New Coding Schemes for the Symmetric K-Description Problem

Chao Tian, *Member, IEEE*, and Jun Chen, *Member, IEEE*

Abstract—We propose novel coding schemes for the K-description problem with symmetric rates and symmetric distortion constraints. There are two main new ingredients in these schemes: the first one is akin to the method seen in the well-known butterfly network of network coding literature, and systematic erasure channel codes are applied on certain carefully chosen source coding component; the second approach is built on the quantization splitting technique which was previously proven useful in the Gaussian CEO problem. We first focus on a special case of the three description problem, where any two descriptions are rate-distortion optimal jointly, referred to as the *no two description excess rate* case. For this special case and the quadratic Gaussian source, we show that the two aforementioned approaches lead to rate-distortion points outside the achievable region based on the source-channel erasure codes, previously proposed by Pradhan, Puri, and Ramchandran. Interestingly, though only the symmetric problem is considered in our work, the proposed schemes in fact benefit from time-sharing several asymmetric rate-distortion points. The insights gained through the no two description excess rate case lead to strategic combination of the new ingredients with the existing coding scheme, yielding new coding schemes for the symmetric K-description problem.

Index Terms—Multiple descriptions, rate-distortion.

I. INTRODUCTION

THE multiple description (MD) source coding problem is well known as a long-standing open problem in information theory. The problem setting is that a source needs to be encoded into several descriptions, such that any subset of these descriptions can be used to reconstruct the source with certain fidelity (see Fig. 1). The problem is practically motivated by applications in efficient distributed storage systems, or communication systems over packet-loss network (see the review article [1] for more detailed discussions). It is easily seen that even for the two description case, a conflict exists that if the individual descriptions are good, then they must be similar to the source

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

J. Chen is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada (e-mail: junchen@ece.mcmaster.ca).
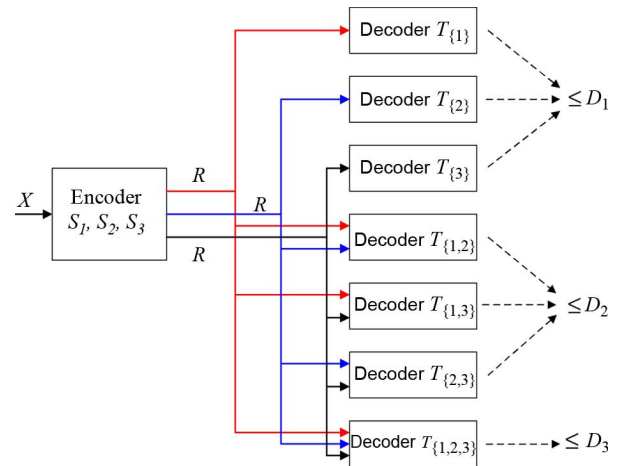
Fig. 1. System diagram for symmetric three description coding.

and thus quite alike, which reduces the efficiency when they are taken jointly.

For the two description problem (i.e., $K = 2$), the only case for which the rate-distortion region has been completely characterized is the quadratic Gaussian problem [2], [3]. Several achievable regions exist [2], [4] for general sources, yet a complete characterization has not been found. Recent effort on the MD problem focuses more on the general K-description problem. An achievable rate region under general distortion constraints was given in [5] based on conditional codebooks. Since many modern packet networks have the property that packets are treated without priority, the symmetric MD problem has received more attention. For this case, the distortion constraints depend only on the number of descriptions available, but not the specific combination of the descriptions; at the same time, the rates of the descriptions are kept the same. An achievable symmetric rate under such symmetric distortion constraints was given in [6] and [7], which utilizes the binning technique often seen in distributed source coding problems. A rather descriptive name was coined for these codes in [6] and [7]—(n, k) source-channel erasure codes (SCEC)—because of their similarity to (n, k) maximum distance separable (MDS) codes. We will sometimes refer to these codes simply as SCEC when the exact values of (n, k) are of no importance; moreover, we will refer to the multilayer coding scheme based on SCEC, which was proposed by Puri, Pradhan, and Ramchandran in [7], as the PPR multilayer scheme.

Wang and Viswanath [8], [9] made important progress on the K-description problem by considering the quadratic vector Gaussian source problem. Applying their result to the symmetric scalar quadratic Gaussian problem shows that the PPR multilayer scheme (in fact a single layer SCEC) is sum-rate

optimal when only two levels of distortion constraints exist and one of them is on the distortion of having the complete set of descriptions. A new sum-rate lower bound was recently discovered by Tian *et al.* [10] with any levels of symmetric distortion constraints, through which it was shown that the (symmetric) individual description rate loss by the PPR multilayer scheme is no larger than 0.92 bit for the Gaussian source. With these results, one may conjecture that the PPR multilayer scheme is indeed sum-rate optimal with any levels of distortion constraints for the Gaussian source. However, as we shall show in this work, this conjecture is in fact not true, and other rate-distortion (R-D) points can be achieved outside the R-D region achievable by the PPR multilayer scheme.

The following simple question motivates our work. It was shown in [6] that for the Gaussian source and a total of three descriptions, using SCEC with a Gaussian codebook, any two descriptions can be made R-D optimal jointly (this requirement will be referred to as *no two description excess rate*), and having all three descriptions can reduce the distortion even further. However, at this R-D point (referred to as the *binning point* due to the coding scheme based on the well-known "binning" technique), we observe that each description by itself is useless, because the bin index alone offers no possible distortion reduction.[1] Our question is the following: is there a coding scheme satisfying the no two description excess rate requirement, for which individual descriptions are still useful?

The answer to this question is in fact positive, and we show that this can be accomplished using two distinct approaches. The first approach requires the addition of a (systematic) erasure channel coding component on top of some carefully chosen source codes, and for $K = 3$ this method is quite similar to the approach used in network coding [11]. The second approach utilizes the quantization splitting technique, which was previously used to design simple yet optimal coding schemes for the Gaussian CEO problem [12] and the Gaussian two description problem [13]. We show that in the K-description problem, quantization splitting can also serve as a useful tool to boost the system performance. Though the quantization splitting technique itself is not new and has been used in other problems, utilizing it in the K-description problem poses particular difficulties. In the existing works [12], [13], quantization splitting can be performed rather conveniently by invoking a sequential encoding and decoding order; however, in the MD problem, due to the rather involved distortion requirements, using such a completely sequential order is no longer appropriate. We circumvent this difficulty by specifying only a partial coding order, instead of a completely sequential one.

The first approach of incorporating channel codes was in fact mentioned briefly in [7] (Section VI) as a possible way to further improve the PPR multilayer scheme. However, it is worth mentioning that one key new finding in our current work is that the channel coding component should operate on some specific source codes, which have to be chosen strategically. Furthermore, the fact that for the Gaussian source there exist R-D points outside the PPR multilayer achievable region is revealed for the first time.

---

[1]This statement is proved in the subsequent section more rigorously.

The remainder of the paper is organized as follows. The problem definition is given in Section II, and then the PPR multilayer scheme and the quantization splitting technique are reviewed briefly. In Section III we introduce two simple coding approaches and show that they are useful under the no two description excess rate requirement, yielding R-D points outside the PPR-multilayer R-D region. In Section IV we generalize these two approaches to the K-description case in the setting of single layer coding. Section V further generalizes the schemes to include multiple layers. Section VI concludes the paper with a few remarks.

II. PROBLEM DEFINITION AND REVIEW

In this section, we first give a few necessary notations and the problem definition, then review briefly the results on the PPR multilayer scheme, and the quantization splitting technique in the context of the Gaussian CEO problem.

A. Notations and Problem Definition

Since we use the Gaussian source as the working example, it is convenient to define the problem for the Gaussian source below. The problem can be defined for general discrete memoryless sources under bounded distortion measures with only minor changes; moreover, the results presented in Sections IV and V hold for both kinds of sources using the techniques in [14], and thus we do not distinguish them in the remainder of the paper unless necessary.

Let $\{X(i)\}_{i=1,2,\ldots}$ be a memoryless and stationary Gaussian source with zero-mean and unit-variance; the source alphabet $\mathcal{X}$ is thus $\mathbb{R}$ in this case. The vector $X(1), X(2), \ldots, X(n)$ will be denoted as X^n. The mean squared error (MSE) distortion $d(x^n, y^n) = \frac{1}{n} \sum_{i=1}^n (x(i) - y(i))^2$ will be used; this implies that the reconstruction alphabet is also $\mathbb{R}$ in this case, i.e., the same as the source alphabet, but the restriction can be relaxed straightforwardly when other sources are considered. The notation $\mathcal{I}_k$ is used to denote the set $\{1, 2, \ldots, k\}$. Throughout the paper, the entropy of a random variable X is denoted as $H(X)$, and if the source is continuous, e.g., the Gaussian source, the notation should be implicitly understood as the differential entropy. For an arbitrary set $\mathcal{A}$, we use $2_+^{\mathcal{A}}$ to denote the set of all nonempty subsets of $\mathcal{A}$, i.e.,

$$2_+^{\mathcal{A}} \triangleq \{\mathcal{S} : \mathcal{S} \subseteq \mathcal{A}, |\mathcal{S}| > 0\} \tag{1}$$

where $|\mathcal{S}|$ is the cardinality of the set $\mathcal{S}$. For a set of random variables with double subscript $\{X_{i,j}, i \in \mathcal{A}, j \in \mathcal{B}\}$, we sometimes write it simply as $\bar{X}_{\mathcal{A},\mathcal{B}}$; similarly the set $\{X_{i,j}, j \in \mathcal{A}\}$ is sometimes written as $\bar{X}_{i,\mathcal{A}}$. All the sets, such as those denoted by $\mathcal{A}$, $\mathcal{B}$, or $\mathcal{C}$ in the remainder of the paper, are inherently understood as subsets of $\mathcal{I}_K$, unless specified otherwise.

For the general K-description problem, a length-n block of the source samples is encoded into K descriptions. A specific decoder $T_{\mathcal{A}}$, where $\mathcal{A} \in 2_+^{\mathcal{I}_K}$, has access to the descriptions in the set $\mathcal{A}$. The symmetric distortion constraints are given such that the decoder $T_{\mathcal{A}}$ can reconstruct the source sequence to satisfy a certain distortion $D_{|\mathcal{A}|}$, i.e., the distortion constraints depend only on the number of descriptions that the decoder has access to, but not the particular combination of descriptions.

Formally, an $(n, (M_i, \ i \in \mathcal{I}_K), (\Delta_{\mathcal{A}}, \ \mathcal{A} \in 2_+^{\mathcal{I}_K}))$ code is defined by encoding functions S_i and decoding functions $T_{\mathcal{A}}$

$$S_i : \mathcal{X}^n \to \mathcal{I}_{M_i}, \quad i \in \mathcal{I}_K,$$
$$T_{\mathcal{A}} : \prod_{i \in \mathcal{A}} \mathcal{I}_{M_i} \to \mathcal{X}^n, \quad \mathcal{A} \in 2_+^{\mathcal{I}_K},$$

and

$$\Delta_{\mathcal{A}} = \mathbb{E} d(X^n, T_{\mathcal{A}}(S_i(X^n), i \in \mathcal{A})), \quad \mathcal{A} \in 2_+^{\mathcal{I}_K}$$

where $\mathbb{E}$ is the expectation operator.

A $(K+1)$ symmetric R-D tuple $(R, D_1, D_2, \ldots, D_K)$ is said to be achievable if for every $\epsilon > 0$, there exists for sufficiently large n an $(n, (M_i, \ i \in \mathcal{I}_K), (\Delta_{\mathcal{A}}, \ \mathcal{A} \in 2_+^{\mathcal{I}_K}))$ code such that

$$\frac{1}{n} \log M_i \leq R + \epsilon, \quad i \in \mathcal{I}_K,$$
$$\Delta_{\mathcal{A}} \leq D_{|\mathcal{A}|} + \epsilon, \quad \mathcal{A} \in 2_+^{\mathcal{I}_K}.$$

The collection of all the achievable $(K+1)$ symmetric R-D tuples $(R, D_1, D_2, \ldots, D_K)$ is of interest. The setting is sufficiently general such that achievable asymmetric R-D tuples or the achievable asymmetric R-D region can also be defined, but they are omitted here for brevity. Occasionally, we need to consider the distortion associated with the descriptions in a particular set $\mathcal{A}$ such that $|\mathcal{A}| = k$, and we write it as $D_{k,\mathcal{A}}$. Throughout this paper, we use base-2 logarithm.

B. Review of the PPR Multilayer Scheme

We first quote the main theorem given in [7] together with a necessary definition, then briefly explain the coding scheme and provide a few new observations; more details can be found in [6] and [7].

Definition 1 ([7]): A joint distribution $p(\bar{y}_{\mathcal{I}_{K-1}, \mathcal{I}_K}, y_K | x)$ is called symmetric if for all $1 \leq n_i \leq K$ where $i \in \mathcal{I}_{K-1}$, it is true that the joint distribution of Y_K and all $(n_1 + n_2 + \cdots + n_{K-1})$ random variables where any n_k are chosen from the set $\{Y_{k,1}, Y_{k,2}, \ldots, Y_{k,K}\}$, conditioned on X, is the same.

Theorem 1 ([7, Th. 2]): For any symmetric probability distribution $p(\bar{y}_{\mathcal{I}_{K-1}, \mathcal{I}_K}, y_K | x)$ over $\mathcal{Y}^{K(K-1)+1}$

$$p(x, \bar{y}_{\mathcal{I}_{K-1}, \mathcal{I}_K}, y_K) = p(x) p(\bar{y}_{\mathcal{I}_{K-1}, \mathcal{I}_K}, y_K | x)$$

and a set of decoding functions

$$g_{\mathcal{A}} : \mathcal{Y}^{|\mathcal{A}||\mathcal{A}|} \to \mathcal{X}, \quad 1 \leq |\mathcal{A}| \leq K - 1,$$
$$g_{\mathcal{I}_K} : \mathcal{Y}^{K(K-1)+1} \to \mathcal{X}$$

such that

$$\mathbb{E} d(X, g_{\mathcal{A}}(\bar{Y}_{\mathcal{I}_{|\mathcal{A}|}, \mathcal{A}})) \leq D_{|\mathcal{A}|}, \quad 1 \leq |\mathcal{A}| \leq K - 1,$$
$$\mathbb{E} d(X, g_{\mathcal{I}_K}(\bar{Y}_{\mathcal{I}_{K-1}, \mathcal{I}_K}, Y_K)) \leq D_K$$

the following rate is achievable:

$$R \geq \sum_{k=1}^{K-1} \frac{1}{k} H(\bar{Y}_{k,\mathcal{I}_k} | \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_k})$$
$$+ \frac{1}{K} H(Y_K | \bar{Y}_{\mathcal{I}_{K-1}, \mathcal{I}_K}) - \frac{1}{K} H(\bar{Y}_{\mathcal{I}_{K-1}, \mathcal{I}_K}, Y_K | X). \quad (2)$$

The PPR multilayer scheme can be described roughly as follows. There are a total of K layers in this coding scheme, and the encoding and decoding can be performed from lower layers to higher layers sequentially. The first $K-1$ layers are encoded and decoded using a different mechanism from the last layer. At the kth layer ($k \in \mathcal{I}_{K-1}$) and for any description $j \in \mathcal{I}_K$, a codebook of size $2^{n\hat{R}_{k,j}}$ is generated using the marginal distribution of $Y_{k,j}$. The codewords in a codebook are then randomly and independently assigned into a total of $2^{nR_{k,j}}$ bins, for all $k \in \mathcal{I}_{K-1} \backslash \{1\}$ and $j \in \mathcal{I}_K$, where $\backslash$ is the set difference operation.

