0, & 1 \le k \le k_0 - 1 \\ \frac{R-R(k_0)}{S(k_0)} + \frac{1}{2}\log_2 \frac{\beta'(k)\alpha'(k_0)}{\alpha'(k)\beta'(k_0)}, & k_0 \le k \le M \end{cases}, \qquad (8)$$

and the minimum value of the objective function is

$$\tau \left(\sum_{k=1}^{k_0-1} \beta'(k) + S(k_0) \frac{\beta'(k_0)}{\alpha'(k_0)} 2^{\frac{-2(R-R(k_0))}{S(k_0)}} \right). \qquad (9)$$

V. Optimal Expected Distortion for Scenarios 1, 2 and 3 under Independent Description Losses

In this section we address the RD optimization problem for scenarios 1-3 assuming that the description losses are independent. Clearly, the value of the optimal expected distortion can be found by applying Proposition 2 once the convex hull of the alpha-beta curve is computed for each problem P(i), $i = 1, 2, 3$. In this section we investigate the convexity properties of the alpha-beta curve for each scenario. In particular, we determine when the alpha-beta curve is convex and for the case when it is not convex we reveal properties of the curve which allow for a simple computation of the convex hull. The proofs of the results are deferred to Appendix B.

Let us denote by $p, 0 < p \le \frac{1}{2}$, the probability that a description is lost. Then for $1 \le k \le L$, one has

$$U(k) = \frac{1}{1-p^L} \binom{L}{k} (1-p)^k p^{(L-k)}. \qquad (10)$$

Proposition 3. *Consider Problem P(1) with coefficients satisfying (10), where* $0 < p \le \frac{1}{2}$ *and* $L \ge 2$. *Then the lower convex hull of the alpha-beta curve consists of the segment lines* $P_k P_{k+1}$, $0 \le k \le S-2$ *and* $P_{S-1}P_L$ *where* S *is the smallest integer from the set* $\{1, 2, \cdots, L\}$ *such that*

$$SU(S) > \sum_{k=S+1}^{L} U(k). \qquad (11)$$

Additionally, for problem DP(1) one has $M = S$ *and, for* $1 \le j \le S-1$,

$$R(j) = \Big(\sum_{k=j+1}^{S-1} \frac{1}{2k} \Big) \log_2 \frac{1-p}{p} + \frac{1}{2S} \log_2 \frac{\sum_{k=S}^{L} U(k)}{(S-1)U(S-1)} + \sum_{k=j+1}^{S-1} \frac{1}{2k} \log_2 \frac{(L-k+1)(k+1)}{(k-1)k}. \qquad (12)$$

Furthermore, the alpha-beta curve of problem P(1) is convex, and hence $S = L$, *if and only if* $p \le \frac{1}{L^2-L+1}$.

Let us turn our attention now to Scenario 2. Then $V(k) = \frac{1}{1-p^L}\binom{L-1}{k}(1-p)^k p^{(L-k)}$ for $1 \le k \le L-1$, and $V(L) = \frac{1-p}{1-p^L}$.

Proposition 4. *Consider Problem P(2) when relations (10) are satisfied, with* $0 < p \le \frac{1}{2}$ *and* $L \ge 2$. *Then the alpha-beta curve is convex if and only if* $p \le \frac{L}{L^2-2L+2}$. *In this case one has* $M = L$ *in problem DP(2) and, for* $1 \le j \le L-1$,

$$R(j) = \Big(\sum_{k=j+1}^{L} \frac{1}{2k} \Big) \log_2 \frac{1-p}{p} + \frac{1}{2L} \log_2 \frac{1}{(L-1)(1-p)^{L-1}} + \sum_{k=j+1}^{L-1} \frac{1}{2k} \log_2 \frac{(L-k)(k+1)}{(k-1)k}. \qquad (13)$$

Additionally, when the alpha-beta curve is not convex its lower convex hull consists of the segment lines $P_k P_{k+1}$, *for* $0 \le k \le T-2$, $P_{T-1}P_{L-1}$ *and* $P_{L-1}P_L$,*where* T *is the smallest integer from the set* $\{1, 2, \cdots, L-2\}$ *such that*

$$\frac{T(L-T-1)}{L} V(T) > \sum_{k=T+1}^{L-1} V(k). \qquad (14)$$

Now let us discuss Scenario 3. Using (10) and (3) one obtains

$$W(j) = \frac{1}{L!}\sum_{k=1}^{j} N(k,j)U(k)$$
$$= \frac{(1-p)p^{L-j}}{1-p^L}\sum_{k=1}^{j}\binom{j-1}{k-1}(1-p)^{k-1}p^{j-1-(k-1)}$$
$$= \frac{(1-p)p^{L-j}}{1-p^L}.$$

It follows that $\frac{\beta(j)}{\alpha(j)} = \frac{L(1-p)p^{L-j}}{1-p^L}$ and clearly $\frac{\beta(j)}{\alpha(j)} \le \frac{\beta(j+1)}{\alpha(j+1)}$, for $1 \le j \le L-1$, when $0 < p < 1$ and $L \ge 2$. This result implies the convexity of the alpha-beta curve of Problem P(3). Further, for problem DP(3), $R(j)$ has a simple form. These observations are stated in the following result, whose proof is immediate.

Proposition 5. *Assume that $0 < p < 1$, $L \ge 2$ and that (10) is satisfied. Then the alpha-beta curve of Problem P(3) is convex. Thus, one has $M = L$ in problem DP(3), and, for $1 \le j \le L-1$,*

$$R(j) = \frac{(L-j)(L-j+1)}{4L}\log_2\frac{1}{p}.$$

VI. Asymptotical Analysis of Performance of Scenarios 1, 2 and 3

In this section we will use the results from Sections IV and V to compare the expected distortions in scenarios 1,2,3 under asymptotical conditions. We will use the notation $D_i(L,R,p)$ for the optimal expected distortion for scenario $i = 1,2,3$, as a function of L, R and p, assuming an independent description loss framework. For each $1 \le i \ne j \le 3$, we will denote $\Delta_{i/j}(L,R,p) = 10\log_{10}\frac{D_j(L,R,p)}{D_i(L,R,p)}$. Furthermore, $R_i(L,p,j)$ will denote the value of $R(j)$ for problem DP(i), $i = 1,2,3$, as a function of L and p and j.

Using Propositions 3-5 it can be easily established that for all $L \ge 2$ and $i = 1,2,3$, one has $\lim_{p\to 0} R_i(L,p,L-1) = \infty$, while $R_i(L,p,L) = 0$. It follows that for fixed $R > 0$ and $L \ge 2$, as p becomes sufficiently small relations $R_i(L,p,L) < R < R_i(L,p,L-1)$ are satisfied for all $i = 1,2,3$, and, consequently, only the L-th layer is non-empty in the optimal rate allocation. More precisely, the inequality $R < R_i(L,p,L-1)$ holds for all $i = 1,2,3$, when $0 < p < \frac{1}{L(L-1)2^{2LR}+1}$. The relation between the performances of the three schemes for this range of p values is spelled out by the following result, following directly from Propositions 2-5.

Corollary 1. *For every $L \ge 2$, $R > 0$ and $0 < p < \frac{1}{L(L-1)2^{2LR}+1}$, one has*

$$D_1(L,R,p) = \tau\Big(1 - p^L - \frac{(1-p)^L}{1-p^L}(1-2^{-2RL})\Big),$$

$$D_2(L,R,p) = D_3(L,R,p) = \tau\Big(1 - \frac{1-p}{1-p^L}(1-2^{-2RL})\Big).$$

This further implies that

$$\lim_{p\to 0} D_i(L,R,p) = \tau 2^{-2RL}, \quad i = 1,2,3.$$

Remark 1. *The above result confirms the expectation that for fixed R and L, as p approaches 0, the redundancy in the system approaches 0 for all three scenarios and their performance becomes identical.*

Next we analyze the performance of the three schemes as $R \to \infty$. The next corollary will be useful in the analysis.

Corollary 2. *For $i = 1,2,3$, $L \ge 2$, $R > R_i(L,p,1)$ and either a) $0 < p \le \frac{1}{2}$ if $(i,L) \ne (1,2)$ or b) $0 < p \le \frac{1}{3}$ if $(i,L) = (1,2)$, one has*

$$D_i(L,R,p) = \tau\frac{\beta(1)}{\alpha(1)}2^{-2(R-R_i(L,p,1))}.$$

Additionally, for $0 < p \le \frac{1}{2}$, $j = 1,3$ and $L \ge 3$, the following holds

$$\lim_{R\to\infty}\Delta_{2/j}(L,R,p) = \frac{10}{\ln 10}\Big(a(j) + 2(R_j(L,p,1) - R_2(L,p,1))\ln 2\Big). \quad (15)$$

where $a(1) = \ln\frac{L}{L-1}$ and $a(3) = \ln\frac{L}{2(L-1)}$.