The rates $\hat{R}_{k,j}$'s should be sufficiently large such that for any typical source sequence, with high probability there exist codewords in these codebooks of a certain layer such that they are jointly typical with the source sequence and the codewords previously found in the lower layers. Using the property of symmetric distribution (see [7]), it can be shown that this can be done with

$$\sum_{j=1}^{K} \hat{R}_{k,j} > \sum_{j=1}^{K} H(Y_{k,j}) - H(\bar{Y}_{k,\mathcal{I}_K} | X, \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_K}),$$
$$k \in \mathcal{I}_{K-1}. \quad (3)$$

As given in [7], the rate can thus be chosen as

$$\hat{R}_{k,j} = \hat{R}_k \triangleq H(Y_{k,1}) - \frac{1}{K} H(\bar{Y}_{k,\mathcal{I}_K} | X, \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_K}) + \delta,$$
$$k \in \mathcal{I}_{K-1} \quad (4)$$

for an arbitrarily small but positive δ. The bin indices (or the codeword indices for the first layer) of the found codewords are then put into the respective descriptions.

At the decoder, with any k descriptions such that $k \in \mathcal{I}_{K-1}$, the lower k layers are decoded sequentially from the lower to the higher layers in k steps. More precisely, the decoder receives descriptions in the set $\mathcal{A}$ such that $|\mathcal{A}| = k$. Taking an induction approach, we can assume the first $k-1$ layers of codewords are correctly decoded, and only need to consider the decoding of the kth layer. If there exists a unique set of codewords in the bins of the corresponding codebooks, specified by the descriptions in the set $\mathcal{A}$, that are jointly typical with each other, and at the same time they are jointly typical with the correctly decoded lower layer codewords, then the decoder reconstructs using the single-letter decoding function $g_{\mathcal{A}}(\cdot)$; otherwise a decoding failure occurs. To succeed with high probability for any $k \in \mathcal{I}_{K-1}$, it can be shown, using the property of symmetric distribution, that the rates $R_{k,j}, j \in \mathcal{I}_K$, only need to satisfy

$$0 \leq R_{k,j} \leq \hat{R}_{k,j}, \quad j \in \mathcal{I}_{K-1} \quad (5)$$

and

$$\sum_{j \in \mathcal{A}} (\hat{R}_{k,j} - R_{k,j}) < k H(Y_{k,1}) - H(\bar{Y}_{k,\mathcal{I}_k} | \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_k}),$$
$$|\mathcal{A}| = k \quad (6)$$

which should hold for any $k \in \mathcal{I}_{K-1}$.

Rewriting (6), we have for all $\mathcal{A}$ such that $|\mathcal{A}| = k$, $k \in \mathcal{I}_{K-1}$, that

$$
\begin{aligned}
\sum_{j\in\mathcal{A}} R_{k,j} &> \sum_{j\in\mathcal{A}} \hat{R}_{k,j} - kH(Y_{k,1}) + H(\bar{Y}_{k,\mathcal{I}_k}|\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_k}) \\
&= H(\bar{Y}_{k,\mathcal{I}_k}|\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_k}) \\
&\quad - \frac{k}{K}H(\bar{Y}_{k,\mathcal{I}_K}|X,\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_K}) + k\delta. \quad (7)
\end{aligned}
$$

The last layer codebook is generated using the more conventional method, i.e., the conditional codebook, and the following condition is sufficient:

$$
\sum_{k=1}^{K} R_{K,k} > I(X;Y_K|\bar{Y}_{\mathcal{I}_{K-1},\mathcal{I}_K}). \quad (8)
$$

The given rate in Theorem 1 can be attained by choosing the symmetric tuple, i.e.,

$$
\begin{aligned}
R_{k,j} = R_k &= \frac{1}{k}H(\bar{Y}_{k,\mathcal{I}_k}|\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_k}) \\
&\quad - \frac{1}{K}H(\bar{Y}_{k,\mathcal{I}_K}|X,\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_K}) + \delta', \\
&\qquad k \in \mathcal{I}_{K-1}, \ j \in \mathcal{I}_K \quad (9)
\end{aligned}
$$

where δ' is a small positive quantity such that $\delta' > \delta$, and by making δ' and δ sufficiently small.

In this coding scheme, if only one layer exists, then it is a single (layer) source channel erasure code (SCEC). For example if only the kth level exists in a K-description PPR multilayer scheme, then it is a (K,k) SCEC; for this case, (3) and (6) only need to hold for a single particular value of k. It is not difficult to see that (4) and (9) can be readily applied to this case by taking the first $k-1$ layers random variables $\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_K}$ to be constants. It is clear that the small quantities δ's are inconsequential, and they will be ignored in the remaining sections.

Two important remarks should be made at this point. Firstly, we note (see also [9]) that it is not necessary to keep the distribution, rates $\hat{R}_{k,j}$'s and $R_{k,j}$'s all symmetric in this coding scheme. In fact, the conditions on the choice of rates can be easily extracted in [6] and [7], when these symmetric conditions are removed, and these conditions can then be reduced to give Theorem 1 when the rates and distribution are chosen to be symmetric. Our second approach to boost the system performance can be understood as utilizing this observation. Secondly, the symmetric distribution requirement can be straightforwardly relaxed to the following one, and Theorem 1 still holds under this relaxed requirement (see also the related discussion in [10]). This observation is useful for the development of the general coding schemes.

Definition 2: A joint distribution $p(\bar{y}_{\mathcal{I}_{K-1},\mathcal{I}_K}, y_K|x)$ is called *generalized symmetric* if for any permutation $\pi(\cdot) : \mathcal{I}_K \rightarrow \mathcal{I}_K$, the joint distribution $p(\bar{y}_{\mathcal{I}_{K-1},\pi(\mathcal{I}_K)}, y_K|x)$ is the same as $p(\bar{y}_{\mathcal{I}_{K-1},\mathcal{I}_K}, y_K|x)$.

In the above definition, though $\bar{y}_{\mathcal{I}_{K-1},\mathcal{I}_K}$ is written as a set, we assume the random variables in this set are inherently ordered in a predefined manner, and thus the definition is
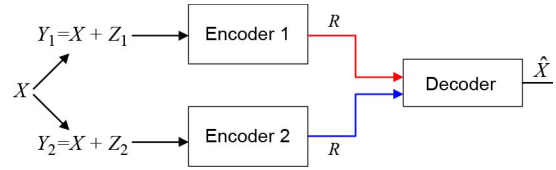

Fig. 2. System diagram for symmetric two encoder Gaussian CEO problem. For simplicity we assume the independent zero-mean Gaussian random variables Z_1 and Z_2 have the same variance.

meaningful; also we slightly abuse the notation for the permutation operator, such that $\pi(\mathcal{I}_K)$ is the resulting sequence after the permutation under this just given convention. The original version of symmetric distribution essentially requires the distribution to be invariant under $K-1$ different permutations $\pi_k(\cdot)$, one for each layer $k = 1, 2, \ldots, K-1$; i.e., if we permute $\{Y_{1,1}, Y_{1,2}, \ldots, Y_{1,K}\}$, and then permute $\{Y_{2,1}, Y_{2,2}, \ldots, Y_{2,K}\}$ differently, and so on for each $k = 1, 2, \ldots, K-1$, the resulting distribution should remain the same as the one before such permutations. This requirement was, however, never completely utilized in the coding scheme in [7], which instead in fact requires only invariance under a single permutation $\pi(\cdot)$ which is applied to all the levels simultaneously, i.e., $\pi_k(\cdot) = \pi(\cdot)$, for $k = 1, 2, \ldots, K-1$.

For the quadratic unit-variance Gaussian source with three descriptions, one particular interesting case is when there is no two description excess rate, i.e., when any two of the descriptions are R-D optimal jointly. For this case with rate R per description, it was shown in [6] that the following R-D quadruple is achievable:

$$
(R, D_1, D_2, D_3) = \left(R, 1, 2^{-4R}, \frac{2 \cdot 2^{-4R}}{3 - 2^{-4R}}\right) \quad (10)
$$

by using only the second layer in the PPR multilayer scheme, and letting

$$
Y_{2,i} = X + N_{2,i}, \quad i = 1, 2, 3 \quad (11)
$$

where $N_{2,i}$'s are zero-mean independent identically distributed Gaussian random variables, also independent of the source X, with variance $2(2^{4R} - 1)^{-1}$. More generally, for the K-description case, when any k descriptions are R-D optimal jointly, SCEC can be shown to achieve the R-D tuple

$$
\begin{aligned}
&(R, D_1, \ldots, D_{k-1}, D_k, \ldots, D_{k+r}, \ldots, D_K) \\
&= \left(R, 1, \ldots, 1, 2^{-2kR}, \ldots, \frac{k2^{-2kR}}{k+r-r2^{-2kR}},\right. \\
&\qquad \left.\ldots, \frac{k2^{-2kR}}{K-(K-k)2^{-2kR}}\right). \quad (12)
\end{aligned}
$$

Another achievable R-D tuple of interest using SCEC is when the combination of all K descriptions is R-D optimal. It was shown in [6] that for the Gaussian source, an SCEC can achieve

$$
\begin{aligned}
&(R, D_1, \ldots, D_j, \ldots, D_K) \\
&= \left(R, 1 - \frac{1}{K}(1 - 2^{-2KR}), \ldots, 1 - \frac{j}{K}(1 - 2^{-2KR}),\right. \\
&\qquad \left.\ldots, 2^{-2KR}\right). \quad (13)
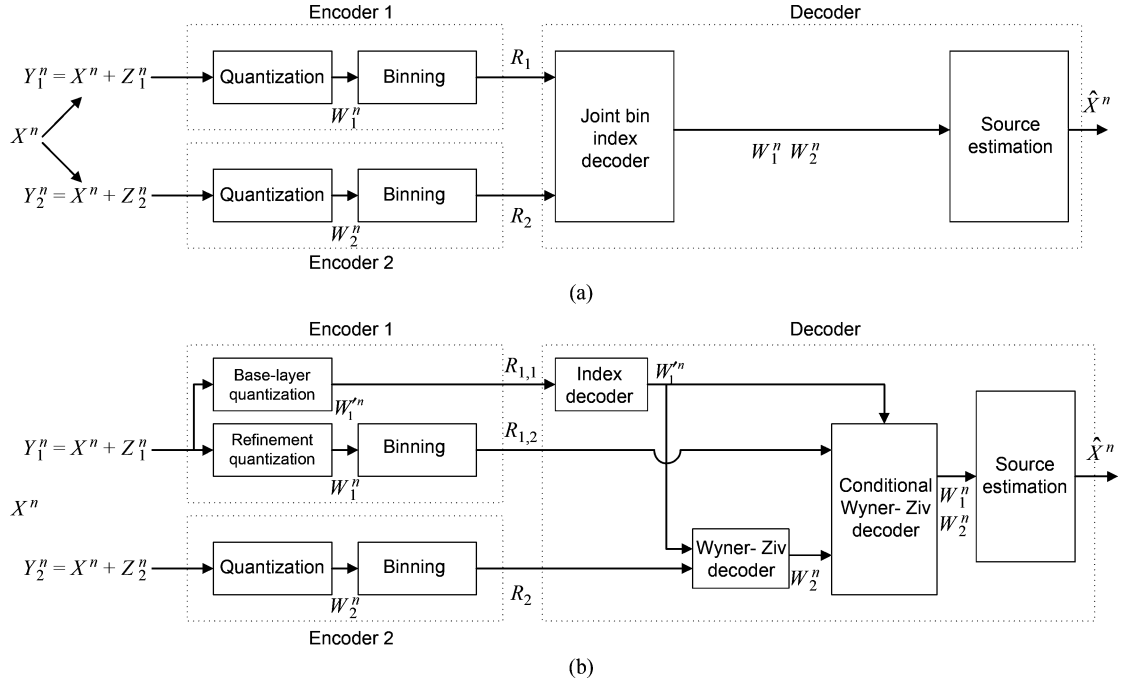\end{aligned}
$$

Fig. 3. Converting the joint decoding procedure into successive decoding through quantization splitting. The binning rates at encoder 1 in (a) and (b) are different.

This R-D tuple is not a special case of (12), i.e., it cannot be deduced from (12) by setting $k = K$ or $k = 1$. In fact the dependence structure among the auxiliary random variables achieving (12) and that achieving (13) are quite different, and we will return to this subtle point later in Section III-C.

C. Review of Quantization Splitting in the Gaussian CEO Problem

Quantization splitting was thoroughly investigated for the Gaussian CEO problem in [12], and was previously also used to construct optimal coding schemes for the Gaussian two description problem [13]. Similar techniques in channel coding are usually referred to as rate splitting [15]–[17], and those in lossless compression are usually referred to as source splitting [18]. Next we give a brief review of the quantization splitting technique in the context of two-encoder Gaussian CEO problem, which is depicted in Fig. 2.

The R-D region characterization for the Gaussian CEO problem was independently discovered by Oohama [19] and Prabhakaran *et al.* [20]. More relevant to us is a special optimal R-D tuple specified by the following equations:

$$R_1 = R_2 = I(W_1; Y_1) - \frac{1}{2} I(W_1; W_2),$$
$$D = \mathbb{E}(X - \mathbb{E}(X|W_1, W_2))^2,$$

where $W_1 = Y_1 + N_1$ and $W_2 = Y_2 + N_2$, N_1 and N_2 are independent Gaussian random variables with variance σ_N^2, also independent of everything else. This R-D triple can be achieved by the so called Berger–Tung scheme [21], [22], depicted in Fig. 3(a), from which we see that the codewords are found by jointly decoding using their bin indices. However, one crucial observation given in [12] is that the scheme can be converted to a successive coding scheme by using the quantization splitting technique, resulting in the coding structure depicted in Fig. 3(b).

Here we choose $W_1' = W_1 + N_1'$, where N_1' is a Gaussian random variable independent of everything else, as an intermediate quantization step at the first encoder, thus the name quantization splitting. The idea is that at the decoder, this quantization information split from W_1 can be decoded by itself, and subsequently the Wyner–Ziv [23] decoding module can be used, instead of the more complicated joint decoding. It is important to note that by properly choosing the variance of N_1', the following two properties can be made true.

- In the quantization splitting scheme, the first encoder rate $R_{1,1} + R_{1,2} = R_1$, and the second encoder rate R_2 remains the same; the distortion is also preserved.
- Encoder 2 in the two coding schemes can be identical.

It is easy to check that we only need to ensure

$$I(W_2; W_1') = \frac{1}{2} I(W_2; W_1) \qquad (14)$$

for which there always exists a valid solution of W_1'.

The second approach given in the next section is based on the quantization splitting technique, and we are motivated partly by the similarity between the SCEC coding structure and the Berger–Tung coding structure in the CEO problem, which will be discussed in more details in the next section.

III. TWO SIMPLE SCHEMES FOR NO TWO DESCRIPTION EXCESS RATE CODING

In this section, we focus on a special case of the three description problem where any two descriptions are R-D optimal jointly. We first establish an outer bound for the PPR multilayer scheme rate region under this requirement, then provide two simple coding schemes, which can achieve R-D points outside the PPR multilayer achievable region. These two schemes are the basis of the general schemes given in later sections, and in this section we present them in the simplest form to facilitate understanding.

A. An Outer Bound for the PPR Multilayer Achievable Region

We first formally define the PPR multilayer achievable region $\mathcal{Q}_{PPR}^{(K)}$, and then provide an outer bound for it.

Definition 3: For the unit variance Gaussian source, PPR multilayer achievable region $\mathcal{Q}_{PPR}^{(K)}$ is the convex closure of the collection of $(R, D_1, D_2, \ldots, D_K)$ tuples which satisfy Theorem 1 for some generalized symmetric probability distribution $p(\bar{y}_{\mathcal{I}_{K-1}, \mathcal{I}_K}, y_K | x)$.

We establish the following theorem for the case of $K = 3$, which essentially states that under the no two description excess rate requirement, each individual description in the PPR multilayer scheme is useless.

Theorem 2: For the quadratic unit-variance Gaussian source, if the R-D quadruple $(R, D_1, D_2, D_3) = (R, D_1, 2^{-4R}, D_3) \in \mathcal{Q}_{PPR}^{(3)}$, then $D_1 \geq 1$.

Proof: Since the conventional single description R-D function for the Gaussian source, denoted as $R(d)$, is strictly convex, it is clear that to achieve $(R, D_1, D_2, D_3) = (R, D_1, 2^{-4R}, D_3)$ by time-sharing several R-D points in $\mathcal{Q}_{PPR}^{(3)}$, each point being time-shared must have the form $(R, D_1', 2^{-4R}, D_3')$. Let us consider any one of these points. It follows from the supposition in the theorem that there exist random variables $\{\bar{Y}_{\mathcal{I}_2, \mathcal{I}_3}, Y_3\}$ such that

$$
\begin{aligned}
\frac{1}{2}R(D_2) &\geq \sum_{k=1}^{2} \frac{1}{k} H(\bar{Y}_{k, \mathcal{I}_k} | \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_k}) \\
&\quad + \frac{1}{3} H(Y_3 | \bar{Y}_{\mathcal{I}_2, \mathcal{I}_3}) - \frac{1}{3} H(\bar{Y}_{\mathcal{I}_2, \mathcal{I}_3}, Y_3 | X) \\
&= H(Y_{1,1}) + \frac{1}{2} H(\bar{Y}_{2, \mathcal{I}_2} | \bar{Y}_{1, \mathcal{I}_2}) \\
&\quad - \frac{1}{3} H(\bar{Y}_{\mathcal{I}_2, \mathcal{I}_3} | X) + \frac{1}{3} I(Y_K; X | Y_{\mathcal{I}_2, \mathcal{I}_3}) \\
&\overset{(a)}{\geq} H(Y_{1,1}) + \frac{1}{2} H(\bar{Y}_{2, \mathcal{I}_2} | \bar{Y}_{1, \mathcal{I}_2}) \\
&\quad - \frac{1}{3} H(\bar{Y}_{\mathcal{I}_2, \mathcal{I}_3} | X) \\
&\overset{(b)}{=} H(Y_{1,1}) - \frac{1}{2} H(\bar{Y}_{1, \mathcal{I}_2}) + \frac{1}{2} H(\bar{Y}_{\mathcal{I}_2, \mathcal{I}_2}) \\
&\quad - \frac{1}{3} H(\bar{Y}_{\mathcal{I}_2, \mathcal{I}_3} | X) \\
&\overset{(c)}{\geq} \frac{1}{2} H(\bar{Y}_{\mathcal{I}_2, \mathcal{I}_2}) - \frac{1}{3} H(\bar{Y}_{\mathcal{I}_2, \mathcal{I}_3} | X)
\end{aligned}
\tag{15}
$$

where (a) is by the nonnegativity of mutual information, (b) is by applying the chain rule, and (c) is by the symmetric assumption $H(Y_{1,1}) = H(Y_{1,2})$ and the nonnegativity of $I(Y_{1,1}; Y_{1,2})$. Continue this chain of inequalities

$$
\begin{aligned}
\frac{1}{2}R(D_2) &\geq \frac{1}{2} H(\bar{Y}_{\mathcal{I}_2, \mathcal{I}_2}) - \frac{1}{3} H(\bar{Y}_{\mathcal{I}_2, \mathcal{I}_3} | X) \\
&= \frac{1}{2} I(\bar{Y}_{\mathcal{I}_2, \mathcal{I}_2}; X) + \frac{1}{2} H(\bar{Y}_{\mathcal{I}_2, \mathcal{I}_2} | X) \\
&\quad - \frac{1}{3} H(\bar{Y}_{\mathcal{I}_2, \mathcal{I}_3} | X) \\
&\overset{(d)}{\geq} \frac{1}{2} I(\bar{Y}_{\mathcal{I}_2, \mathcal{I}_2}; X)
\end{aligned}
\tag{16}
$$

where the equality is by adding and subtracting the same term, and (d) follows from the conditional version of Han's inequality [25] and the symmetry. Since $\bar{Y}_{\mathcal{I}_2, \mathcal{I}_2}$ can be used to recover X within distortion D_2, by the conventional R-D theorem

$$
I(\bar{Y}_{\mathcal{I}_2, \mathcal{I}_2}; X) \geq R(D_2).
\tag{17}
$$

We thus have

$$
\frac{1}{2} R(D_2) \geq \frac{1}{2} I(\bar{Y}_{\mathcal{I}_2, \mathcal{I}_2}; X) \geq \frac{1}{2} R(D_2).
\tag{18}
$$

Since the beginning and the end of this chain of inequalities are the same, all the inequalities have to be equalities. To keep (c) and (d) holding with equality, we must have

$$
I(Y_{1,1}; Y_{1,2}) = 0, \quad I(\bar{Y}_{\mathcal{I}_2, 1}; \bar{Y}_{\mathcal{I}_2, 2} \bar{Y}_{\mathcal{I}_2, 3} | X) = 0.
\tag{19}
$$

The second equality in (19) leads to

$$
I(Y_{1,1}; Y_{1,2} | X) = 0.
\tag{20}
$$

If we assume the joint distribution among X and $\{\bar{Y}_{\mathcal{I}_2, \mathcal{I}_3}, Y_3\}$ is jointly Gaussian, i.e., a Gaussian codebook is used, then the proof is rather straightforward from here on, as was given in [26]. Next we provide a proof without the Gaussianity assumption. In the following, we assume $Y_{1,1}, Y_{1,2}, Y_{1,3}$ are real-valued continuous random variables which have probability density functions and the differential entropies are well defined; the case when they are discrete random variables can be treated similarly. For these two cases, the (differential) entropies are well defined, and Theorem 1 is meaningful.

Since $I(Y_{1,1}; Y_{1,2}) = 0$ implies that $Y_{1,1}$ and $Y_{1,2}$ are independent, we can write their joint density $f_{Y_{1,1}, Y_{1,2}}(\cdot, \cdot)$ as

$$
f_{Y_{1,1}, Y_{1,2}}(y_a, y_b) = f_{Y_{1,1}}(y_a) f_{Y_{1,2}}(y_b)
\tag{21}
$$

for any $y_a, y_b \in \mathbb{R}$. We thus have

$$
\begin{aligned}
&f_{Y_{1,1}, Y_{1,2}}(y_a, y_b) f_{Y_{1,1}, Y_{1,2}}(y_b, y_a) \\
&= f_{Y_{1,1}}(y_a) f_{Y_{1,1}}(y_b) f_{Y_{1,2}}(y_b) f_{Y_{1,2}}(y_a) \\
&= f_{Y_{1,1}, Y_{1,2}}(y_a, y_a) f_{Y_{1,1}, Y_{1,2}}(y_b, y_b).
\end{aligned}
\tag{22}
$$

Since $I(Y_{1,1}; Y_{1,2} | X) = 0$ implies the Markov string $Y_{1,1} \leftrightarrow X \leftrightarrow Y_{1,2}$, and by the symmetry in the distribution of Theorem 1, we can write the joint density of $f_{X, Y_{1,1}, Y_{1,2}}(\cdot, \cdot, \cdot)$ as

$$
f_{X, Y_{1,1}, Y_{1,2}}(x, y_a, y_b) = f_X(x) f_{Y|X}(y_a | x) f_{Y|X}(y_b | x)
\tag{23}
$$

for any triple $x, y_a, y_b \in \mathbb{R}$. However, this implies that

$$
\begin{aligned}
&f_{Y_{1,1}, Y_{1,2}}(y_a, y_b) \\
&= \int_{-\infty}^{\infty} f_{X, Y_{1,1}, Y_{1,2}}(x, y_a, y_b) dx \\
&= \int_{-\infty}^{\infty} f_X(x) f_{Y|X}(y_a | x) f_{Y|X}(y_b | x) dx.
\end{aligned}
\tag{24}
$$

Recall that Cauchy–Schwarz inequality states that for any square-integrable real-valued functions $g_1(x)$ and $g_2(x)$, we have

$$
\left(\int g_a(x) g_b(x) dx \right)^2 \leq \int (g_a(x))^2 dx \int (g_b(x))^2 dx
$$

where equality holds if and only if $g_a(x)$ and $g_b(x)$ are linear dependent. Now with fixed y_a and y_b, take

$$g_a(x) = \sqrt{f_X(x)} f_{Y|X}(y_a|x)$$
$$g_b(x) = \sqrt{f_X(x)} f_{Y|X}(y_b|x)$$

it follows from (24) that

$$f_{Y_{1,1},Y_{1,2}}(y_a, y_b) f_{Y_{1,1},Y_{1,2}}(y_b, y_a)$$
$$= \left(\int g_a(x) g_b(x) dx \right)^2$$
$$\leq \int (g_a(x))^2 dx \int (g_b(x))^2 dx$$
$$= f_{Y_{1,1},Y_{1,2}}(y_a, y_a) f_{Y_{1,1},Y_{1,2}}(y_b, y_b).$$

However, this inequality can only hold with equality because of (22), and thus $g_a(x) = \sqrt{f_X(x)} f_{Y|X}(y_a|x)$ and $g_b(x) = \sqrt{f_X(x)} f_{Y|X}(y_b|x)$ are linearly dependent, and it follows that $g_a^* \triangleq f_{Y|X}(y_a|x)$ and $g_b^* \triangleq f_{Y|X}(y_b|x)$ are also linearly dependent, for any fixed y_a and y_b. This is equivalent to the independence between the random variables X and $Y_{1,1}$ (and $Y_{1,2}$, $Y_{1,3}$ by symmetry). Thus, by the converse to the conventional rate distortion theorem, we conclude that $D_1' \geq 1$ and thus $D_1 \geq 1$, and the proof is complete. ∎

Note that the symmetric distribution assumption in Theorem 1 was used in the proof at several places. It will become clear in the sequel that this is a severe limitation of the PPR multilayer scheme, and asymmetry can be introduced in various ways to improve the system performance, even when only the symmetric MD problem is considered.

B. The Coding Scheme Based on Channel Codes

The complete R-D region for the Gaussian two description problem was established in [3], from which it can be easily checked that the following symmetric R-D triple is achievable:

$$(R, D_1, D_2) = \left(R, \frac{1}{2}(1 + 2^{-4R}), 2^{-4R} \right) \qquad (25)$$

i.e., the two descriptions are R-D optimal jointly.

Now for this R-D point, there is a sufficiently long code that operates near it, and we denote the resulting codeword indices as S_1 and S_2, respectively; moreover, we assume they are written as binary sequences. Now the third description is simply given as $S_1 \oplus S_2$, where $\oplus$ is the modulo two addition (binary XOR operation). With this simple scheme, it is straightforwardly seen that these three descriptions can achieve the following asymmetric distortions using rate R per description:

$$D_{1,\{1\}} = D_{1,\{2\}} = \frac{1}{2}(1 + 2^{-4R}), \quad D_{1,\{3\}} = 1, \quad (26)$$
$$D_2 = 2^{-4R}, \quad D_3 = 2^{-4R}. \qquad (27)$$

By a simple time-sharing argument, it is seen that the following symmetric R-D quadruple is achievable:

$$(R, D_1, D_2, D_3) = \left(R, \frac{1}{3} 2^{-4R} + \frac{2}{3}, 2^{-4R}, 2^{-4R} \right). \quad (28)$$

The modulo two addition operation introduces a channel erasure code, and it is extremely similar in spirit to network coding in
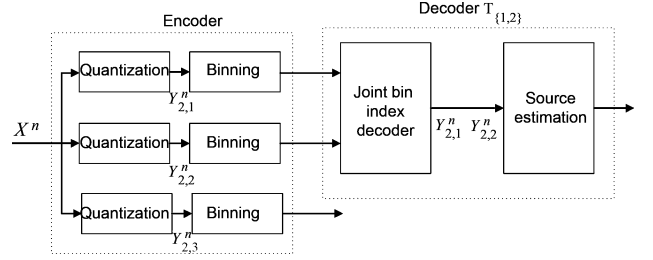

Fig. 4. The SCEC coding scheme. Though the quantization noises can be correlated in general, in the special case discussed in this section, they are in fact independent.

the well known butterfly network [11]. We thus refer to this specific R-D point as the *network coding point*. This channel code introduced by the modulo two addition is essentially a $(3,2)$ systematic maximum distance separable (MDS) code, and this view becomes important for the general K-description problem considered in later sections.

The following theorem now follows from Theorem 2 and the performance of the network coding point.

Theorem 3: For the unit variance Gaussian source, the network coding point R-D quadruple $(R, \frac{1}{3} 2^{-4R} + \frac{2}{3}, 2^{-4R}, 2^{-4R})$ is achievable, and it is outside $\mathcal{R}_{PPR}^{(3)}$.

One may wonder whether in the above scheme, using asymmetric distortion triples within the two description D-R region instead of the symmetric one given in (25), can lead to better performance. However, it can be easily checked that in order to satisfy the no two description excess rate requirement, asymmetric choice of distortion triples cannot yield better performance for the Gaussian three description problem considered here.

C. The Scheme Based on Quantization Splitting

For the network coding point, though individual descriptions are useful, the three description distortion D_3 does not improve upon the two description distortion D_2, in contrast to the binning point for which D_3 does improve upon D_2. Thus, the network coding point R-D tuple does not improve upon that of the binning point in a component-wise manner. Next, we introduce a new coding scheme based on quantization splitting, which indeed gives performance strictly dominating that of the binning point.