Proof. We only need to prove that $\frac{\beta'(1)}{\alpha'(1)} = \frac{\beta(1)}{\alpha(1)}$ for $i = 1,2$, and the conclusion follows according to Propositions 1, 2 and 5. For this we will show that $S > 1$ and $T > 1$. It can be easily seen that $U(1) \le U(2)$ and $\frac{L-2}{L}V(1) \le V(2)$, which yield $S > 1$, respectively, $T > 1$. □

The following proposition, proved in Appendix C, quantifies the gain of MUEP versus UEP.

Proposition 6. *For $L \ge 3$ denote*

$$A(L) = \sum_{k=1}^{L-1}\frac{1}{k}\ln\frac{L-k+1}{L-k} - \frac{\ln L}{L}.$$

Then the following statements hold.

6.a) *For $0 < p \le \frac{1}{L^2-L+1}$ and $L \ge 3$, one has*

$$\lim_{R\to\infty}\Delta_{2/1}(L,R,p) = \frac{10}{\ln 10}\Big(A(L) + \frac{L-1}{L}\ln(1-p)\Big). \quad (16)$$

6.b) *For $L \ge 3$, the following relations hold*

$$\lim_{p\to 0}\lim_{R\to\infty}\Delta_{2/1}(L,R,p) = \frac{10}{\ln 10}A(L), \quad (17)$$

$$\lim_{L\to\infty}\frac{\lim_{p\to 0}\lim_{R\to\infty}\Delta_{2/1}(L,R,p)}{\frac{\ln L}{L}} = 1. \quad (18)$$

6.c) *There is some constant $c > 0$ such that $\lim_{R\to\infty}\Delta_{2/1}(L,R,p) \le c$ for all $L \ge 2$ and*

$0 < p \leq \frac{1}{2}$.

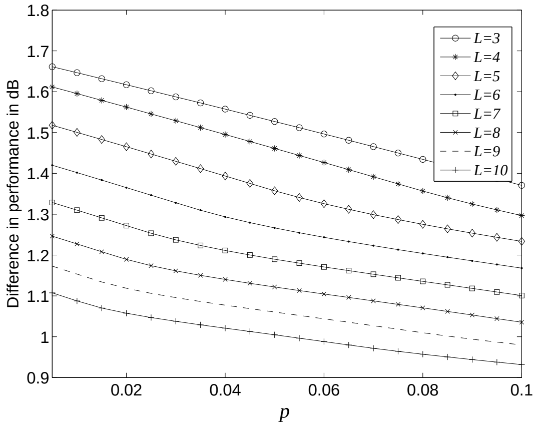

Fig. 2. Asymptotical gain in performance of MUEP versus UEP as $R \to \infty$, for $3 \leq L \leq 10$ and $0.005 \leq p \leq 0.1$.

Remark 2. *While (17) gives the exact expression of $\lim_{p\to 0}\lim_{R\to\infty}\Delta_{2/1}(L,R,p)$, relation (18) provides a simpler approximation of this quantity as $L \to \infty$. An immediate consequence is that*

$$\lim_{L\to\infty}\lim_{p\to 0}\lim_{R\to\infty}\Delta_{2/1}(L,R,p) = 0.$$

Figure 2 plots the value of $\lim_{R\to\infty}\Delta_{2/1}(L,R,p)$ for $p \in [0.005, 0.1]$, and $2 \leq L \leq 10$, according to Propositions 1-4. We see that the highest improvement is 1.68 dB, achieved for $L = 3$ and p approaching 0, while the lowest is 0.93 dB, achieved when $L = 10$ and $p = 0.1$. We would like, however, to emphasize that the application of these results to practical situations has to be done with caution because they rely on a strong assumption which may be difficult to satisfy in practice. In particular, it is assumed that the operational RD function of the SRC for each of the L subsets of transform coefficients of scenario 2, coincides with the operational RD function of the whole set, assumption which holds when the size of each subset is large enough. In practical situations the total number of samples is limited and, as L grows, the size of each subset decreases and the assumption becomes inaccurate. The analysis of the impact of such inaccuracies on the relation between scenarios 1 and 2 will be addressed in our future work.

Let us compare now scenarios 2 and 3. From the description of the two schemes it is clear that they have identical performance for $L = 2$. However, it is not clear how they compare when $L \geq 3$. Figure 3 plots the value of $\Delta_{2/3}(L,R,p)$ for $L = 3$ and $R = 1$ obtained by applying Propositions 1,2, 4 and 5. The figure shows that scenario 2 is better for smaller values of p, but it is worse for larger p. On the other hand, the improvement achieved in the former case highly offsets the loss obtained in the latter case. The two schemes also have identical performance for values of p very close to 0. This corresponds to the case identified in Corollary 1 when only the last layer is nonempty. The following proposition, proved in Appendix C, clarifies the relation for general L as $R \to \infty$.

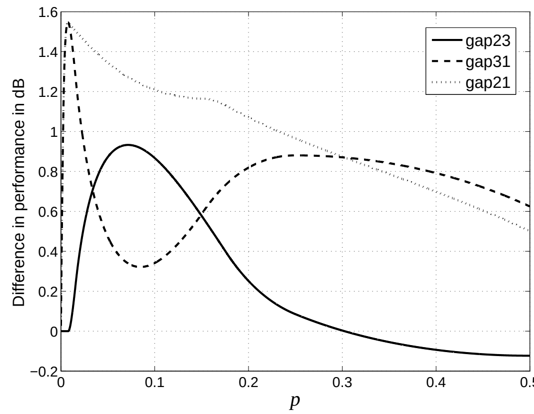

Fig. 3. Value of $\Delta_{i/j}(L,R,p)$ (denoted "gapij" in the legend) versus p, for $L = 3$, and $R = 1$, $(i,j) = (2,1), (2,3), (3,1)$.

Proposition 7. *For any integer $L \geq 3$ denote*

$$\mathcal{H}(L) = \sum_{k=1}^{L}\frac{1}{k},$$

$$\omega(L) = \sum_{k=2}^{L-1}\frac{1}{k}\ln\frac{(L-k)(k+1)}{(k-1)k} + \frac{L-1}{L}\ln(L-1) - \ln\frac{L}{2}.$$

7.a) *For all $L \geq 3$ and $0 < p \leq \frac{L}{L^2-2L+2}$, one has*

$$\lim_{R\to\infty}\Delta_{2/3}(L,R,p) = \frac{10}{\ln 10}\left(\left(\mathcal{H}(L) - 2 + \frac{1}{L}\right)\ln\frac{1}{1-p} + \left(\frac{L+1}{2} - \mathcal{H}(L)\right)\ln\frac{1}{p} - \omega(L)\right). \quad (19)$$

Furthermore, for any $L \geq 3$, one has

$$\lim_{p\to 0}\lim_{R\to\infty}\Delta_{2/3}(L,R,p) = \infty. \quad (20)$$

7.b) *For any $L \geq 3$ and $0 < p \leq \frac{1}{2}$, the following inequality is valid*

$$\lim_{R\to\infty}\Delta_{2/3}(L,R,p) \geq \frac{10}{\ln 10}\sigma(L), \quad (21)$$

where

$$\sigma(L) \triangleq \left(\frac{L-1}{2} - \frac{L-1}{k_1+1}\right)\ln 2 + \ln\frac{L}{2(L-1)} + \frac{1}{k_1+1}\ln\frac{k_1\binom{L-1}{k_1}}{2} - \sum_{k=2}^{k_1}\frac{1}{k}\ln\frac{(L-k)(k+1)}{(k-1)k},$$

and $k_1 = 1$ *if* $L = 3$*, and* $k_1 = \lfloor \frac{L}{4} \rfloor$ *otherwise. Additionally, for all* $0 < p \leq \frac{1}{2}$*, we have*

$$\lim_{L\to\infty} \lim_{R\to\infty} \Delta_{2/3}(L,R,p) = \infty. \quad (22)$$

7.c) *There is some* $L_0 > 3$ *such that for all* $L \geq L_0$ *and* $0 < p \leq \frac{1}{2}$*, one has* $\lim_{R\to\infty} \Delta_{2/3}(L,R,p) > 0$*. Moreover, there is some* $c_0 > 0$ *such that* $\lim_{R\to\infty} \Delta_{3/2}(L,R,p) \leq c_0$ *for all* $L \geq 3$ *and* $0 < p \leq \frac{1}{2}$*.*