To understand the coding scheme in its simplest form, let us again consider the no two description excess rate case, and recall the PPR multilayer scheme with only the second layer SCEC. In Fig. 4, the encoding structure is given, and one of the decoder using the first and the second descriptions is also shown. Observe that Figs. 3(a) and 4 are strikingly similar, with $Y_{2,1}$ and $Y_{2,2}$ replacing the role of W_1 and W_2, and the noisy observations of the source in the CEO problem in fact become direct observation of the source. It is thus conceivable that the quantization splitting technique used in the scheme given in Fig. 3(b) can also be used to convert the SCEC in the MD problem. Indeed, now we can introduce another random variable defined as

$$X_{2,1} = Y_{2,1} + N_{2,1}' \qquad (29)$$

where $N'_{2,1}$ is a zero-mean Gaussian random variable independent of everything else, with variance $\sigma^2_{N'}$. Using this new random variable, an almost identical quantization splitting method given in Fig. 3 can be used. More precisely, the coding scheme can be given as follows with the specified joint distribution of $(X, Y_{2,1}, X_{2,1}, Y_{2,2}, Y_{2,3})$ defined in (11) and (29).

- Codebook generation: Four codebooks $C^y_{2,1}, C^x_{2,1}, C^y_{2,2}, C^y_{2,3}$ of sizes $2^{n\hat{R}^y_{2,1}}, 2^{nR^x_{2,1}}, 2^{n\hat{R}^y_{2,2}}, 2^{n\hat{R}^y_{2,3}}$ are generated independently according to the n-fold product of the marginal distribution of $Y_{2,1}, X_{2,1}, Y_{2,2}, Y_{2,3}$, respectively.
- Random binning: each codeword in $C^y_{2,1}$ is randomly assigned to one of $2^{nR^y_{2,1}}$ bins; similarly each codeword in $C^y_{2,2}$ and $C^y_{2,3}$ is randomly assigned to one of $2^{nR^y_{2,2}}$ and $2^{nR^y_{2,3}}$ bins, respectively.
- Encoding: The encoder attempts to find codewords $y^n_{2,1}, x^n_{2,1}, y^n_{2,2}, y^n_{2,3}$ in the codebooks $C^y_{2,1}, C^x_{2,1}, C^y_{2,2}, C^y_{2,3}$, respectively, such that they are jointly typical with a typical source sequence x^n. If this cannot be done, an encoding error is declared; otherwise the codeword index of $x^n_{2,1}$ (denoted as $i^x_{2,1}$) and the bin index of $y^n_{2,1}$ (denoted as $i^y_{2,1}$) are concatenated together as the first description information, and the bin indices of $y^n_{2,2}, y^n_{2,3}$ (denoted as $i^y_{2,2}$ and $i^y_{2,3}$) are used as the second and third description information, respectively.
- Decoding: At decoder $T_{\{1,2\}}$, the codeword $x^n_{2,1}$ is first decoded using its index $i^x_{2,1}$, and then $y^n_{2,2}$ is decoded by finding the unique codeword in the $i^y_{2,2}$th bin of $C^y_{2,2}$ codebook that is jointly typical with $x^n_{2,1}$, and finally $y^n_{2,1}$ is decoded by finding the unique codeword in the $i^y_{2,1}$th bin of $C^y_{2,1}$ that is jointly typical with the decoded $(x^n_{2,1}, y^n_{2,2})$. At decoder $T_{\{1,3\}}$, the decoding procedure is similar to the one used in decoder $T_{\{1,2\}}$. For decoder $T_{\{2,3\}}$, it tries to find a unique codeword in the $i^y_{2,2}$th bin of $C^y_{2,2}$ and one in the $i^y_{2,3}$th bin of $C^y_{2,3}$ that are jointly typical with each other. In the above decoding steps, if more than one codeword (or one pair of codewords) is found, then a decoding error is declared. The decoders eventually estimate the source sequence using the available decoded codeword sequences, if no error occurs.

In the codebook generation step, the sizes of the codebooks should be sufficiently large, such that jointly typical codewords can be found for any typical source sequence with high probability. During decoding, to ensure the decoders can decode with high probability, the sizes of bins cannot be too large. Though the exact conditions can be derived in a rigorous manner, here we only list one possible choice of such rates; the formal proof is delayed to the next section when a more general setting is considered in order to avoid repetition. The following rates are valid to guarantee reliable encoding (ignoring the small quantities δ's)

$$\hat{R}^x_{2,1} = R^x_{2,1} = I(X; X_{2,1}), \tag{30}$$

$$\hat{R}^y_{2,1} = I(X; Y_{2,1}|X_{2,1}) + I(X_{2,1}; Y_{2,1}), \tag{31}$$

$$\hat{R}^y_{2,2} = \hat{R}^y_{2,3} = I(X; Y_{2,2}) = I(X; Y_{2,3}). \tag{32}$$

Using the same converting technique of Wyner–Ziv decoding as in the CEO problem, it is easily checked that the following rates suffice for reliable decoding:

$$R^y_{2,1} = I(X; Y_{2,1}|X_{2,1}) + I(X_{2,1}; Y_{2,1}) - I(Y_{2,1}; X_{2,1}Y_{2,2}), \tag{33}$$

$$R^y_{2,2} = R^y_{2,3} = \frac{1}{2}I(Y_{2,1}, Y_{2,2}; X). \tag{34}$$

Take $K = 3$ and $k = 2$ in (9), and compare it with (34); it is clear that for the second and third descriptions, the rates remain the same as the original SCEC-based scheme when no splitting is used.

Next we check the rate for the first description is also the same as when splitting is not used

$$
\begin{aligned}
R^x_{2,1} + R^y_{2,1} &= I(X; X_{2,1}) + I(X; Y_{2,1}|X_{2,1}) \\
&\quad + I(X_{2,1}; Y_{2,1}) - I(Y_{2,1}; X_{2,1}Y_{2,2}) \\
&\overset{(a)}{=} I(X; Y_{2,1}) + I(X_{2,1}; Y_{2,1}) \\
&\quad - I(Y_{2,1}; X_{2,1}Y_{2,2}) \\
&\overset{(b)}{=} I(X; Y_{2,1}) - I(Y_{2,1}X_{2,1}; Y_{2,2}) \\
&\quad + I(X_{2,1}; Y_{2,2}) \\
&\overset{(c)}{=} I(X; Y_{2,1}) - I(Y_{2,1}; Y_{2,2}) \\
&\quad + I(X_{2,1}; Y_{2,2})
\end{aligned}
\tag{35}
$$

where (a) and (c) are due to the Markov string $(X, Y_{2,2}) \leftrightarrow Y_{2,1} \leftrightarrow X_{2,1}$, and (b) is by applying the chain rule twice. Recall the condition given in (14) when the Berger–Tung scheme is converted to the successive decoding scheme using quantization splitting, and take $W_1 = Y_{2,1}$, $W'_1 = X_{2,1}$ and $W_2 = Y_{2,2}$, the condition (14) is thus equivalent to

$$\frac{1}{2}I(Y_{2,1}; Y_{2,2}) = I(X_{2,1}; Y_{2,2}). \tag{36}$$

With this condition satisfied, it follows that

$$
\begin{aligned}
R^x_{2,1} + R^y_{2,1} &= I(X; Y_{2,1}) - \frac{1}{2}I(Y_{2,1}; Y_{2,2}) \\
&\overset{(d)}{=} \frac{1}{2}H(Y_{2,1}, Y_{2,2}) - H(N_1) \\
&= \frac{1}{2}I(Y_{2,1}, Y_{2,2}; X)
\end{aligned}
\tag{37}
$$

where (d) is by the definition of random variables in (11). Thus, indeed this new scheme based on quantization splitting preserves the rates of the original scheme. It is also seen that when any two descriptions (or all three descriptions) are available, the distortions are also preserved, i.e., it satisfies the no two description excess rate requirement. However, observe that if only the first description is available, the codeword based on $X_{2,1}$ can be decoded, which can be used to estimate the source.

To quantify the distortion with only the first description, the variance of $N'_{2,1}$ needs to be determined. From (36), and the

choice of variances of $N_{2,1}$ and $N_{2,2}$ being $2(2^{4R} - 1)^{-1}$, we derive that

$$\sigma_{N'}^2 = \frac{2}{2^{2R} - 1} - \frac{2}{2^{4R} - 1}. \qquad (38)$$

Thus, the linear minimum mean squared error of estimating X using $X_{2,1} = X + N_{2,1} + N_{2,1}'$ can be computed to be $2(2^{2R} + 1)^{-1}$. Summarizing the above derivation, it is clear now that the following asymmetric distortions are achievable using rate R per description:

$$D_{1,\{1\}} = 2(2^{2R} + 1)^{-1}, \quad D_{1,\{2\}} = D_{1,\{3\}} = 1,$$
$$D_2 = 2^{-4R}, \quad D_3 = \frac{2 \cdot 2^{-4R}}{3 - 2^{-4R}}. \qquad (39)$$

This implies the following symmetric R-D quadruple can be achieved by a time-sharing argument:

$$(R, D_1, D_2, D_3) = \left(R, \frac{2 \cdot 2^{2R} + 4}{3(2^{2R} + 1)}, 2^{-4R}, \frac{2 \cdot 2^{-4R}}{3 - 2^{-4R}} \right). \qquad (40)$$

We will refer to this R-D quadruple as the splitting point, and the following theorem is now straightforward.

Theorem 4: For the unit variance Gaussian source, the splitting point R-D quadruple $(R, \frac{2 \cdot 2^{2R} + 4}{3(2^{2R} + 1)}, 2^{-4R}, \frac{2 \cdot 2^{-4R}}{3 - 2^{-4R}})$ is achievable, and it is outside $\mathcal{R}_{PPR}^{(3)}$.

Note that the splitting point R-D quadruple does not dominate the network coding point R-D quadruple in a component-wise manner, because the individual description distortion D_1 in the splitting scheme is not as good as in the network coding point, though the three description distortion D_3 in the splitting scheme is indeed better.

In the special case considered above, the quantization noises $N_{2,i}$'s are mutually independent, thus the quantization splitting technique used in the Gaussian CEO problem can be used without much change. However, since the MD problem is a centralized coding problem, the quantization noise can in fact be correlated in the PPR multilayer scheme, unlike in the distributed CEO problem where quantization noises are always independent in the Berger–Tung scheme; see [24] for related discussions. In fact, the depiction in Fig. 4 is no longer accurate in the general case, since the overall MD encoder cannot be divided into three separate encoders, and the encoding operation has to be better coordinated. With such correlated quantization noises, the technique used in the Gaussian CEO problem cannot be directly applied any more, because the choice of splitting random variables is now more constrained. In the next section, we show that even when the quantization noises are correlated, the quantization splitting technique can still be used. In fact, though we made the analogy to the Gaussian CEO problem to facilitate understanding here, the coding scheme outlined for Theorem 4 already does not rely on whether the encoders can operate in a separate manner.

Another notable difference between the quantization splitting used in the Gaussian CEO problem in [12] and that used here is

the decoding order. In the Gaussian CEO problem, since there is only one central decoder, we can employ an explicit sequential decoding order for quantization splitting. However, since there exist multiple decoders in the MD problem, it is difficult to find a completely sequential decoding order which is good for all the decoders, particularly when the number of descriptions is large. To circumvent this difficulty, we can alternatively view the coding scheme given for the above example as one designed for a four-description problem. Using this alternative interpretation, the conditions for the codebooks can be represented in a simpler manner, and we will explore this alternative route in the next section. This point of view is reminiscent to rate splitting in the K-user multiple access channel, where the rate splitting scheme can be understood as a scheme for more than K users, but some of the users are paired together.

Finally, we note that only one random variable is being split in the $(3, 2)$ SCEC case, and in general we shall split $k - 1$ random variables for any (K, k) SCEC in the next section. It is indeed possible to split more random variables; however, it appears impossible to do so without jeopardizing the original (K, k) SCEC performance, and thus we shall focus on quantization splitting that only splits $k - 1$ random variables.

D. Outline on Generalizing the Two Approaches

By considering the no two description excess rate case, we have provided two distinct approaches to boost the performance of symmetric MD coding, which yield R-D points outside $\mathcal{R}_{PPR}^{(3)}$. In the remainder of the paper, these two approaches are developed further in order to provide general coding schemes. We proceed in two steps: first single layer schemes similar to a single layer SCEC are considered, and then multilayer schemes are built.

IV. SINGLE LAYER SCHEMES FOR K DESCRIPTIONS

In this section, we generalize the approaches given in the previous section to the K-description case, however, only for single layer systems. Such single layer systems are important step stones toward general schemes, as in the PPR multilayer scheme for which the single layer SCEC is essential.

It is not immediately clear what the requirements are for a system to qualify as "single layer," since even for a single layer SCEC, more than one level of distortion can be considered and they can be traded off among each other. Indeed such a classification is rather informal, and cannot be defined in a precise manner. Nevertheless we provide a heuristic perspective which may convince the readers that the schemes given in this section are indeed "single layer." Note that a single (K, k) SCEC can be understood as a code for which any k descriptions can be used to achieve a certain distortion (even being jointly R-D optimal), and moreover, more than k descriptions (say $k + 1$ descriptions) can reduce the distortion even more. With an SCEC, fewer than k descriptions are useless, but the approaches given in the previous section have the property that fewer than k descriptions (more precisely, $k = 2$ in the previous section) can still be useful. Thus, we focus on the schemes with which any k descriptions can be used to satisfy a certain distortion constraint D_k, and at the same time we are also interested in the distortions

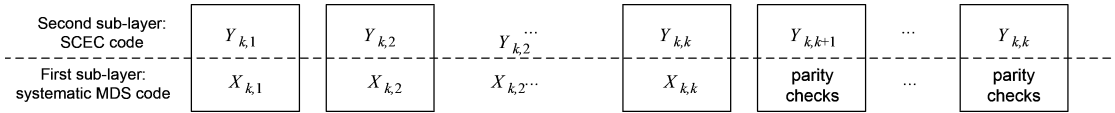| Second sub-layer: SCEC code	$Y_{k,1}$	$Y_{k,2}$	$Y_{k,2}^{\cdots}$	$Y_{k,k}$	$Y_{k,k+1}$	$\cdots$	$Y_{k,k}$
First sub-layer: systematic MDS code	$X_{k,1}$	$X_{k,2}$	$X_{k,2}\cdots$	$X_{k,k}$	parity checks	$\cdots$	parity checks

Fig. 5. The single layer coding structure based on systematic MDS channel code and SCEC. Each rectangle stands for a description.

achieved with $k-1$ and $k+1$ descriptions, however, to a some-what lesser degree; the R-D tuple is thus conveniently denoted as $(R, D_{k-1}, D_k, D_{k+1})$ with other distortions ignored. With the single layer codes ready, the construction of general multi-layer schemes become rather natural in the next section.

Following this heuristic perspective, we will now focus on the combination of channel codes with SCEC for the first approach, and the generalization of quantization splitting scheme under more relaxed quantization noise dependence structure for the second approach. One particularly interesting question is whether such codes can achieve the extremal R-D tuples where any k descriptions are jointly optimal, at least for the Gaussian source, and we will also treat this aspect in some detail; similar to the condition of no two description excess rate, we refer to this condition as no k description excess rate. In the remainder of the paper, familiarity with the properties of the typical sequences is assumed, and familiarity with [2], [6], and [7] is particularly helpful.

A. Single Layer Scheme Based on Channel Codes and SCEC

Let us first recall the no two description excess rate case for $K = 3$, i.e., $D_2 = 2^{-4R}$ for the Gaussian source, and consider tradeoff between the distortions D_1 and D_3. From the result given in the previous section, we know that the network coding point is outside $\mathcal{R}_{PPR}^{(3)}$. Thus, a direct time-sharing between the network coding point and the binning point can achieve certain (D_1, D_3) tradeoff, under the no two description excess rate requirement. However, by strategically combining the coding schemes, instead of time-sharing them, we can construct a scheme with an even better performance.