Remark 3. a) *It is worth discussing relation (20) in contrast to the result of Corollary 1. Relation (20) says that for every integer* $N > 0$*, there is some* $0 < \epsilon(N) \leq \frac{1}{2}$*, such that* $\Delta_{2/3}(R,L,p) > N$ *for all* $0 < p < \epsilon(N)$ *and* $R > \max\{R_2(L,p,1), R_3(L,p,1)\}$*. This statement does not contradict the conclusion of Corollary 1 that* $\Delta_{2/3}(R,L,p) = 0$ *when* $0 < p < \frac{1}{L(L-1)2^{2LR}+1}$ *because no matter how small p is, the fact that* $R > \max\{R_2(L,p,1), R_3(L,p,1)\}$ *implies that* $p > \frac{1}{L(L-1)2^{2LR}+1}$*.*

b) *Proposition 7 states that in the limit of* $R \to \infty$*, MUEP has superior performance than scenario 3 when* p *is sufficiently small or if* L *is sufficiently large. Additionally, the gain of MUEP versus scenario 3 becomes unbounded for fixed* L *as* p *goes to* 0*, or if* p *is fixed and* L *goes to* ∞*. In order to illustrate how big the gap can be we plot in Fig. 4 the value of* $\lim_{R\to\infty} \Delta_{2/3}(L,R,p)$ *as given in (19) for* $p \in [0.005, 0.1]$ *and* $3 \leq L \leq 10$*.*

c) *It can be shown analytically that the looser lower bound which is developed for* $\sigma(L)$ *in the proof of Proposition 7 is positive for all* $L \geq 64$*. Further numerical evaluation of* $\sigma(L)$ *using MATLAB shows that* $\sigma(L) < 0$ *for* $3 \leq L \leq 6$ *and* $\sigma(L) > 0$ *for* $7 \leq L \leq 63$*. It follows that* $L_0 \leq 7$ *and* $c_0 \leq \max_{3\leq L\leq 6}(-\frac{10}{\ln 10}\sigma(L)) \approx 1.25$*. We conclude that as* $R \to \infty$*, it could be possible for scenario 3 to outperform MUEP for some* $p \in (0, 1/2]$*, only if* $3 \leq L \leq 6$*, and if this happens then the gain is no larger than* 1.25 *dB.*

The next result, which establishes the relation between UEP and scenario 3, follows easily from Propositions 6 and 7.

Corollary 3. *For any* $L \geq 3$*, one has*

$$\lim_{p\to 0} \lim_{R\to\infty} \Delta_{1/3}(L,R,p) = \lim_{L\to\infty} \lim_{R\to\infty} \Delta_{1/3}(L,R,p) = \infty.$$

Furthermore, there is some $L_1 > 3$ *such that for all* $L \geq L_1$ *and* $0 < p \leq \frac{1}{2}$*, relation* $\lim_{R\to\infty} \Delta_{1/3}(L,R,p) > 0$ *holds, and there is some* $c_1 > 0$ *such that* $\lim_{R\to\infty} \Delta_{3/1}(L,R,p) \leq c_1$ *for all* $L \geq 3$ *and* $0 < p \leq \frac{1}{2}$*.*

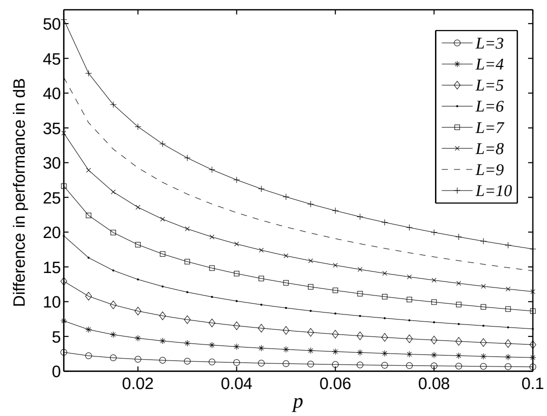

Fig. 4. Asymptotical improvement in performance of scenario 2 versus 3 as $R \to \infty$ for $3 \leq L \leq 10$ and $0.005 \leq p \leq 0.1$.

VII. Asymptotical Comparison between Scenarios 1, 2, 3 and 4, 5

The aim of this section is to compare the asymptotical performance of the three MD schemes based on SRC with MDLVQ and the framework of [43]. For the latter two frameworks the high rate analysis of performance is available and the problem of computing the optimal expected distortion when the description losses are independent was also addressed in previous work. However, in the case of MDLVQ an essential constraint was omitted from the formulation of the optimization problem, namely the condition that the entropy of the central quantizer cannot exceed RL. On the other hand, in the case of [43], the explicit expression of the optimal distortion is not given. Therefore, for both scenarios 4 and 5 we first derive the optimal expected distortion relying on the high rate analysis from prior work.

A. MDLVQ

The existing MDLVQ framework for $L \geq 3$ was introduced in [18] and further investigated in [19]–[21]. The high-rate analysis of performance was carried on in [18] and further refined in [19], [21]. Here we will use the derivation provided in [21], which is more concise in our opinion.

Consider an MDLVQ consisting of a lattice vector quantizer of dimension $n \geq 1$, with a lattice codebook $\Lambda_c \subseteq \mathbb{R}^n$, and an index assignment which maps each central lattice point λ_c to an L-tuple of points from a sublattice Λ_s of Λ_c. Let K denote the index of Λ_s with respect to Λ_c and let ν denote the volume of the fundamental region of Λ_c. We will assume that the MDLVQ is applied to the transform coefficients as in scenarios 1-3. The same assumption as in Section IV holds for the random variable modelling the transform coefficients. Assuming that the rate R of one description grows to ∞, the following relation holds [21, Eq.(44)]

$$(K\nu)^{\frac{2}{n}} = 2^{2(h-R)}, \tag{23}$$

where h is as defined in Section IV. Let $D_4(k, K, \nu)$ denote the average distortion when k descriptions are received. Let us denote by $G(\Lambda_c)$ the normalized second moment of the central lattice, and by $G(S_t)$ the normalized second moment of a sphere in $\mathbb{R}^t$. According to [21], when R and K approach ∞, one has

$$\begin{aligned} D_4(L, K, \nu) &= G(\Lambda_c) 2^{2(h-R)} K^{-\frac{2}{n}}, \\ D_4(k, K, \nu) &= D_4(L, K, \nu) + 2^{2(h-R)} \Big(\frac{G(S_n)}{L^2} + \\ &\quad \frac{L-k}{k} L^{-\frac{L}{L-1}} G(S_{Ln-n}) K^{\frac{2}{(L-1)n}} \Big), \end{aligned}$$

for $1 \le k \le L-1$. We mention that the latter relation follows from [21, Eq.(50),(58),(59)] and (23).

By minimizing the expected distortion (conditioned on one description being received) for fixed R, subject to the condition that the entropy of the central quantizer does not exceed RL, one obtains the optimal value of K as

$$K_{opt} = \min \left(2^{n(L-1)R}, \left(\frac{G(\Lambda_c) L^{\frac{L}{L-1}} (L-1)}{p z(p) G(S_{Ln-n})} \right)^{\frac{n(L-1)}{2L}} \right),$$

where $z(p) \triangleq \frac{1}{1-p^L} \sum_{k=1}^{L-1} \binom{L}{k} (1-p)^k p^{L-1-k} \frac{L-k}{k}$. Let

$$R_4(L, p) \triangleq \frac{1}{2L} \log_2 \frac{G(\Lambda_c) L^{\frac{L}{L-1}} (L-1)}{p z(p) G(S_{Ln-n})}.$$

Then the optimal expected distortion for scenario 4 is

$$\begin{aligned} D_4(L, R, p) &= 2^{2h} \Big(G(\Lambda_c) 2^{-2LR} + p y(p) \frac{G(S_n)}{L^2} 2^{-2R} + \\ &\quad p z(p) G(S_{Ln-n}) L^{-\frac{L-1}{L}} \Big), \text{ if } R < R_4(L, p), \\ D_4(L, R, p) &= 2^{2(h-R)} p^{\frac{L-1}{L}} \left(p^{\frac{1}{L}} y(p) \frac{G(S_n)}{L^2} + \right. \\ &\quad \left. G(\Lambda_c)^{\frac{1}{L}} \Big(\frac{z(p) G(S_{Ln-n})}{L-1} \Big)^{\frac{L-1}{L}} \right) \text{ if } R \ge R_4(L, p), \end{aligned} \tag{24}$$

where $y(p) \triangleq \frac{1}{1-p^L} \sum_{k=1}^{L-1} \binom{L}{k} (1-p)^k p^{L-1-k}$. Notice that

$$\lim_{p \to 0} D_4(L, R, p) = 2^{2h} G(\Lambda_c) 2^{-2LR}. \tag{25}$$

B. Scenario 5

Scenario 5 [43] is an improvement upon [40], which uses staggered scalar quantizers instead of repetition codes. We will assume that the same transform as in scenarios 1-3 is first used to decorrelate the signal, and that the version of the MD scheme developed in [43] for memoryless sources is applied to the transform coefficients. According to [43], under the high resolution assumption, the expected distortion of this scheme (conditioned on at least one description being received) is

$$\tau \Big(\frac{p\gamma(p)}{L} 2^{-2R_1} + \frac{\mu(p)}{L} 2^{-2R_2} \Big),$$

where

$$\gamma(p) \triangleq \frac{L(1-p)}{(1-p^L)} \left(p^{L-2}(L-1) + \sum_{k=2}^{L-1} \frac{(1-p)^{k-1} p^{L-1-k}}{(L-1)^2} \sum_{\ell=1}^{L-k} \binom{L-2-\ell}{k-2} \ell^3 \right),$$

and $\mu(p) \triangleq \frac{L(1-p)}{1-p^L}$. Recall that $\tau = \frac{1}{12} 2^{2h}$. We mention that the rates R_1 and R_2 satisfy the conditions $0 \le R_1 \le R_2$ and $(L-1)R_1 + R_2 = LR$. Then the problem of minimizing the expected distortion has the form of problem P with $L = 2$. It is easy to verify that when $0 < p \le \frac{1}{2}$ the alpha-beta curve of the problem is convex. Thus, by applying Propositions 1 and 2 one obtains the optimal expected distortion as

$$D_5(R, L, p) = \tau \Big(\frac{p\gamma(p)}{L} + \frac{1-p}{1-p^L} 2^{-2RL} \Big), \quad \text{if } R < R_5(L, p),$$

$$D_5(R, L, p) = \tau \Big(\frac{p\gamma(p)}{L-1} \Big)^{\frac{L-1}{L}} \mu(p)^{\frac{1}{L}} 2^{-2R}, \tag{26}$$

$$\text{if } R \ge R_5(L, p), \tag{27}$$

where

$$R_5(L, p) \triangleq \frac{1}{2L} \log_2 \frac{\mu(p)(L-1)}{p\gamma(p)}.$$

Further, notice that

$$\lim_{p \to 0} D_5(R, L, p) = \tau 2^{-2RL}. \tag{28}$$

C. Comparison between Scenarios 1,2,3 and 4,5

We will extend the notation $\Delta_{i/j}(R, L, p)$ to include scenarios 4 and 5 as well, i.e., for $1 \le i, j \le 5$. The results of the asymptotical comparison of the MD schemes based on SRC with scenarios 4 and 5 are summarized next.

Proposition 8. *Let* $L \ge 3$.

8.a) *For any* $R > 0$ *and* $i = 1, 2, 3$, *the following hold:*

$$\lim_{p \to 0} \Delta_{5/i}(R, L, p) = 0,$$

$$\lim_{p \to 0} \Delta_{4/i}(R, L, p) = 10 \log_{10} \frac{1}{12 G(\Lambda_c)}.$$

8.b) *Additionally, one has*

$$\lim_{p \to 0} \lim_{R \to \infty} \Delta_{5/4}(R, L, p) = 10 \log_{10} \frac{12 G(\Lambda_c)^{\frac{1}{L}} \Big((L-1) G(S_{Ln-n}) \Big)^{\frac{L-1}{L}}}{L^{\frac{1}{L}}}, \tag{29}$$

which further implies that

$$10\Big(\log_{10}(L-1)^{\frac{L-3}{L}}-\log_{10}\frac{2\pi e}{12}\Big)<$$
$$\lim_{p\to 0}\lim_{R\to\infty}\Delta_{5/4}(R,L,p)<10\log_{10}L^{\frac{L-2}{L}}, \quad (30)$$

and

$$\lim_{p\to 0}\lim_{R\to\infty}\Delta_{5/4}(R,L,p)>0 \text{ for all } L\geq 4. \quad (31)$$

Furthermore,

$$\lim_{p\to 0}\lim_{R\to\infty}\Delta_{i/j}(R,L,p)=\infty, i=1,2,3, j=4,5. \quad (32)$$

Remark 4. a) *For $L\geq 3$ and fixed R, as p approaches 0, the MDLVQ with $n\geq 2$ has a slight advantage over the other four schemes (because $G(\Lambda_c)\leq\frac{1}{12}$), which is solely due to the use of vector quantization over scalar quantization. On the other hand, as $R\to\infty$, both scenarios 4 and 5 are highly inferior to any of the first three scenarios, for p sufficiently small. This result is not surprising since the latter three schemes have more degrees of freedom than the former two ($L-1$ versus 1).*

b) *We point out that relation (31) is in agreement with the numerical results reported in [43, Fig. 8(b)] for a unit variance Gaussian source, which show that for $L=4$ and $R=5$ scenario 5 strictly outperforms scenario 4 in terms of expected distortion for all $0<p<\frac{1}{10}$.*

Proof of Proposition 8: Claim 8.a) follows from relations (25), (28) and Corollary 1. To prove 8.b) notice first that for fixed L and $p, 0<p\leq\frac{1}{2}$, relations (24) and (27) hold as $R\to\infty$. Therefore, one has

$$\lim_{R\to\infty}10\log_{10}\frac{D_4(R,L,p)}{D_5(R,L,p)}=10\log_{10}\frac{12B(p)}{\left(\frac{\gamma(p)}{L-1}\right)^{\frac{L-1}{L}}\mu(p)^{\frac{1}{L}}},$$

where $B(p)$ denotes the expression enclosed in parentheses in (24). Computing further the limit as $p\to 0$ relation (29) follows. Applying further the inequalities $\frac{1}{12}\geq G(\Lambda_c)$ and $\frac{1}{12}>G(S_{Ln-n})$, relation (30) follows, which in turn, implies (31).

For $0<p\leq\frac{1}{2}$ and $R\geq R_3(L,p,1)$, one has $D_3(R,L,p)=\tau 2^{-2R}L\frac{1-p}{1-p^L}p^{\frac{L-1}{2}}$. Using further (27), one gets

$$\lim_{R\to\infty}\frac{D_5(R,L,p)}{D_3(R,L,p)}=p^{-\frac{(L-1)(L-2)}{2L}}\Big(\frac{\gamma(p)}{(L-1)\mu(p)}\Big)^{\frac{L-1}{L}}.$$

For $L\geq 3$, one has $\lim_{p\to 0}p^{-\frac{(L-1)(L-2)}{2L}}=\infty$, while $\lim_{p\to 0}\frac{\gamma(p)}{\mu(p)}>0$. Thus, (32) holds for $i=3$ and $j=5$. The validity of (32) for the other values of i and j follows using (20), (29), and Corollary 3. ■

VIII. Conclusion

This work compares analytically five representative MD scenarios for practical $L\geq 3$ symmetric descriptions. The first scenario is the PET or UEP scheme, which uses unequal erasure protection via Reed Solomon codes. The second scenario, also known as MUEP, is a recently proposed improvement to the UEP scheme, which uses domain partitioning into L subsets and applies the SRC code to each subset. Additionally, the FEC is applied using permuted RS codes. The third scheme partitions the domain into $L!$ subsets and applies the FEC via repetition codes of various lengths. The existing framework for MD lattice vector quantization and another recently proposed MD scheme are also included in the comparison.

The comparison is performed in terms of the minimum squared error achieved given a constraint on the rate, under the independent description loss assumption. Our analysis shows that in the limit of $R\to\infty$ and $p\to 0$ the gain of the second scenario upon the first one is bounded reaching up to 1.68 dB, and it gradually approaches 0 as $L\to\infty$. Additionally, we prove that UEP and MUEP always outperform scenario 3 as $R\to\infty$ if L is sufficiently large or if p is sufficiently low, and that the difference in performance is unbounded as $L\to\infty$ or as $p\to 0$. Further, our analysis reveals that the first three scenarios are superior to the other two as $R\to\infty$ with gains which become unbounded as $p\to 0$.