Consider the K-description case: the main idea is to encode the source in two sublayers. Roughly speaking, the first sublayer uses a (K, k) systematic maximum distance separable (MDS) code on certain source coding component, and the second sublayer then encodes the quantization noise from the first layer using a (K, k) SCEC. Most of the components being standard with existing technique [6], the only part that is not clear yet is on what source coding component the (K, k) MDS code should be used. Recall that our goal here is similar to that in the three description case, but instead of requiring some individual descriptions to be useful, we require some combination of $(k-1)$ descriptions to be useful. Since the source coding component has k portions before the (K, k) MDS coding step, it is natural to use a $(k, k-1)$ SCEC for the source coding component in this sublayer. Now the scheme is rather straightforward, at least for the Gaussian case (see Fig. 5).

- We first construct a $(k, k-1)$ SCEC on which a (K, k) systematic MDS channel code is applied to yield the first sublayer;
- Then the quantization noise is further encoded using a (K, k) SCEC as the second sublayer.

Next we provide an outline of the general scheme through random coding, which helps to avoid the explicit discussion of quantization noise. Though considering quantization noise in the Gaussian case is meaningful, it can be confusing for general discrete memoryless sources. We summarize the result in the following theorem, after a necessary definition.

Definition 4: A joint distribution $p(\bar{y}_{k,\mathcal{I}_K}, \bar{x}_{k,\mathcal{I}_k}|x)$ is called *mixed symmetric* if for any permutations $\pi(\cdot) : \mathcal{I}_K \to \mathcal{I}_K$ and $\pi_k(\cdot) : \mathcal{I}_k \to \mathcal{I}_k$ the joint distribution $p(\bar{y}_{k,\pi(\mathcal{I}_K)}, \bar{x}_{k,\pi_k(\mathcal{I}_k)}|x)$ is the same as $p(\bar{y}_{k,\mathcal{I}_K}, \bar{x}_{k,\mathcal{I}_k}|x)$.

Let D_o be the distortion achieved using a conventional R-D code at $R = 0$, i.e., $D_o = \min_{x_o:x_o \in \mathcal{X}} \mathbb{E}d(X, x_o)$. We have Theorem 5 given below. Note that though Theorem 5 is given only for discrete sources, the result holds for the Gaussian source by replacing entropies with differential entropies, replacing the discrete alphabets by reals and replacing the bounded distortion measure by the squared error distortion measure; this remark also holds for the theorems presented in later sections though they are also given only for discrete sources.

Theorem 5: For any $2 \le k \le K-1$, any discrete memoryless source X, any mixed symmetric distribution $p(\bar{y}_{k,\mathcal{I}_K}, \bar{x}_{k,\mathcal{I}_k}|x)$ over finite alphabets $\mathcal{Y}_k^K \times \mathcal{X}_k^k$

$$p(x, \bar{y}_{k,\mathcal{I}_K}, \bar{x}_{k,\mathcal{I}_k}) = p(\bar{y}_{k,\mathcal{I}_K}, \bar{x}_{k,\mathcal{I}_k}|x)p(x) \quad (41)$$

and a set of decoding functions

$$g_\mathcal{A} : \mathcal{X}_k^{k-1} \to \mathcal{X}, \quad |\mathcal{A}| = k - 1, \quad (42)$$
$$g_\mathcal{B} : \mathcal{Y}_k^k \times \mathcal{X}_k^k \to \mathcal{X}, \quad |\mathcal{B}| = k, \quad (43)$$
$$g_\mathcal{C} : \mathcal{Y}_k^{k+1} \times \mathcal{X}_k^k \to \mathcal{X}, \quad |\mathcal{C}| = k + 1 \quad (44)$$

such that

$$\lambda_{k-1}^* \sum_{\mathcal{A}:|\mathcal{A}|=k-1, \mathcal{A} \subset \mathcal{I}_k} \mathbb{E}d(X, g_\mathcal{A}(\bar{X}_{k,\mathcal{A}}))$$
$$+ (1 - k\lambda_{k-1}^*)D_o \le D_{k-1}, \quad (45)$$
$$\mathbb{E}d(X, g_\mathcal{B}(\bar{X}_{k,\mathcal{I}_k}, \bar{Y}_{k,\mathcal{B}})) \le D_k, \quad |\mathcal{B}| = k, \quad (46)$$
$$\mathbb{E}d(X, g_\mathcal{C}(\bar{X}_{k,\mathcal{I}_k}, \bar{Y}_{k,\mathcal{C}})) \le D_{k+1}, \quad |\mathcal{C}| = k + 1 \quad (47)$$

where $\lambda_{k-1}^* = \binom{K}{k-1}^{-1}$, the symmetric R-D tuple $(R, D_{k-1}, D_k, D_{k+1})$ is achievable where

$$R \ge \frac{1}{k-1}H(\bar{X}_{k,\mathcal{I}_{k-1}}) - \frac{1}{k}H(\bar{X}_{k,\mathcal{I}_k}|X)$$
$$+ \frac{1}{k}H(\bar{Y}_{k,\mathcal{I}_k}|\bar{X}_{k,\mathcal{I}_k}) - \frac{1}{K}H(\bar{Y}_{k,\mathcal{I}_K}|X, \bar{X}_{k,\mathcal{I}_k}). \quad (48)$$

Remark: We use $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ to label the decoding functions with $k-1$ descriptions, k descriptions, and $k+1$ descriptions, respectively. In Theorem 5, there is in fact no need to distinguish the decoding functions $g_\mathcal{B}(\cdot)$ (as well as $g_\mathcal{C}(\cdot)$) by their subscript

$\mathcal{B}$ (as well as $\mathcal{C}$) among those sets with the same cardinality, i.e., they can be written as $g_{\mathcal{I}_k}(\cdot)$ and $g_{\mathcal{I}_{k+1}}(\cdot)$, respectively, in (43) and (44); this is because the symmetric distribution requirement automatically ensures that a single decoding function can be used. However, we keep the current form in this simple single layer setting for better clarity.

Proof: We provide an outline for the random coding scheme with some details omitted in order to emphasize the structure of the code, since the proof techniques to bound the error probability are similar to those used in [2] and [5]–[7]. Fix a joint distribution $p(x, \bar{y}_{k, \mathcal{I}_K}, \bar{x}_{k, \mathcal{I}_k})$, the coding scheme is as follows.

Codebook generation: For each $i \in \mathcal{I}_k$, a codebook of size $2^{n\hat{R}_k^x}$ is generated using the marginal distribution of $X_{k,i}$, denoted as $C_{k,i}^x$; similarly, for each $i \in \mathcal{I}_K$, a codebook of size $2^{n\hat{R}_k^y}$ is generated using the marginal distribution of $Y_{k,i}$, denoted as $C_{k,i}^y$.

Random binning: For $i \in \mathcal{I}_k$, each codeword in $C_{k,i}^x$ is randomly assigned into one of $2^{nR_k^x}$ bins; similarly for $i \in \mathcal{I}_K$, each codeword in $C_{k,i}^y$ is randomly assigned into one of $2^{nR_k^y}$ bins.

Encoding: For a typical source sequence x^n, the encoder first finds a codeword in each of the codebooks $C_{k,i}^x$, $i \in \mathcal{I}_k$, such that they are jointly typical with x^n; let us denote their bin indices as $S_{k,i}^x$, and assume these indices are written in a sufficiently large alphabet. The index $S_{k,i}^x$ is written in the ith description, $i \in \mathcal{I}_k$, as the x sublayer. A systematic (K,k) MDS code is then applied on $S_{k,i}^x$, $i \in \mathcal{I}_k$, and the parity check portion is split into $K - k$ parts as the x sublayer of the $(k+1)$th description to the Kth description.

Next the encoder finds a codeword in each of the codebooks $C_{k,i}^y$, $i \in \mathcal{I}_K$, such that they are jointly typical with the source sequence x^n and the previously found codewords in x sublayer; the bin indices of these codewords are written in the y sublayer of the ith description.

The $C_{k,i}^x$ codebooks should be sufficiently large such that for any typical source sequence x^n, we can find jointly typical codewords; using the proof technique in [2], [5], [6] (and ignoring the small quantities δ's as aforementioned), we only need

$$\hat{R}_k^x \geq H(X_{k,1}) - \frac{1}{k}H(\bar{X}_{k,\mathcal{I}_k}|X). \tag{49}$$

Similarly the $C_{k,i}^y$ codebooks should be sufficiently large such that for any typical source sequence x^n and the codewords previously found in x sublayer, we can find jointly typical codewords; for this we only need

$$\hat{R}_k^y \geq H(Y_{k,1}) - \frac{1}{K}H(\bar{Y}_{k,\mathcal{I}_K}|X, \bar{X}_{k,\mathcal{I}_k}). \tag{50}$$

Other events incur an error, and a special symbol can be sent to signal it.

Decoding: When any $k - 1$ descriptions in the set $\mathcal{A}$ are received, such that $\mathcal{A} \subset \mathcal{I}_k$, the x sublayer bin indices in these descriptions are available, and the decoder finds in each of these bins within the respective codebooks a unique codeword such that they are jointly typical; if there is more than one choice, an error is declared. Then the decoding function $g_{\mathcal{A}}(\cdot)$ is invoked on these codewords. For any other combination of $(k-1)$ de-

scriptions, the source sequence is reconstructed using a default letter. When any k descriptions in the set $\mathcal{B}$ are received, i.e., $|\mathcal{B}| = k$, the x sublayer is decoded by decoding the (K, k) MDS channel code to recover the bin indices, then finding in each of these bins a unique codeword such that they are jointly typical; if there is more than one choice, an error is declared. Next the y sublayer is decoded by finding a unique codeword in each of y sublayer bins specified, such that they are jointly typical, also jointly typical with the decoded x sublayer codewords previously found. The decoding function $g_{\mathcal{B}}(\cdot)$ is then used. The decoding procedure for $k + 1$ descriptions is similar to the case of k descriptions. Clearly, the size of the bins cannot be too large, and we need to ensure

$$(k-1)(\hat{R}_k^x - R_k^x) \leq (k-1)H(X_{k,1}) - H(\bar{X}_{k,\mathcal{I}_{k-1}}), \tag{51}$$

$$k(\hat{R}_k^y - R_k^y) \leq kH(Y_{k,1}) - H(\bar{Y}_{k,\mathcal{I}_k}|\bar{X}_{k,\mathcal{I}_k}) \tag{52}$$

for the x sublayer and y sublayer, respectively.

With (49)–(52), it is not difficult to see the rate given in (48) is indeed achievable. We only need to verify that the distortion expression given in (45) is achievable, but this is obvious by observing that it is the result of time-sharing the distortions induced by the two kinds of $(k-1)$ description combinations. The proof is thus complete. ∎

Next let us return to the Gaussian source, and more specifically consider the no k description excess rate case. For the x sublayer, we can use the result for the $(k, k-1)$ SCEC under the no k description excess rate condition given in [6]; and for the y sublayer, we only need to consider encoding the quantization noise $X - \mathbb{E}(X|\bar{X}_{k,\mathcal{I}_k})$, and apply the result for the (K, k) SCEC also under the no k description excess rate in [6]. Let the rate used in the x sublayer be R_a per description, and the remaining rate $R - R_a$ be used in the y sublayer, it is then straightforward to use (12) to derive that

$$D_k = 2^{-2kR}, \quad D_{k+1} = \frac{k \cdot 2^{-2kR}}{k + 1 - 2^{-2k(R-R_a)}}. \tag{53}$$

Using (13) and (45) for the first layer $(k, k-1)$ SCEC, we can also derive that

$$\begin{aligned} D_{k-1} &= \frac{k}{\binom{K}{k-1}}\left(1 - \frac{k-1}{k}(1 - 2^{-2kR_a})\right) \\ &\quad + \left(1 - \frac{k}{\binom{K}{k-1}}\right) \\ &= 1 - \frac{k-1}{\binom{K}{k-1}}(1 - 2^{-2kR_a}). \end{aligned} \tag{54}$$

Using (53) and (54), it is easy to check that in this scheme D_{k+1} is a convex function of D_{k-1} for a fixed R, and thus the performance of this scheme is better than a direct time-sharing between the MDS channel code scheme and the SCEC scheme; the extreme schemes for $K = 3$ in fact lead to the network coding point $(R_a = R)$ and the binning point $(R_a = 0)$. In Fig. 6, we plot the tradeoff between D_1 and D_3 for the no two description excess rate case using the combined scheme given above. It is seen that the tradeoff curve is convex, implying the performance
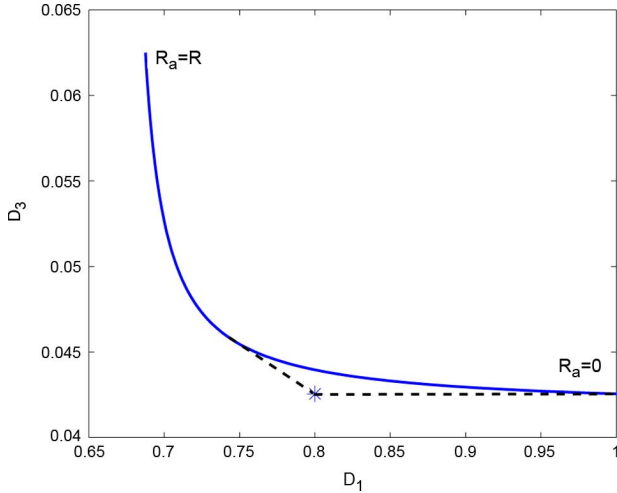
Fig. 6. The solid line is the tradeoff between D_1 and D_3 using the combined scheme of channel codes and SCEC, under the no two description excess rate requirement for the unit variance Gaussian source with $R = 1$. The star denotes the splitting point, and the dashed line forms a portion of the convex hull of the distortion region achievable by the two methods.

is better than a direct time-sharing. The single point labeled with a star is the splitting point as given in (39).

The distortion D_{k-1} for using $k-1$ descriptions can be further improved. An interesting effect observed in [6] is that though a (K,k) SCEC is designed to be only decodable when at least k descriptions are available, for certain special cases fewer than k descriptions are also decodable. For the no k description excess rate considered above, this observation can be used. In fact, it is clearly seen that for the R-D tuple in (13), any number of descriptions is useful. Using a code achieving the performance in (13) in the x sublayer of the given scheme, let us assume, for example, the first $k-2$ descriptions are received, but another description in the last $K-k$ description is also available which has only parity check portion, then it is clear that a distortion reduction can still be achieved by using the x sublayer of the first $k-2$ descriptions. Taking this effect into consideration, for the Gaussian source the distortion D_{k-1} can be further reduced to those given in (55), shown at the bottom of the page, where

$$D^*_{k-1,\mathcal{I}_j} \triangleq 1 - \frac{j}{k}(1 - 2^{-2kR_a}) \qquad (56)$$

and for convenience we define $\binom{m}{n} \triangleq 0$ if $m < n$ and $\binom{m}{0} \triangleq 1$. However, note that such a phenomenon that fewer than $k-1$ descriptions can be decodable is only a byproduct of the special R-D point, but not a mechanism inherent in the given coding scheme. Thus, from here on, we will not consider this aspect

further. Nevertheless, it is worth noting that this implies an increased difficulty in finding an optimal scheme, since an extra tension appears to exist between the general coding scheme and the scheme for some special R-D points.