On the other hand, we would like to acknowledge that for these results to be useful when designing practical systems, an extensive experimental validations is needed, which is deferred to future work.

Appendix A
Proof of Proposition 2

Proof of Proposition 2: The Lagrangian associated to problem DP is

$$J(R'_1,\cdots,R'_M,\lambda_1,\cdots\lambda_M,\nu)=$$
$$\sigma^2\sum_{k=1}^{M}\beta'(k)2^{-2R'_k}-\sum_{k=1}^{M}\lambda_k R'_k+\nu(\sum_{k=1}^{M}\alpha'(k)R'_k-R).$$

Because the problem is convex, the KKT conditions, which are given below, are necessary and sufficient for optimality.

$$\lambda_k\geq 0,\quad 1\leq k\leq M, \quad (33)$$
$$R'_{k,opt}\geq 0,\quad 1\leq k\leq M, \quad (34)$$
$$\lambda_k R'_{k,opt}=0,\quad 1\leq k\leq M, \quad (35)$$
$$\sum_{k=1}^{M}\alpha'(k)R'_{k,opt}=R, \quad (36)$$
$$-2\tau\beta'(k)2^{-2R'_{k,opt}}\ln 2-\lambda_k+\nu\alpha'(k)=0,\quad 1\leq k\leq M. \quad (37)$$

Note that (37) and (33) imply that

$$\lambda_k=-2\tau\beta'(k)2^{-2R'_{k,opt}}\ln 2+\alpha'(k)\nu\geq 0,$$

which further leads to

$$R'_{k,opt} \geq \frac{1}{2}\log_2 \frac{2\tau(\ln 2)\beta'(k)}{\nu\alpha'(k)},$$

which is satisfied with equality if and only if $\lambda_k = 0$. Corroborating with (34) and the fact that at least one of $R'_{k,opt}$ and λ_k must equal 0 by (35), it follows that

$$R'_{k,opt} = \max\{0, \frac{1}{2}\log_2 \frac{2\tau(\ln 2)\beta'(k)}{\nu\alpha'(k)}\}.$$

Now let us define an integer n_0 as follows. If there is some $k \in \{2, \cdots, M\}$ such that

$$\log_2 \frac{2\tau(\ln 2)\beta'(k)}{\alpha'(k)} \geq \log_2 \nu > \log_2 \frac{2\tau(\ln 2)\beta'(k-1)}{\nu\alpha'(k-1)}, \tag{38}$$

then $n_0 \triangleq k$. Notice that if such a k exists then it is unique by (6). If such a k does not exist, but the first inequality in (38) holds for $k = 1$ then $n_0 \triangleq 1$. Finally, if the second inequality in (38) holds for $k = M+1$ then $n_0 \triangleq M+1$. It further follows that

$$R'_{k,opt} = \begin{cases} 0, & 1 \leq k \leq n_0 - 1 \\ \frac{1}{2}\log_2 \frac{2\tau(\ln 2)\beta'(k)}{\nu\alpha'(k)}, & n_0 \leq k \leq M \end{cases}. \tag{39}$$

Substituting (39) in (36) and solving for ν leads to $n_0 < M+1$ (since $R > 0$) and

$$\log_2 \nu = \log_2(2\tau \ln 2) - \frac{2R}{S(n_0)} + \frac{1}{S(n_0)} \sum_{k=n_0}^{M} \alpha'(k) \log_2 \frac{\beta'(k)}{\alpha'(k)}. \tag{40}$$

Using (7) it can be shown that

$$\sum_{k=n_0}^{M} \alpha'(k) \log_2 \frac{\beta'(k)}{\alpha'(k)} = S(n_0) \log_2 \frac{\beta'(n_0)}{\alpha'(n_0)} + 2R(n_0).$$

Substituting the above in (40) leads to

$$\log_2 \nu = \frac{2(R(n_0) - R)}{S(n_0)} + \log_2 \frac{2\tau(\ln 2)\beta'(n_0)}{\alpha'(n_0)}. \tag{41}$$

Replacing $\log_2 \nu$ from the above equality in (38) when n_0 equals k satisfying (38), leads after some algebra to

$$R(n_0) \leq R < R(n_0 - 1).$$

The above relation also follows when $n_0 = 1$. Thus, it follows that n_0 equals the integer k_0 defined in the statement of Proposition 2. Using this observation and replacing (41) in (39), relation (8) follows. Relation (9) is further obtained by replacing $R'_{k,opt}$ from (8) in the objective function. ■

APPENDIX B
PROOF OF RESULTS OF SECTION V

Proof of Proposition 3: The case $L = 2$ is straightforward. Assume now that $L \geq 3$. Let $0 \leq k \leq L-3$ and let us analyze when the inequality $slope(P_k P_{k+1}) \leq slope(P_{k+1} P_{k+2})$ holds. This inequality is equivalent to $(k+1)(k+2)U(k+1) \leq (k+2)(k+3)U(k+2)$, which, by (10), after straightforward algebraic manipulations, becomes equivalent to $f_L(k) \leq 0$, where

$$f_L(k) \triangleq k^2 + k\big(4 - p - (1-p)L\big) + 3 - p - 3(1-p)L. \tag{42}$$

Recall that $0 < p \leq 1/2$. Then it can be easily shown that $f_L((1-p)L - 1) < 0$, $f_L((1-p)(L-1)) > 0$ and $f_L(0) < 0$. Using further the facts that $f_L(k)$ is convex and that $(1-p)L-1 > 0$, it follows that there is a real value x_0 such that $(1-p)L-1 < x_0 < (1-p)(L-1)$, $f_L(x_0) = 0$, and $f_L(k) \leq 0$ for all integers $k \in \{0, \cdots, \lfloor x_0 \rfloor\}$, while $f_L(k) > 0$ for all integers $k \in \{\lfloor x_0 \rfloor + 1, \cdots, L-3\}$. This further implies that the portion of the alpha-beta curve between the points P_0 and $P_{\min\{L-1, \lfloor x_0 \rfloor + 2\}}$ is convex, while the remaining portion up to P_{L-1} (if non empty) is concave.

By using straightforward algebra it can be shown that relation $slope(P_{L-2} P_{L-1}) \leq slope(P_{L-1} P_L)$ is equivalent to $p \leq \frac{1}{L^2 - L + 1}$. Additionally, notice that $\frac{1}{L^2-L+1} \leq \frac{1}{L}$, while the inequality $p \leq \frac{1}{L}$ implies that $(1-p)L - 1 \geq L-2$, and further that $x_0 > L-2$. We conclude that when $p \leq \frac{1}{L^2-L+1}$ the alpha-beta curve is convex.

On the other hand, when $p > \frac{1}{L^2-L+1}$ the alpha-beta curve is non convex since at least P_{L-1} is not on its lower convex hull. However, the portion of the curve between the points P_0 and $P_{\min\{L-1, \lfloor x_0 \rfloor + 2\}}$ is convex, while the remaining portion is concave.

Notice now that for $S < L$ relation (11) is equivalent to

$$slope(P_{S-1} P_S) > slope(P_{S-1} P_L). \tag{43}$$

Thus, the smallest integer $S \in \{1, \cdots, L-1\}$ satisfying (11), assuming that such an integer exists, obeys the conditions $slope(P_{S-1} P_S) > slope(P_{S-1} P_L) \geq slope(P_{S-2} P_{S-1})$, where, by convention, $slope(P_{-1} P_0) = 0$. Then, according to the properties of the alpha-beta curve highlighted above, it follows that its lower convex hull satisfies the claim made in Proposition 3. On the other hand, when (43) is violated for all $1 \leq S \leq L-1$, the alpha-beta curve must be convex and the only integer satisfying (11) is $S = L$ (when $S = L$ the right hand side of (11) is 0), in which case the claim also holds.

It follows that problem DP(1) has $M = S$ and by applying (7) relation (12) follows. ■

Proof of Proposition 4: The case $L = 2$ is trivial. Assume now that $L \geq 3$. For $0 \leq k \leq L-3$, the inequality $slope(P_kP_{k+1}) \le slope(P_{k+1}P_{k+2})$ is equivalent to $f_{L-1}(k) \le 0$, where $f_{L-1}(k)$ is obtained by replacing L by $L-1$ in (42). Following similar steps as in the proof of Proposition 3 it follows that there is a real value x_1 such that $(1-p)(L-1)-1 < x_1 < (1-p)(L-2)$, $f_{L-1}(x_1)=0$, and $f_{L-1}(k) \le 0$ for all integers $k \in \{0,\cdots,\lfloor x_1 \rfloor\}$, while $f_{L-1}(k) > 0$ for all integers $k \in \{\lfloor x_1 \rfloor+1,\cdots,L-3\}$. This implies that the portion of the alpha-beta curve between the points P_0 and $P_{\min\{L-1,\lfloor x_1\rfloor+2\}}$ is convex, while the remaining portion up to P_{L-1} (if non empty) is concave. Moreover, notice that the whole portion between the points P_0 and P_{L-1} is convex if and only if $x_1 \ge L-3$, which is further equivalent to $f_{L-1}(L-3) \le 0$, which is in turn equivalent to $p \le \frac{L}{L^2-2L+2}$.

Next we will show that the inequality $slope(P_{L-2}P_{L-1}) \le slope(P_{L-1}P_L)$ holds for all $L \ge 3$ and $0 < p \le 1$. After some algebra the above relation becomes equivalent to $h(p) \le 0$, where $h(p) = (1-p)^{L-2}p - \frac{1}{L-1}$. A simple analysis of the sign of the derivative $h'(p)$ implies that $\frac{1}{L-1}$ is the unique point of maximum of the function $h(\cdot)$ on the interval $(0,1]$. Since $h(\frac{1}{L-1}) < 0$ the claim follows.

We conclude that for $L \ge 3$ and $p \le \frac{L}{L^2-2L+2}$, the alpha-beta curve is convex and applying Proposition 2 to problem DP(2) yields (13).

When $L \ge 3$ and $\frac{L}{L^2-2L+2} < p \le \frac{1}{2}$ the alpha-beta curve is non convex because at least the point P_{L-2} is not an extreme point on the lower convex hull. Then there must exist an integer $T \in \{1,\cdots,L-2\}$ satisfying the relation $slope(P_{T-1}P_T) > slope(P_{T-1}P_{L-1})$, which is equivalent to (14). Considering the smallest T with this property and using an argument as in the proof of Proposition 3, it follows that the lower convex hull of the portion of the alpha-beta curve up to the point P_{L-1} is the union of segment lines P_kP_{k+1}, $0 \le k \le T-2$, and $P_{T-1}P_{L-1}$.

To complete the proof of the claim it is sufficient to show that $slope(P_{T-1}P_{L-1}) \le slope(P_{L-1}P_L)$. This inequality is equivalent to $\sum_{k=T}^{L-1} V(k) \le \frac{L-T}{T}V(L)$. Since $V(L) = \frac{1-p}{1-p^L} > 1-p$, while $\sum_{k=T}^{L-1} V(k) \le 1-V(L) < p$, it is sufficient to prove that $p \le \frac{L-T}{T}(1-p)$, which is equivalent to $T \le L(1-p)$. When $T=1$ the previous inequality clearly holds. Let us consider now the case $T > 1$. The definition of T implies that $slope(P_{T-2}P_{T-1}) \le slope(P_{T-1}P_T)$, fact which implies that $T \le \lfloor x_1 \rfloor + 2 < (1-p)(L-2)+2 = (1-p)L+2p \le (1-p)L+1$, thus $T \le (1-p)L$. Thus, the proof of the claim on the convex hull is complete.

When $L=2$, the alpha-beta curve is clearly convex for all $0 < p \le 1/2$. Because in this case $\frac{L}{L^2-2L+2} = 1$ the statement of the proposition holds, too. ∎

Appendix C
Proof of Claims of Section VI

Proof of Proposition 6: Relation (16) follows from (15) and Propositions 3 and 4, leading immediately to (17). Further, in order to prove (18) we need to show that $\lim_{L\to\infty} \frac{A(L)}{\frac{\ln L}{L}} = 1$. Notice first that by the mean value theorem there are $\mu_k \in (L-k, L+1-k)$ such that

$$\ln \frac{L+1-k}{L-k} = \ln(L+1-k) - \ln(L-k) = \frac{1}{\mu_k},$$

for $1 \le k \le L-1$. Using further the inequalities $\frac{1}{L-k+1} < \frac{1}{\mu_k} < \frac{1}{L-k}$ and equalities $\frac{1}{k}\frac{1}{L-k+1} = \frac{1}{L+1}(\frac{1}{k} + \frac{1}{L-k+1})$ and $\frac{1}{k}\frac{1}{L-k} = \frac{1}{L}(\frac{1}{k}+\frac{1}{L-k})$ for $1 \le k \le L-1$, one obtains

$$\frac{1}{L+1}(\mathcal{H}(L-1)+\mathcal{H}(L)-1) - \frac{\ln L}{L} < A(L) < \frac{2}{L}\mathcal{H}(L-1) - \frac{\ln L}{L}, \tag{44}$$

where $\mathcal{H}(n) \triangleq \sum_{i=1}^{n}\frac{1}{i}$, for any integer $n > 0$. Next we apply the well known relations $\ln(n+1) < \mathcal{H}(n) < \ln n + 1 < \ln(n+1)+1$ to (44) and obtain

$$\frac{2\ln L - 1}{L+1} - \frac{\ln L}{L} < A(L) < \frac{2}{L}(\ln L + 1) - \frac{\ln L}{L}.$$

Using further the fact that

$$\lim_{L\to\infty} \frac{\frac{2\ln L-1}{L+1} - \frac{\ln L}{L}}{\frac{\ln L}{L}} = \lim_{L\to\infty} \frac{\frac{2}{L}(\ln L+1) - \frac{\ln L}{L}}{\frac{\ln L}{L}} = 1.$$

relation (18) follows. Finally, the last claim follows from Theorem 4.1 in [38]. ∎

Proof of Proposition 7: Using (15) and replacing $R_2(L,p,1)$ from (13) (since $p \le \frac{L}{L^2-2L+2}$) and $R_3(L,p,1)$ from Proposition 5, equation (19) follows. Further, it can be easily proved by induction that the coefficient of $\ln\frac{1}{p}$ in the right hand side of (19) is positive for $L \ge 3$. The fact that $\lim_{p\to 0}\ln\frac{1}{p} = \infty$ and $\lim_{p\to 0}\ln\frac{1}{1-p} = 0$ imply relation (20).

In order to prove claim 7.b) we will establish an upper bound for $D_2(L,R,p)$ for all L, R and $p \in (0, 1/2]$. For this, consider problem P'(2) which is obtained from problem P by replacing L by k_1+1 and letting $\alpha(k) = \frac{1}{k(k+1)}$, $\beta(k) = V(k)$, for $1 \le k \le k_1$, and $\alpha(k_1+1) = \frac{1}{k_1+1}$, $\beta(k_1+1) = \sum_{k=k_1+1}^{L} V(k)$. Clearly, any feasible solution $(R_1,\cdots,R_{k_1+1})$ of problem P'(2) can be extended to a feasible solution $(R_1,\cdots,R_L)$ to problem P(2) by letting $R_k = R_{k_1+1}$ for all $k_1+2 \le k \le L$. Therefore, the optimal value of the cost function of P'(2) is an upper bound for $D_2(L,R,p)$ for all R, p, L.

To determine the solution of problem P'(2) we will first show that its alpha-beta curve is convex. Consider first $k_1 \ge 2$ and recall that in the proof of Proposition 4 it was established that relation $\frac{V(k+1)}{\frac{1}{(k+1)(k+2)}} \le \frac{V(k+2)}{\frac{1}{(k+2)(k+3)}}$ holds for $0 \le k \le (1-p)(L-1)-1$. These inequalities hold for $0 \le k \le k_1 - 2$ (because $p \le \frac{1}{2}$). It follows that the portion of the alpha-beta curve up to point P_{k_1} is convex. It remains to show now that $\frac{V(k_1)}{\frac{1}{k_1(k_1+1)}} \le \frac{\sum_{k=k_1+1}^{L} V(k)}{\frac{1}{k_1+1}}$, which is equivalent to

$$k_1 V(k_1) \le \sum_{k=k_1+1}^{L} V(k). \tag{45}$$

Notice that relation $V(k) \le V(k+1)$ is equivalent to $k \le L(1-p)-1$, and, consequently, it is true for $k \le \lfloor \frac{L+1}{2} \rfloor - 2$. Since $k_1 - 1 \le \lfloor \frac{L+1}{2} \rfloor - 1 - k_1$ (due to $k_1 = \lfloor \frac{L}{4} \rfloor$), it further follows that $(k_1-1)V(k_1) \le \sum_{k=k_1+1}^{\lfloor \frac{L+1}{2} \rfloor - 1} V(k)$. Finally, corroborating with the fact that $V(k_1) \le \sum_{k=1}^{L-1} V(k) = \frac{p}{1-p^L} \le V(L)$, relation (45) follows.