Readers may wonder whether the quantization splitting technique can also be combined together with the systematic MDS channel codes and SCEC to yield an even better performance. Indeed, from the plot for $K = 3$ shown in Fig. 6, it is clear that even a simple time-sharing of the scheme given above and the splitting point scheme given in the previous section can provide further performance improvement. There are in fact many ways to combine various techniques; however, we did not find a satisfactory combination which leads to a natural multilayer scheme. Moreover, such combinations can quickly become overly complicated. Thus, in this work, we instead only present the schemes with clear structures that we believe are most natural, and forgo the task of finding a complete general scheme with all the possible components.

B. Single Layer Scheme Based on Quantization Splitting

Now we consider a single layer system based on the quantization splitting approach. From the no two description excess rate case, we see that one key difficulty is to find some appropriate splitting random variables. The space of the splitting random variables is constrained, and only within this space, the quantization splitting approach leads to meaningful coding schemes. The following definition formalizes these constraints; for convenience, we assume $K \geq 3$ and $2 \leq k \leq K - 1$, since the other cases are degenerate.

Definition 5: Let $p(\bar{y}_{k,\mathcal{I}_K}|x)$ be a symmetric distribution. The random variables $\bar{X}_{k,\mathcal{I}_{k-1}}$ are *admissible for symmetric splitting* with respect to $p(\bar{y}_{k,\mathcal{I}_K}|x)$, if they are jointly distributed with X and $\bar{Y}_{k,\mathcal{I}_K}$ such that $X_{k,i} \leftrightarrow Y_{k,i} \leftrightarrow (X, \bar{Y}_{k,\mathcal{I}_K\backslash\{i\}}, \bar{X}_{k,\mathcal{I}_{k-1}\backslash\{i\}})$ is a Markov string and $p(x_{k,i}|y_{k,i})$ is the same for all $i = 1, 2, \ldots, k-1$; furthermore, the following conditions are satisfied:

$$\frac{l}{k-1}H(\bar{X}_{k,\mathcal{I}_{k-1}}|X) - H(\bar{X}_{k,\mathcal{I}_l}|\bar{Y}_{k,\mathcal{I}_{l+m}\backslash\mathcal{I}_l}, X)$$
$$\leq H(\bar{Y}_{k,\mathcal{I}_m}|X) - \frac{m}{K}H(\bar{Y}_{k,\mathcal{I}_K}|X),$$
$$0 \leq l \leq k-1, \ 0 \leq l+m \leq K \qquad (57)$$

and

$$\frac{l}{k-1}I(\bar{Y}_{k,\mathcal{I}_{k-1}};\bar{X}_{k,\mathcal{I}_{k-1}})$$
$$\leq \frac{m}{k}H(\bar{Y}_{k,\mathcal{I}_k}) - H(\bar{Y}_{k,\mathcal{I}_m}|\bar{Y}_{k,\mathcal{I}_k\backslash\mathcal{I}_m}, \bar{X}_{k,\mathcal{I}_l}),$$
$$0 \leq l \leq m < k \qquad (58)$$

$$D^*_{k-1} = \frac{1}{\binom{K}{k-1}}\left(\sum_{j=\max(1,2k-K-1)}^{k-1}\binom{k}{j}\binom{K-k}{k-j-1}D^*_{k-1,\mathcal{I}_j} + \binom{K-k}{k-1}D_o\right) \qquad (55)$$
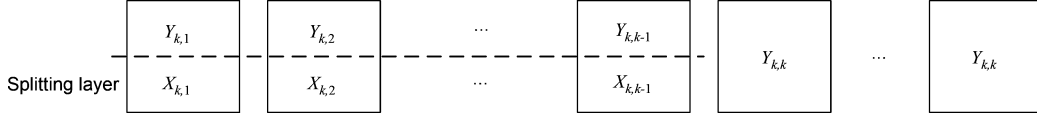
Fig. 7.   The coding structure for the single layer scheme based on quantization splitting.

where l, m are nonnegative integers, and

$$\frac{1}{k-1}I(\bar{X}_{k,\mathcal{I}_{k-1}};X) \leq \frac{1}{k}H(\bar{Y}_{k,\mathcal{I}_k}) - \frac{1}{K}H(\bar{Y}_{k,\mathcal{I}_K}|X). \quad (59)$$

It is not immediately clear whether there indeed exist nontrivial random variables that are admissible for symmetric splitting. After the proof the following theorem is given, we will show that the space of random variables that are admissible for symmetric splitting is generally not trivial or empty, which is important because otherwise the theorem would be equivalent to the known result in [6].

Theorem 6: Let $p(\bar{y}_{k,\mathcal{I}_K}|x)$ be a symmetric distribution, and let $\bar{X}_{k,\mathcal{I}_{k-1}}$ be random variables admissible for symmetric splitting with respect to $p(\bar{y}_{k,\mathcal{I}_K}|x)$. If there exist decoding functions

$$g_{\mathcal{I}_{k-1}} : \mathcal{X}_k^{k-1} \to \mathcal{X},$$
$$g_{\mathcal{B}} : \mathcal{Y}_k^k \to \mathcal{X}, \quad |\mathcal{B}| = k,$$
$$g_{\mathcal{C}} : \mathcal{Y}_k^{k+1} \to \mathcal{X}, \quad |\mathcal{C}| = k+1$$

such that

$$\lambda_{k-1}^* \mathbb{E}d(X, g_{\mathcal{I}_{k-1}}(\bar{X}_{k,\mathcal{I}_{k-1}})) + (1 - \lambda_{k-1}^*)D_o \leq D_{k-1}, \quad (60)$$

$$\mathbb{E}d(X, g_{\mathcal{B}}(\bar{Y}_{k,\mathcal{B}})) \leq D_k, \qquad |\mathcal{B}| = k, \qquad (61)$$

$$\mathbb{E}d(X, g_{\mathcal{C}}(\bar{Y}_{k,\mathcal{C}})) \leq D_{k+1}, \qquad |\mathcal{C}| = k+1 \qquad (62)$$

where again $\lambda_{k-1}^* = \binom{K}{k-1}^{-1}$, the symmetric R-D tuple $(R, D_{k-1}, D_k, D_{k+1})$ is achievable where

$$R \geq \frac{1}{k}H(\bar{Y}_{k,\mathcal{I}_k}) - \frac{1}{K}H(\bar{Y}_{k,\mathcal{I}_K}|X). \quad (63)$$

Remark: By letting the random variables $\bar{X}_{k,\mathcal{I}_{k-1}}$ be constants, we almost recover the main theorem (Theorem 1) in [6], with the only difference that the distortions $D_{k+2}, \ldots, D_K$ are chosen to be ignored in Theorem 6. But as long as the splitting random variables $\bar{X}_{k,\mathcal{I}_{k-1}}$ are not trivial, Theorem 6 strictly improves upon the performance of a (K, k) SCEC.

In the proof of this theorem, we give an outline for the random coding scheme, but again do not bound the error probability rigorously. As aforementioned, this coding scheme can be understood as a scheme for $K+k-1$ descriptions, but some of the descriptions are paired together; see Fig. 7 for an illustration. Our focus is on reducing the rather involved conditions into more manageable ones. We will initially describe the coding scheme with a more relaxed symmetric requirement and then specialize it to the result given in the theorem; the scheme itself is in fact more general as we will discuss after the proof.

Proof: First fix the joint distribution of $(X, \bar{Y}_{k,\mathcal{I}_K}, \bar{X}_{k,\mathcal{I}_{k-1}})$. For each random variable $X_{k,i}$, $i \in \mathcal{I}_{k-1}$, a codebook of size $2^{n\hat{R}_{k,i}^x}$ is generated using the marginal distribution of $X_{k,i}$; similarly, for each random variable $Y_{k,i}$, $i \in \mathcal{I}_K$, a codebook of size $2^{n\hat{R}_{k,i}^y}$ is generated using the marginal distribution of $Y_{k,i}$. Denote these codebooks as $C_{k,i}^x$ and $C_{k,i}^y$, respectively. For each single $C_{k,i}^x$ codebook, $i \in \mathcal{I}_{k-1}$, each of its codewords is uniformly at random assigned into $2^{nR_{k,i}^x}$ bins; similarly for each single $C^{y_{k,i}}$ codebook, $i \in \mathcal{I}_K$, each of its codewords is uniformly at random assigned into $2^{nR_{k,i}^y}$ bins.

During encoding, the encoder finds in each of these codebooks a codeword such that they are jointly typical with the (typical) source sequence as well as with each other. For the ith description, $i \in \mathcal{I}_{k-1}$, the bin index of the found codeword in the codebook $C_{k,i}^x$, and the bin index of the found codeword in the codebook $C_{k,i}^y$ are concatenated as the information for this description; for the jth description, $j \in \mathcal{I}_K \backslash \mathcal{I}_{k-1}$, the bin index of the found codeword in codebook $C_{k,j}^y$ is used as the information for this description.

When $k-1$ descriptions are received, if they are those in $\mathcal{I}_{k-1}$, the decoder finds a unique tuple of codewords, one in each bin specified by the bin index within the codebook $C_{k,i}^x$, $k \in \mathcal{I}_{k-1}$, such that they are jointly typical; if there is more than one such tuple, an error is declared. For other $k-1$ description combinations, the decoder uses the default letter as its reconstruction. When any k descriptions are received, which are those in the set $\mathcal{S} \cup \mathcal{T}$, where $\mathcal{S} \subseteq \mathcal{I}_{k-1}$ and $\mathcal{T} \subseteq \mathcal{I}_K \backslash \mathcal{I}_{k-1}$, the decoder tries to find a unique tuple of codewords, one in each bin specified by the bin indices within $C_{k,i}^x$ and $C_{k,i}^y$, $i \in \mathcal{S}$, and one in each bin specified by the bin indices within $C_{k,j}^y$, $j \in \mathcal{T}$, such that they are jointly typical; if there is more than one such tuple, an error is declared. When any $k+1$ descriptions are received, two k-description decoders can be used to find the intended codewords. Given the above outline, it is clear that the rates for the descriptions are given by

$$R_{k,i} = R_{k,i}^x + R_{k,i}^y, \quad i \in \mathcal{I}_{k-1}, \qquad (64)$$
$$R_{k,i} = R_{k,i}^y, \quad i \in \mathcal{I}_K \backslash \mathcal{I}_{k-1}. \qquad (65)$$

In this scheme, the codebook sizes should be sufficiently large, such that jointly typical codewords can be found with high probability. The sizes of the bins, on the other hand, should be sufficiently small, such that the joint typicality decoders succeed with high probability. Using the analysis given in [6], and taking the above coding scheme as one for $K + k - 1$ descriptions, where some of the descriptions are paired together, we have the following constraints on $(\hat{R}_{k,1}^x, \ldots, \hat{R}_{k,k-1}^x, \hat{Y}_{k,1}^y, \ldots, \hat{R}_{k,K}^y) \in \mathbb{R}_+^{K+k-1}$ and

$(R_{k,1}^x, \ldots, R_{k,k-1}^x, R_{k,1}^y, \ldots, R_{k,K}^y) \in \mathbb{R}_+^{K+k-1}$ (ignoring the small quantities δ's).

1) Codebook constraints:

$$\sum_{i \in \mathcal{A}} \hat{R}_{k,i}^y + \sum_{i \in \mathcal{B}} \hat{R}_{k,i}^x$$
$$\geq \sum_{i \in \mathcal{A}} H(Y_{k,i}) + \sum_{i \in \mathcal{B}} H(X_{k,i}) - H(\bar{Y}_{k,\mathcal{A}}, \bar{X}_{k,\mathcal{B}}|X),$$
$$\mathcal{A} \subseteq \mathcal{I}_K, \mathcal{B} \subseteq \mathcal{I}_{k-1}. \qquad (66)$$

2) Binning constraints for k description decoders: for any $\mathcal{A}$ such that $|\mathcal{A}| = k$, let $\mathcal{B} = \mathcal{A} \cap \mathcal{I}_{k-1}$, then

$$0 \leq \sum_{i \in \mathcal{A}'} (\hat{R}_{k,i}^y - R_{k,i}^y) + \sum_{i \in \mathcal{B}'} (\hat{R}_{k,i}^x - R_{k,i}^x)$$
$$\leq \sum_{i \in \mathcal{A}'} H(Y_{k,i}) + \sum_{i \in \mathcal{B}'} H(X_{k,i})$$
$$\quad - H(\bar{Y}_{k,\mathcal{A}'}, \bar{X}_{k,\mathcal{B}'}|\bar{Y}_{k,\mathcal{A}\backslash\mathcal{A}'}, \bar{X}_{k,\mathcal{B}\backslash\mathcal{B}'}),$$
$$\text{any } \mathcal{A}' \subseteq \mathcal{A}, \mathcal{B}' \subseteq \mathcal{B}. \qquad (67)$$

3) Binning constraints for the $k-1$ description decoder:

$$0 \leq \sum_{i \in \mathcal{B}} (\hat{R}_{k,i}^x - R_{k,i}^x)$$
$$\leq \sum_{i \in \mathcal{B}} H(X_{k,i}) - H(\bar{X}_{k,\mathcal{B}}|\bar{X}_{k,\mathcal{I}_{k-1}\backslash\mathcal{B}}),$$
$$\text{any } \mathcal{B} \subseteq \mathcal{I}_{k-1}. \qquad (68)$$

Until this point, we have not introduced much symmetry into the coding scheme yet, and the constraints given above are completely general without such a requirement. Next, we choose the following rates with certain symmetry built in. The codebook rates are set as

$$\hat{R}_{k,i}^x = \hat{R}_k^x \triangleq H(X_{k,1}) - \frac{1}{k-1} H(\bar{X}_{k,\mathcal{I}_{k-1}}|X),$$
$$i \in \mathcal{I}_{k-1}, \qquad (69)$$
$$\hat{R}_{k,i}^y = \hat{R}_k^{y*} \triangleq H(Y_{k,1}) - \frac{1}{K} H(\bar{Y}_{k,\mathcal{I}_K}|X)$$
$$\quad + \frac{1}{k-1} I(\bar{X}_{k,\mathcal{I}_{k-1}}; \bar{Y}_{k,\mathcal{I}_{k-1}}|X), \quad i \in \mathcal{I}_{k-1}, \qquad (70)$$
$$\hat{R}_{k,i}^y = \hat{R}_k^y \triangleq H(Y_{k,1}) - \frac{1}{K} H(\bar{Y}_{k,\mathcal{I}_K}|X),$$
$$i \in \mathcal{I}_K \backslash \mathcal{I}_{k-1} \qquad (71)$$

and the binning rates are set as

$$R_{k,i}^x = R_k^x \triangleq \frac{1}{k-1} I(\bar{X}_{k,\mathcal{I}_{k-1}}; X), \quad i \in \mathcal{I}_{k-1} \qquad (72)$$
$$R_{k,i}^y = R_k^{y*} \triangleq \frac{1}{k} H(\bar{Y}_{k,\mathcal{I}_k}) - \frac{1}{K} H(\bar{Y}_{k,\mathcal{I}_K}|X)$$
$$\quad - \frac{1}{k-1} I(\bar{X}_{k,\mathcal{I}_{k-1}}; X), \quad i \in \mathcal{I}_{k-1}, \qquad (73)$$
$$R_{k,i}^y = R_k^y \triangleq \frac{1}{k} H(\bar{Y}_{k,\mathcal{I}_k})$$
$$\quad - \frac{1}{K} H(\bar{Y}_{k,\mathcal{I}_K}|X), \quad i \in \mathcal{I}_K \backslash \mathcal{I}_{k-1}. \qquad (74)$$

Our plan is to prove that when $\bar{X}_{k,\mathcal{I}_{k-1}}$ are admissible for symmetric splitting, the rates given in (69)–(74) are indeed nonnegative rates which satisfy the conditions (66)–(68), and

$$R_{k,i} = \frac{1}{k} H(\bar{Y}_{k,\mathcal{I}_k}) - \frac{1}{K} H(\bar{Y}_{k,\mathcal{I}_K}|X). \qquad (75)$$

Once this is proved, Theorem 6 is established.