When $k_1 = 1$ in order to establish the convexity of the alpha-beta curve of problem P'(2) one only needs to prove (45), which holds in view of the inequality $V(1) \le V(L)$.

Using Propositions 1 and 2 to solve problem P'(2), and since its optimal cost is an upper bound for $D_2(L,R,p)$, one obtains

$$D_2(L,R,p) \le \frac{2\tau(L-1)(1-p)p^{L-1}}{1-p^L} 2^{-2(R-R_2^*(L,p))},$$

for all $L \ge 3$, $0 < p \le \frac{1}{2}$, and $R > R_2^*(L,p)$, where $R_2^*(L,p)$ is defined as $R(1)$ in (7) for problem P'(2), i.e.,

$$R_2^*(L,p) = \frac{1}{2}\left(\sum_{k=2}^{k_1} \frac{1}{k} \log_2 \frac{V(k)(k+1)}{V(k-1)(k-1)} + \frac{1}{k_1+1} \log_2 \frac{\sum_{k=k_1+1}^{L} V(k)}{k_1 V(k_1)} \right).$$

Using further Corollary 2 it follows that

$$\lim_{R\to\infty} \Delta_{2/3}(L,R,p) \ge \frac{10}{\ln 10}\Big(\ln \frac{L}{2(L-1)} + 2(R_3(L,p,1) - R_2^*(L,p)) \ln 2 \Big)$$

for any $L \ge 3$ and $0 < p \le \frac{1}{2}$. Corroborating with the fact that $\sum_{k=k_1+1}^{L-1} V(k) \le \sum_{k=1}^{L-1} V(k) = \frac{p}{1-p^L} < V(L)$, after some algebra, one obtains

$$\lim_{R\to\infty} \Delta_{2/3}(L,R,p) \ge \frac{10}{\ln 10}\Bigg(\ln \frac{L}{2(L-1)} + \frac{1}{k_1+1} \ln \frac{k_1\binom{L-1}{k_1}}{2} - \sum_{k=2}^{k_1} \frac{1}{k} \ln \frac{(L-k)(k+1)}{k(k-1)} + \Big(\frac{2k_1}{k_1+1} - \mathcal{H}(k_1) \Big) \ln(1-p) + \Big(\mathcal{H}(k_1) + \frac{L-2k_1-1}{k_1+1} - \frac{L-1}{2} \Big) \ln p \Bigg).$$

Let $\phi(p)$ denote the expression on the right hand side of the above inequality, for $0 < p \le \frac{1}{2}$. We will show that $\phi(p)$ achieves its minimum over the interval $(0, \frac{1}{2}]$ in $p = \frac{1}{2}$. This statement trivially holds for $k_1 = 1$ since the coefficients of $\ln(1-p)$ and of $\ln p$ are 0. Assume now that $k_1 \ge 2$, i.e., $L \ge 8$. Let a denote the coefficient of $\ln p$, and b denote the coefficient of $\ln(1-p)$ in $\phi(p)$. Notice that it can be proved easily by induction that $a \le b < 0$. The derivative of $\phi(p)$ with respect to p is $\phi'(p) = \frac{10}{\ln 10} \frac{a-(a+b)p}{p(1-p)}$ and its unique zero point is $p_0 = \frac{a}{a+b} \ge \frac{1}{2}$. It follows that $\phi'(p) < 0$ for $0 < p \le \frac{1}{2}$, fact which implies that $\phi(p)$ is decreasing over the interval $(0, \frac{1}{2}]$. Thus, $\phi(p) \ge \phi(\frac{1}{2}) = \frac{10}{\ln 10}\sigma(L)$ leading to (21).

In order to prove relation (22) it is sufficient to show that $\lim_{L\to\infty} \sigma(L) = \infty$. To this end we will first find a simpler lower bound for $\sigma(L)$. Notice that

$$\sum_{k=2}^{k_1} \frac{1}{k} \ln \frac{(L-k)(k+1)}{(k-1)k} = \sum_{k=2}^{k_1} \frac{1}{k} \ln \frac{L-k}{k-1} + \sum_{k=2}^{k_1} \frac{1}{k} \ln \frac{k+1}{k} \le \ln L \sum_{k=2}^{k_1} \frac{1}{k} + \sum_{k=2}^{k_1} \frac{1}{k(k-1)} \le (\ln L - 2\ln 2)\ln L + 1,$$

where the first inequality follows from $\ln \frac{L-k}{k-1} \le \ln L$ and $\ln \frac{k+1}{k} = \ln(k+1) - \ln k = \frac{1}{\nu_k} < \frac{1}{k-1}$, for some $\nu_k \in (k, k+1)$ by the mean value theorem. Meanwhile, the second inequality follows from $\mathcal{H}(k_1) \le \ln k_1 + 1 \le \ln(L/4) + 1$ and from $\sum_{k=2}^{k_1} \frac{1}{k(k-1)} = 1 - \frac{1}{k_1} < 1$. Using further the fact that $\ln \frac{L}{2(L-1)} \ge \ln \frac{1}{2}$, $\frac{1}{k_1+1} \ln \frac{k_1\binom{L-1}{k_1}}{2} \ge 0$, and that $\frac{L-1}{2} - \frac{L-1}{k_1+1} \le \frac{L-1}{2} - 4$ when $k_1 \le 2$, it follows that $\sigma(L) \ge (\frac{L-11}{2} + 2\ln L)\ln 2 - \ln^2 L - 1$ for $k_1 \le 2$. Since $\lim_{L\to\infty}(\frac{L-11}{2} + 2\ln L)\ln 2 - \ln^2 L - 1 = \infty$ the claim follows.

Finally, claim 2.c) follows from (21) and from the fact that $\Delta_{3/2}(L,R,p) = -\Delta_{2/3}(L,R,p)$. ■

References

[1] V. A. Vaishampayan, "Design of multiple-description scalar quantizers", *IEEE Trans. Inform. Th.*, vol. 39, no. 3, pp. 821–834, May 1993.

[2] C. Tian and S. S. Hemami, "Universal multiple description scalar quantization: analysis and design", *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 2089–2102, Sept. 2004.

[3] S. Dumitrescu and X. Wu, "Optimal two-description scalar quantizer design", *Algorithmica*, vol. 41, no. 4, pp. 269-287, Feb. 2005.

[4] S. Dumitrescu and X. Wu, "Lagrangian optimization of two-description scalar quantizers", *IEEE Trans. Inform. Theory*, vol. 53, no. 11, pp. 3990–4012, Nov. 2007.

[5] D. Muresan and M. Effros, "Quantization as histogram segmentation: optimal scalar quantizer design in network systems", *IEEE Trans. Inform. Theory*, vol. 54, no. 1, pp. 344–366, Jan. 2008.

[6] M. Fleming, Q. Zhao, and M. Effros, "Network vector quantization", *IEEE Trans. Inform. Theory*, vol. 50, no. 8, pp. 1584–1604, Aug. 2004.

[7] V. A. Vaishampayan, N. Sloane, and S. Servetto, "Multiple description vector quantization with lattice codebooks: design and analysis," *IEEE Trans. Inform. Theory*, vol. 47, no. 5, pp. 1718–1734, Jul. 2001.

[8] Y. Wang, M. T. Orchard, V. A. Vaishampayan, and A. R. Reibman, "Multiple description coding using pairwise correlating transforms," *IEEE Trans. Image Proc.*, vol. 10, no. 3, pp. 351–366, Mar. 2001.

[9] V. K Goyal and J. Kovacevic, "Generalized multiple description coding with correlating transforms," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2199–2224, Sept. 2001.

[10] W. Jiang and A. Ortega, "Multiple description coding via polyphase transform and selective quantization", *Proc. SPIE: Visual Comm. and Image Proc.*, Jan. 1999.

[11] K. P. Subbalakshmi and S. Somasundaram, "Multiple description image coding framework for EBCOT", *IEEE Intern. Conf. Image Proc.*, Sept. 2002, Rochester, NY.

[12] I. V. Bajic and J. W. Woods, "Domain-based multiple description coding of images and video," *IEEE Trans. Image Proc.*, vol. 12, no. 10, pp. 1211-1225, Oct. 2003.

[13] T. Tillo, M. Grangetto, and G. Olmo, "Multiple description image coding based on Lagrangian rate allocation", *IEEE Trans. Image Proc.*, vol. 16, no. 3, pp. 673–683, Mar. 2007.