To see that the rates (69)–(74) are indeed nonnegative, we only need to apply Han's inequalities [25] and the condition (59); it is also clear (75) is satisfied by (64) and (65). By using the exact same proof as in [6], it can be shown straightforwardly that (68) is indeed satisfied. Thus, we only need to focus on (66) and (67) from here on.

We first rewrite the condition (66) by substituting the given choice of $\hat{R}_{k,i}^x$ and $\hat{R}_{k,i}^y$. The conditions in (66) are then equivalent to that for any $\mathcal{S}_x$, $\mathcal{S}_y$ and $\mathcal{T}$ such that $\mathcal{S}_x \subseteq \mathcal{I}_{k-1}$, $\mathcal{S}_y \subseteq \mathcal{I}_{k-1}$ and $\mathcal{T} \subseteq \mathcal{I}_K \backslash \mathcal{I}_{k-1}$, the following condition holds:

$$\frac{|\mathcal{S}_y| + |\mathcal{T}|}{K} H(\bar{Y}_{k,\mathcal{I}_K}|X) + \frac{|\mathcal{S}_y|}{k-1} H(\bar{X}_{k,\mathcal{I}_{k-1}}|\bar{Y}_{k,\mathcal{I}_{k-1}})$$
$$\quad - \frac{|\mathcal{S}_y| - |\mathcal{S}_x|}{k-1} H(\bar{X}_{k,\mathcal{I}_{k-1}}|X)$$
$$\leq H(\bar{Y}_{k,\mathcal{S}_y}, \bar{Y}_{k,\mathcal{T}}|X) + H(\bar{X}_{k,\mathcal{S}_x}|\bar{Y}_{k,\mathcal{S}_y}, \bar{Y}_{k,\mathcal{T}}, X). \qquad (76)$$

Next we consider the two cases $|\mathcal{S}_x| < |\mathcal{S}_y|$ and $|\mathcal{S}_x| \geq |\mathcal{S}_y|$ separately. For the first case, we write

$$H(\bar{X}_{k,\mathcal{S}_x}|\bar{Y}_{k,\mathcal{S}_y}, \bar{Y}_{k,\mathcal{T}}, X) \overset{(a)}{\geq} H(\bar{X}_{k,\mathcal{S}_x}|\bar{Y}_{k,\mathcal{S}_y \cup \mathcal{S}_x}, \bar{Y}_{k,\mathcal{T}}, X)$$
$$\overset{(b)}{=} H(\bar{X}_{k,\mathcal{S}_x}|\bar{Y}_{k,\mathcal{S}_x})$$
$$\overset{(c)}{\geq} |\mathcal{S}_x| H(X_{k,1}|Y_{k,1}) \qquad (77)$$

where (a) is because conditioning reduces entropy, and (b) and (c) are by the symmetry and the Markov string $X_{k,i} \leftrightarrow Y_{k,i} \leftrightarrow (X, \bar{Y}_{k,\mathcal{I}_K \backslash \{i\}}, \bar{X}_{k,\mathcal{I}_{k-1}\backslash\{i\}})$; by the same Markov string, we have also

$$\frac{|\mathcal{S}_y|}{k-1} H(\bar{X}_{k,\mathcal{I}_{k-1}}|\bar{Y}_{k,\mathcal{I}_{k-1}}) = |\mathcal{S}_y| H(X_{k,1}|Y_{k,1}). \qquad (78)$$

Thus, if we can prove

$$\frac{|\mathcal{S}_y| + |\mathcal{T}|}{K} H(\bar{Y}_{k,\mathcal{I}_K}|X) + |\mathcal{S}_y| H(X_{k,1}|Y_{k,1})$$
$$\quad - \frac{|\mathcal{S}_y| - |\mathcal{S}_x|}{k-1} H(\bar{X}_{k,\mathcal{I}_{k-1}}|X)$$
$$\leq H(\bar{Y}_{k,\mathcal{S}_y}, \bar{X}_{k,\mathcal{T}}|X) + |\mathcal{S}_x| H(X_{k,1}|Y_{k,1}) \qquad (79)$$

then (76) clearly holds. But notice that

$$(|\mathcal{S}_x| - |\mathcal{S}_y|) H(X_{k,1}|Y_{k,1}) + \frac{|\mathcal{S}_y| - |\mathcal{S}_x|}{k-1} H(\bar{X}_{k,\mathcal{I}_{k-1}}|X)$$
$$= \frac{|\mathcal{S}_x| - |\mathcal{S}_y|}{k-1} H(\bar{X}_{k,\mathcal{I}_{k-1}}|\bar{Y}_{k,\mathcal{I}_{k-1}}, X)$$
$$\quad + \frac{|\mathcal{S}_y| - |\mathcal{S}_x|}{k-1} H(\bar{X}_{k,\mathcal{I}_{k-1}}|X)$$
$$= \frac{|\mathcal{S}_y| - |\mathcal{S}_x|}{k-1} I(\bar{X}_{k,\mathcal{I}_{k-1}}; \bar{Y}_{k,\mathcal{I}_{k-1}}|X)$$
$$\overset{(d)}{\geq} \frac{|\mathcal{S}_y| + |\mathcal{T}|}{K} H(\bar{Y}_{k,\mathcal{I}_K}|X) - H(\bar{Y}_{k,\mathcal{S}_y}, \bar{X}_{k,\mathcal{T}}|X) \qquad (80)$$

where (d) is true because the right-hand side is nonpositive due to Han's inequality and the symmetry in the distribution, but the left-hand side is nonnegative when $|\mathcal{S}_x| < |\mathcal{S}_y|$. Since (80) is equivalent to (79), it is clear that for the case $|\mathcal{S}_x| < |\mathcal{S}_y|$, (66) indeed holds with the chosen rates.

Next consider the case $|\mathcal{S}_x| \geq |\mathcal{S}_y|$, and rewrite the right-hand side of (76) using the chain rule and the Markov string $X_{k,i} \leftrightarrow Y_{k,i} \leftrightarrow (X, \bar{Y}_{k,\mathcal{I}_K \setminus \{i\}}, \bar{X}_{k,\mathcal{I}_{k-1} \setminus \{i\}})$ as

$$
\begin{aligned}
H(\bar{Y}_{k,\mathcal{S}_y}, & \bar{Y}_{k,\mathcal{T}}|X) + H(\bar{X}_{k,\mathcal{S}_x}|\bar{Y}_{k,\mathcal{S}_y}, \bar{Y}_{k,\mathcal{T}}, X) \\
&= H(\bar{Y}_{k,\mathcal{S}_y}, \bar{Y}_{k,\mathcal{T}}|X) \\
&\quad + H(\bar{X}_{k,\mathcal{S}_x \setminus \mathcal{S}_y}|\bar{Y}_{k,\mathcal{S}_y}, \bar{Y}_{k,\mathcal{T}}, X) \\
&\quad + H(\bar{X}_{k,\mathcal{S}_x \cap \mathcal{S}_y}|\bar{Y}_{k,\mathcal{S}_y}, \bar{Y}_{k,\mathcal{T}}, \bar{X}_{k,\mathcal{S}_x \setminus \mathcal{S}_y}, X) \\
&= H(\bar{Y}_{k,\mathcal{S}_y}, \bar{Y}_{k,\mathcal{T}}|X) \\
&\quad + H(\bar{X}_{k,\mathcal{S}_x \setminus \mathcal{S}_y}|\bar{Y}_{k,\mathcal{S}_y}, \bar{Y}_{k,\mathcal{T}}, X) \\
&\quad + |\mathcal{S}_x \cap \mathcal{S}_y| H(\bar{X}_{k,1}|\bar{Y}_{k,1}). \qquad (81)
\end{aligned}
$$

Note that (66) needs to hold for any $\mathcal{S}_x \subseteq \mathcal{I}_{k-1}$, and the left-hand side of (76) depends only on $|\mathcal{S}_x|$ for any given $\mathcal{S}_y$ and $\mathcal{T}$. This implies that for fixed $\mathcal{S}_y$, $\mathcal{T}$ and $|\mathcal{S}_x|$, we can limit our consideration to the choice of $\mathcal{S}_x$ that minimizes the right-hand side of (76), for which we can write

$$
\begin{aligned}
|\mathcal{S}_x \cap \mathcal{S}_y| & H(\bar{X}_{k,1}|\bar{Y}_{k,1}) + H(\bar{X}_{k,\mathcal{S}_x \setminus \mathcal{S}_y}|\bar{Y}_{k,\mathcal{S}_y}, \bar{Y}_{k,\mathcal{T}}, X) \\
&\geq |\mathcal{S}_y| H(\bar{X}_{k,1}|\bar{Y}_{k,1}) \\
&\quad + H(\bar{X}_{k,\mathcal{I}_{|\mathcal{S}_x| - |\mathcal{S}_y|}}|\bar{Y}_{k,\mathcal{I}_{|\mathcal{S}_x| + |\mathcal{T}|} \setminus \mathcal{I}_{|\mathcal{S}_x| - |\mathcal{S}_y|}}, X) \qquad (82)
\end{aligned}
$$

because the Markov string $X_{k,i} \leftrightarrow Y_{k,i} \leftrightarrow (X, \bar{Y}_{k,\mathcal{I}_K \setminus \{i\}}, \bar{X}_{k,\mathcal{I}_{k-1} \setminus \{i\}})$ implies

$$
H(\bar{X}_{k,1}|\bar{Y}_{k,1}) \leq H(\bar{X}_{k,1}|\bar{X}_{k,\mathcal{I}_{k-1} \setminus \{1\}}, \bar{Y}_{k,\mathcal{I}_K \setminus \{1\}}, X). \qquad (83)
$$

In other words, the dominant inequality in the set of inequalities of (76) is the case when $\mathcal{S}_y \subseteq \mathcal{S}_x$, and for this reason, as long as (76) holds for $\mathcal{S}_y \subseteq \mathcal{S}_x$, the condition (66) is satisfied with the given rates. With this observation and (82), it is clear that we only need the following inequality to hold:

$$
\begin{aligned}
\frac{|\mathcal{S}_y| + |\mathcal{T}|}{K} & H(\bar{Y}_{k,\mathcal{I}_K}|X) + \frac{|\mathcal{S}_x| - |\mathcal{S}_y|}{k-1} H(\bar{X}_{k,\mathcal{I}_{k-1}}|X) \\
&\leq H(\bar{Y}_{k,\mathcal{S}_y}, \bar{Y}_{k,\mathcal{T}}|X) \\
&\quad + H(\bar{X}_{k,\mathcal{I}_{|\mathcal{S}_x| - |\mathcal{S}_y|}}|\bar{Y}_{k,\mathcal{I}_{|\mathcal{S}_x| + |\mathcal{T}|} \setminus \mathcal{I}_{|\mathcal{S}_x| - |\mathcal{S}_y|}}, X). \qquad (84)
\end{aligned}
$$

Now define $l \triangleq |\mathcal{S}_x| - |\mathcal{S}_y|$ and $m \triangleq |\mathcal{S}_y| + |\mathcal{T}| = |\mathcal{S}_y \cup \mathcal{T}|$, and notice that $\mathcal{S}_y \cup \mathcal{T} \subseteq \mathcal{I}_K$ and $k - 1 \geq |\mathcal{S}_x| - |\mathcal{S}_y| \geq 0$, (84) is exactly the admissible condition given in (57).

Next we turn to the condition in (67). First note that $\hat{R}_k^y \geq R_k^y$, $\hat{R}_k^{y*} \geq R_k^{y*}$ and $\hat{R}_k^x \geq R_k^x$, which imply that the first inequality in (67) holds for the chosen rates. Let $\mathcal{S} \subseteq \mathcal{I}_{k-1}$ and $\mathcal{T} \subseteq \mathcal{I}_K \setminus \mathcal{I}_{k-1}$, such that $|\mathcal{S} \cup \mathcal{T}| = k$. After substituting the chosen rates, the condition (67) is equivalent to that for any such $\mathcal{S}$ and

$\mathcal{T}$, and their subsets $\mathcal{S}_x$, $\mathcal{S}_y$, $\mathcal{T}'$ such that $\mathcal{S}_x \subseteq \mathcal{S}$, $\mathcal{S}_y \subseteq \mathcal{S}$ and $\mathcal{T}' \subseteq \mathcal{T}$

$$
\begin{aligned}
\frac{|\mathcal{S}_x|}{k-1} & H(\bar{X}_{k,\mathcal{I}_{k-1}}) + \frac{|\mathcal{S}_y| + |\mathcal{T}'|}{k} H(\bar{Y}_{k,\mathcal{I}_k}) \\
&\quad - \frac{|\mathcal{S}_y|}{k-1} I(\bar{X}_{k,\mathcal{I}_{k-1}}; \bar{Y}_{k,\mathcal{I}_{k-1}}) \\
&\geq H(\bar{X}_{k,\mathcal{S}_x}, \bar{Y}_{k,\mathcal{S}_y}, \bar{Y}_{k,\mathcal{T}'}|\bar{X}_{k,\mathcal{S} \setminus \mathcal{S}_x}, \bar{Y}_{k,\mathcal{S} \setminus \mathcal{S}_y}, \bar{Y}_{k,\mathcal{T} \setminus \mathcal{T}'}) \\
&= H(\bar{X}_{k,\mathcal{S}_x}|\bar{Y}_{k,\mathcal{S}}, \bar{Y}_{k,\mathcal{T}}, \bar{X}_{k,\mathcal{S} \setminus \mathcal{S}_x}) \\
&\quad + H(\bar{Y}_{k,\mathcal{S}_y}, \bar{Y}_{k,\mathcal{T}'}|\bar{X}_{k,\mathcal{S} \setminus \mathcal{S}_x}, \bar{Y}_{k,\mathcal{S} \setminus \mathcal{S}_y}, \bar{Y}_{k,\mathcal{T} \setminus \mathcal{T}'}) \\
&= |\mathcal{S}_x| H(X_{k,1}|Y_{k,1}) \\
&\quad + H(\bar{Y}_{k,\mathcal{S}_y}, \bar{Y}_{k,\mathcal{T}'}|\bar{X}_{k,\mathcal{S} \setminus \mathcal{S}_x}, \bar{Y}_{k,\mathcal{S} \setminus \mathcal{S}_y}, \bar{Y}_{k,\mathcal{T} \setminus \mathcal{T}'}) \qquad (85)
\end{aligned}
$$

where the last two equalities are due to the chain rule and the Markov string $X_{k,i} \leftrightarrow Y_{k,i} \leftrightarrow (X, \bar{Y}_{k,\mathcal{I}_K \setminus \{i\}}, \bar{X}_{k,\mathcal{I}_{k-1} \setminus \{i\}})$. We thus arrive at the equivalent condition

$$
\begin{aligned}
\frac{|\mathcal{S}_x| - |\mathcal{S}_y|}{k-1} & H(\bar{X}_{k,\mathcal{I}_{k-1}}) + (|\mathcal{S}_y| - |\mathcal{S}_x|) H(X_{k,1}|Y_{k,1}) \\
&\geq H(\bar{Y}_{k,\mathcal{S}_y}, \bar{Y}_{k,\mathcal{T}'}|\bar{X}_{k,\mathcal{S} \setminus \mathcal{S}_x}, \bar{Y}_{k,\mathcal{S} \setminus \mathcal{S}_y}, \bar{Y}_{k,\mathcal{T} \setminus \mathcal{T}'}) \\
&\quad - \frac{|\mathcal{S}_y| + |\mathcal{T}'|}{k} H(\bar{Y}_{k,\mathcal{I}_k}). \qquad (86)
\end{aligned}
$$