[14] Y. Zhang, S. Dumitrescu, J. Chen, and Z. Sun, " LDGM-based multiple description coding for finite alphabet sources", *IEEE Trans. Commun.*, vol. 60, no. 12, pp. 3671–3682, Dec. 2012.

[15] T. Y. Berger-Wolf and E. M. Reingold, "Index assignment for multichannel communication under failure", *IEEE Trans. Inform. Theory*, vol. 48, no. 10, pp. 2656-2668, Oct. 2002.

[16] S. Dumitrescu and X. Wu, "On properties of locally optimal multiple description scalar quantizers with convex cells", *IEEE Trans. Inform. Theory*, vol. 55, no. 12, pp. 5591–5606, Dec. 2009.

[17] C. Tian and S. S. Hemami, "Sequential design of multiple description scalar quantizers", *Proc. IEEE Data Compression Conf.*, Snowbird, UT, pp. 32–41, Mar. 2004.

[18] J. Ostergaard, J. Jensen, and R. Heusdens, "n-channel entropy-constrained multiple description lattice vector quantization", *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 1956–1973, May 2006.

[19] X. Huang and X. Wu, "Optimal index assignment for multiple description lattice vector quantization," *Proc. IEEE Data Compression Conf.*, pp. 272–281, Snowbird, UT, Mar. 2006.

[20] M. Liu and C. Zhu, "M-description lattice vector quantization: index assignment and analysis", *IEEE Trans. Signal Proc.*, vol. 57, no. 6, pp. 2258–2274, Jun. 2009.

[21] G. Zhang, J. Ostergaard, J. Klejsa, and W. B. Kleijn, "High-rate analysis of symmetric L-channel multiple description coding", *IEEE Trans. Commun.*, vol. 59, no. 7, pp. 1846–1856, Jul. 2011.

[22] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Trans. Inform. Theory*, vol. IT-33, no. 4, pp. 502-521, Jul. 1987.

[23] J. Chen, Y. Zhang, and S. Dumitrescu, "Gaussian multiple description coding with low-density generator matrix codes", *IEEE Trans. Commun.*, vol. 60, no. 3, pp. 676–687, Mar. 2012.

[24] G. Davis and J. Danskin, "Joint source and channel coding for image transmission over lossy packet networks", *Proc. SPIE Conf. Wavelet Applications of Digital Image Proc. XIX* , pp. 376–387, Denver, August 1996.

[25] R. Puri and K. Ramchandran, "Multiple description source coding through forward error correction codes," *Proc. 33rd Asilomar Conf. Signals, Systems, and Computers*, vol. 1, pp. 342–346, Oct. 1999.

[26] A. E. Mohr, E. A. Riskin, and R. E. Ladner, "Unequal loss protection: graceful degradation over packet erasure channels through forward error correction", *IEEE J. Selected Areas Commun.*, vol. 18, no. 7, pp. 819-828, Jun. 2000.

[27] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoding transmission", *IEEE Trans. Inform. Theory*, vol. 42, no. 6, pp. 1737–1744, Nov. 1996.

[28] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees", *IEEE Trans. Circuits Systems Video Tech.*, vol. 6, no. 3, pp. 243–250, Jun. 1996.

[29] D. Taubman, "High performance scalable image compression with EBCOT", *IEEE Trans. Image Proc.*, vol. 9, no. 7, pp. 1158–1170, Jul. 2000.

[30] B.-J. Kim, Z. Xiong, and W. A. Pearlman, "Very low bit-rate embedded video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT)", *IEEE Trans. Circuits Systems Video Tech.*, vol. 10, no. 8, pp. 1374-1385, Dec. 2000.

[31] Heiko Schwarz, Detlev Marpe, and Thomas Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Systems Video Tech., Special Issue on Scalable Video Coding*, vol. 17, no. 9, pp. 1103–1120, Sept. 2007.

[32] D. G. Sachs, R. Anand, and K. Ramchandran, "Wireless image transmission using multiple-description based concatenated codes", *Proc. SPIE 2000*, vol. 3974, pp. 300–11, Jan. 2000.

[33] T. Stockhammer and C. Buchner, "Progressive texture video streaming for lossy packet networks", *Proc. 11th Intern. Packet Video Workshop*, Kyongju, May 2001.

[34] S. Dumitrescu, X. Wu, and Z. Wang, "Globally optimal uneven error-protected packetization of scalable code streams," *IEEE Trans. Multimedia*, vol. 6, no. 2, pp. 230–239, Apr. 2004.

[35] V. Stankovic, R. Hamzaoui, and Z. Xiong, "Efficient channel code rate selection algorithms for forward error correction of packetized multimedia bitstreams in varying channels," *IEEE Trans. Multimedia*, vol. 6, no. 2, pp. 240–248, Apr. 2004.

[36] J. Thie and D. Taubman, "Optimal erasure protection strategy for scalably compressed data with tree-structured dependencies," *IEEE Trans. Image Proc.*, vol. 14, no. 12, pp. 2002–2011, Dec. 2005.

[37] S. Dumitrescu, X. Wu, and Z. Wang, "Efficient algorithms for optimal uneven protection of single and multiple scalable code streams against packet erasures," *IEEE Trans. Multimedia*, vol. 9, no. 7, pp. 1466–1474, Nov. 2007.

[38] C. Tian, S. Mohajer, and S. N. Diggavi, "Approximating the Gaussian multiple description rate region under symmetric distortion constraints", *IEEE Trans. Inform. Theory*, vol. 55, no. 8, pp. 3869–3891, Aug. 2009.

[39] S. Dumitrescu, G. Rivers, and S. Shirani, "Unequal erasure protection technique for scalable multi-streams", *IEEE Trans. Image Proc.*, vol. 19, no. 2, pp. 422-434, Feb. 2010.

[40] E. Baccaglini, T. Tillo, and G. Olmo, "A flexible R-D based multiple description scheme for JPEG 2000", *IEEE Signal Proc. Letters*, vol. 14, no. 3, pp. 197–200, Mar. 2007.

[41] E. Akyol, A. M. Tekalp, and M. R. Civanlar, "A flexible multiple description coding framework for adaptive peer-to-peer video streaming", *IEEE J. Selected Topics Signal Proc.*, vol. 1, no. 2, pp. 231–245, Aug. 2007.

[42] T. Tillo, E. Baccaglini, and G. Olmo, "Multiple descriptions based on multirate coding for JPEG 2000 and H.264/AVC", *IEEE Trans. Image Proc.*, vol. 19, no. 7, pp. 1756–1767, July 2010.

[43] U. Samarawickrama, J. Liang, and C. Tian, "M-channel multiple description coding with two-rate coding and staggered quantization", *IEEE Trans. Circuits Systems Video Tech.*, vol. 20, no. 7, pp. 933–944, Jul. 2010.

[44] S. Dumitrescu and T. Zheng, "Improved multiple description framework based on successively refinable quantization and uneven erasure protection", *Proc. IEEE Data Compression Conf.*, pp. 514-514, Mar. 2008, Snowbird, UT.

[45] S. Dumitrescu and X. Wu, "Optimal multiresolution quantization for scalable multimedia coding", *Proc. IEEE Inform. Theory Workshop*, pp. 139–142, Bangalore, Oct. 2002.

[46] H. Gish and J. N. Pierce, "Asymptotically efficient quantizing", *IEEE Trans. Inform. Theory*, vol. IT-14, no.5, pp. 676–683, Sept. 1968.

PLACE PHOTO HERE

Sorina Dumitrescu (M'05-SM'13) received the B.Sc. and Ph.D. degrees in mathematics from the University of Bucharest, Romania, in 1990 and 1997, respectively. From 2000 to 2002 she was a Postdoctoral Fellow in the Department of Computer Science at the University of Western Ontario, London, Canada. Since 2002 she has been with the Department of Electrical and Computer Engineering at McMaster University, Hamilton, Canada, where she held Postdoctoral, Research Associate, and Assistant Professor positions, and where she is currently an Associate Professor. Her current research interests include multimedia coding and communications, network-aware data compression, multiple description codes, joint source-channel coding, signal quantization. Her earlier research interests were in formal languages and automata theory. Dr. Dumitrescu held an NSERC University Faculty Award during 2007-2012.

PLACE
PHOTO
HERE

Ting Zheng received her B.Eng degree in electronic information engineering from Beijing University of Technology, Beijing, China, in 2006, and her MASc. in electrical and computer engineering from McMaster University, Ontario, Canada, in 2008. She is currently working at IBM China in Beijing.