However, notice that

$$
\begin{aligned}
\frac{|\mathcal{S}_x| - |\mathcal{S}_y|}{k-1} & H(\bar{X}_{k,\mathcal{I}_{k-1}}) + (|\mathcal{S}_y| - |\mathcal{S}_x|) H(X_{k,1}|Y_{k,1}) \\
&= \frac{|\mathcal{S}_x| - |\mathcal{S}_y|}{k-1} I(\bar{X}_{k,\mathcal{I}_{k-1}}; \bar{Y}_{k,\mathcal{I}_{k-1}}) \qquad (87)
\end{aligned}
$$

and

$$
\begin{aligned}
H(\bar{Y}_{k,\mathcal{S}_y}, & \bar{Y}_{k,\mathcal{T}'}|\bar{X}_{k,\mathcal{S} \setminus \mathcal{S}_x}, \bar{Y}_{k,\mathcal{S} \setminus \mathcal{S}_y}, \bar{Y}_{k,\mathcal{T} \setminus \mathcal{T}'}) \\
&\quad - \frac{|\mathcal{S}_y| + |\mathcal{T}'|}{k} H(\bar{Y}_{k,\mathcal{I}_k}) \\
&\overset{(e)}{\leq} H(\bar{Y}_{k,\mathcal{S}_y}, \bar{Y}_{k,\mathcal{T}'}|\bar{Y}_{k,\mathcal{S} \setminus \mathcal{S}_y}, \bar{Y}_{k,\mathcal{T} \setminus \mathcal{T}'}) \\
&\quad - \frac{|\mathcal{S}_y| + |\mathcal{T}'|}{k} H(\bar{Y}_{k,\mathcal{I}_k}) \overset{(f)}{\leq} 0 \qquad (88)
\end{aligned}
$$

where (e) is because conditioning reduces entropy and (f) is due to Han's inequality. Thus, (86) is satisfied when $|\mathcal{S}_x| > |\mathcal{S}_y|$, and we only need to consider the case $|\mathcal{S}_x| \leq |\mathcal{S}_y|$ in the sequel. For this case notice that for any fixed $|\mathcal{S}_x|, |\mathcal{S}_y|, |\mathcal{T}'|$

$$
\begin{aligned}
H(\bar{Y}_{k,\mathcal{S}_y}, & \bar{Y}_{k,\mathcal{T}'}|\bar{X}_{k,\mathcal{S} \setminus \mathcal{S}_x}, \bar{Y}_{k,\mathcal{S} \setminus \mathcal{S}_y}, \bar{Y}_{k,\mathcal{T} \setminus \mathcal{T}'}) \\
&= H(\bar{Y}_{k,\mathcal{S}_y}, \bar{Y}_{k,\mathcal{T}'}|\bar{X}_{k,(\mathcal{S} \setminus \mathcal{S}_x) \cap \mathcal{S}_y}, \bar{Y}_{k,\mathcal{S} \setminus \mathcal{S}_y}, \bar{Y}_{k,\mathcal{T} \setminus \mathcal{T}'}) \\
&\leq H(\bar{Y}_{k,\mathcal{I}_{|\mathcal{S}_y| + |\mathcal{T}'|}}|\bar{X}_{k,\mathcal{I}_{|\mathcal{S}_y| - |\mathcal{S}_x|}}, \bar{Y}_{k,\mathcal{I}_k \setminus \mathcal{I}_{|\mathcal{S}_y| + |\mathcal{T}'|}}) \qquad (89)
\end{aligned}
$$

where the inequality is by observing that the left-hand side of (89) is maximized when the set $(\mathcal{S} \setminus \mathcal{S}_x) \cap \mathcal{S}_y$ is minimized, i.e., when $\mathcal{S}_x \subseteq \mathcal{S}_y$ (noting also the symmetry in the distribution). By defining $l \triangleq |\mathcal{S}_y| - |\mathcal{S}_x|$, and $m \triangleq |\mathcal{S}_y| + |\mathcal{T}'|$, it is seen that the most stringent conditions in (85) are exactly those defined in the admissible condition of (58). To exclude the case $m = k$,

observe that under this condition (89) is trivially true. The proof is thus complete. ∎

Now we show the important fact that the space of random variables admissible for symmetric splitting is generally not trivial or empty, by investigating the conditions (57), (58) and (59) one by one. First observe that the left-hand side of (57) satisfies

$$\frac{l}{k-1}H(\bar{X}_{k,\mathcal{I}_{k-1}}|X) - H(\bar{X}_{k,\mathcal{I}_l}|\bar{Y}_{k,\mathcal{I}_{l+m}\setminus\mathcal{I}_l}, X)$$
$$\leq \frac{l}{k-1}H(\bar{X}_{k,\mathcal{I}_{k-1}}|X) - H(\bar{X}_{k,\mathcal{I}_l}|\bar{Y}_{k,\mathcal{I}_{l+m}}, X)$$
$$\overset{(a)}{\leq} \frac{l}{k-1}H(\bar{X}_{k,\mathcal{I}_{k-1}}|X)$$
$$\quad - \frac{l}{k-1}H(\bar{X}_{k,\mathcal{I}_{k-1}}|\bar{Y}_{k,\mathcal{I}_{k-1}}, X)$$
$$= \frac{l}{k-1}I(\bar{X}_{k,\mathcal{I}_{k-1}}; \bar{Y}_{k,\mathcal{I}_{k-1}}|X)$$
$$\leq \frac{l}{k-1}I(\bar{X}_{k,\mathcal{I}_{k-1}}; \bar{Y}_{k,\mathcal{I}_{k-1}}, X)$$
$$\overset{(b)}{=} \frac{l}{k-1}I(\bar{X}_{k,\mathcal{I}_{k-1}}; \bar{Y}_{k,\mathcal{I}_{k-1}}) \tag{90}$$

where both (a) and (b) are by the Markov string $X_{k,i} \leftrightarrow Y_{k,i} \leftrightarrow (X, \bar{Y}_{k,\mathcal{I}_K\setminus\{i\}}, \bar{X}_{k,\mathcal{I}_{k-1}\setminus\{i\}})$. For the right-hand side of (57), we have

$$H(\bar{Y}_{k,\mathcal{I}_m}|X) - \frac{m}{K}H(\bar{Y}_{k,\mathcal{I}_K}|X)$$
$$= m\left[\frac{1}{m}H(\bar{Y}_{k,\mathcal{I}_m}|X) - \frac{1}{K}H(\bar{Y}_{k,\mathcal{I}_K}|X)\right]$$
$$\overset{(c)}{\geq} m\left[\frac{1}{K-1}H(\bar{Y}_{k,\mathcal{I}_{K-1}}|X) - \frac{1}{K}H(\bar{Y}_{k,\mathcal{I}_K}|X)\right]$$
$$\overset{(d)}{\geq} 0 \tag{91}$$

where both (c) and (d) are by Han's inequality; note that for the case that $m = K$, (57) is true trivially, and thus it can be ignored. The bracket in (91) can be zero; however, following the proof of Han's inequality [25, p. 491], it is easily verified that this only occurs when $Y_{k,i}$'s are mutually independent conditioned on X. Thus, (91) holds with strict inequality unless $m = 0$ or $Y_{k,i}$'s are mutually independent conditioned on X. It is clear that if (91) holds with strict inequality, by making $I(\bar{X}_{k,\mathcal{I}_{k-1}}; \bar{Y}_{k,\mathcal{I}_{k-1}})$ sufficiently small but positive, (57) can be made true. Moreover, when $m = 0$, the left-hand side of (57) is clearly nonpositive by Han's inequality; when $Y_{k,i}$'s are independent conditioned on X, the left-hand side of (57) is now in fact zero. Thus, for these two special cases, (57) is satisfied without requiring $I(\bar{X}_{k,\mathcal{I}_{k-1}}; \bar{Y}_{k,\mathcal{I}_{k-1}})$ to be small.

Next, notice that the right-hand side of (58) satisfies

$$\frac{m}{k}H(\bar{Y}_{k,\mathcal{I}_k}) - H(\bar{Y}_{k,\mathcal{I}_m}|\bar{Y}_{k,\mathcal{I}_k\setminus\mathcal{I}_m}, \bar{X}_{k,\mathcal{I}_l})$$
$$\geq \frac{m}{k}H(\bar{Y}_{k,\mathcal{I}_k}) - H(\bar{Y}_{k,\mathcal{I}_m}|\bar{Y}_{k,\mathcal{I}_k\setminus\mathcal{I}_m})$$
$$= (k-m)\left[\frac{1}{k-m}H(\bar{Y}_{k,\mathcal{I}_{k-m}}) - \frac{1}{k}H(\bar{Y}_{k,\mathcal{I}_k})\right]$$

$$\overset{(e)}{\geq} (k-m)\left[\frac{1}{k-1}H(\bar{Y}_{k,\mathcal{I}_{k-1}}) - \frac{1}{k}H(\bar{Y}_{k,\mathcal{I}_k})\right]$$
$$\overset{(f)}{\geq} 0 \tag{92}$$

where both (e) and (f) are again by Han's inequality, and the last inequality (f) holds with equality only when $Y_{k,i}$'s are mutually independent. If (92) holds with strict inequality, by making $I(\bar{X}_{k,\mathcal{I}_{k-1}}; \bar{Y}_{k,\mathcal{I}_{k-1}})$ sufficiently small but positive, (58) can be made true. Moreover, when $Y_{k,i}$'s are indeed mutually independent, using the Markov string $X_{k,i} \leftrightarrow Y_{k,i} \leftrightarrow (X, \bar{Y}_{k,\mathcal{I}_K\setminus\{i\}}, \bar{X}_{k,\mathcal{I}_{k-1}\setminus\{i\}})$, it is clear that

$$\frac{m}{k}H(\bar{Y}_{k,\mathcal{I}_k}) - H(\bar{Y}_{k,\mathcal{I}_m}|\bar{Y}_{k,\mathcal{I}_k\setminus\mathcal{I}_m}, \bar{X}_{k,\mathcal{I}_l})$$
$$= \frac{m}{k}H(\bar{Y}_{k,\mathcal{I}_k}) - H(\bar{Y}_{k,\mathcal{I}_m}|\bar{X}_{k,\mathcal{I}_l})$$
$$= mH(Y_{k,1}) - (m-l)H(Y_{k,1}) - H(\bar{Y}_{k,\mathcal{I}_l}|\bar{X}_{k,\mathcal{I}_l})$$
$$= H(\bar{Y}_{k,\mathcal{I}_l}) - H(\bar{Y}_{k,\mathcal{I}_l}|\bar{X}_{k,\mathcal{I}_l})$$
$$= lI(X_{k,1}; Y_{k,1}) = \frac{l}{k-1}I(\bar{X}_{k,\mathcal{I}_{k-1}}; \bar{Y}_{k,\mathcal{I}_{k-1}}). \tag{93}$$

Thus, for this case (58) holds without requiring $I(\bar{X}_{k,\mathcal{I}_{k-1}}; \bar{Y}_{k,\mathcal{I}_{k-1}})$ to be small.

It is also straightforward to show that the right-hand side of (59) satisfies

$$\frac{1}{k}H(\bar{Y}_{k,\mathcal{I}_k}) - \frac{1}{K}H(\bar{Y}_{k,\mathcal{I}_K}|X)$$
$$= \frac{1}{k}H(\bar{Y}_{k,\mathcal{I}_k}) - \frac{1}{K}H(\bar{Y}_{k,\mathcal{I}_K}) + \frac{1}{K}H(\bar{Y}_{k,\mathcal{I}_K})$$
$$\quad - \frac{1}{K}H(\bar{Y}_{k,\mathcal{I}_K}|X)$$
$$= \left[\frac{1}{k}H(\bar{Y}_{k,\mathcal{I}_k}) - \frac{1}{K}H(\bar{Y}_{k,\mathcal{I}_K})\right]$$
$$\quad + \frac{1}{K}I(\bar{Y}_{k,\mathcal{I}_K}; X) > 0 \tag{94}$$

because the first bracket is not-negative by Han's inquality, and the second term is strictly positive since the set of random variables $\bar{Y}_{k,\mathcal{I}_K}$ is dependent on X; moreover, the left-hand side of (59) satisfies

$$\frac{1}{k-1}I(\bar{X}_{k,\mathcal{I}_{k-1}}; X) \leq \frac{1}{k-1}I(\bar{X}_{k,\mathcal{I}_{k-1}}; \bar{Y}_{k,\mathcal{I}_{k-1}}). \tag{95}$$

Thus, (59) can also be made true by making $I(\bar{X}_{k,\mathcal{I}_{k-1}}; \bar{Y}_{k,\mathcal{I}_{k-1}})$ small.

Summarize the above discussion, we have that the following cases may occur.

- $(Y_{k,1}, \ldots, Y_{k,K})$ are conditional independent given X, but $Y_{k,i}$'s are not independent. The right-hand side of (57) is zero, but (57) always holds. Clearly the right-hand side of (58) and (59) are strictly positive by (92) and (94). By making $I(\bar{X}_{k,\mathcal{I}_{k-1}}; \bar{Y}_{k,\mathcal{I}_{k-1}})$ sufficiently small, nontrivial symmetric splitting random variables can be found.
- $Y_{k,i}$'s are independent. In fact, for this case the sizes of bins in the coding scheme are zero, implying no binning is used. This implies that each individual description is decodable in the original scheme, and the quantization splitting technique is not necessary. Using the admissible con-

dition, we see that the right-hand side of (92) is zero, but (58) always holds; the right-hand side of (57) and (59) are strictly positive. This means there exist splitting random variables, and it can be verified that even $X_{k,i} = Y_{k,i}$ is a valid choice. This is exactly the reason why the R-D tuple given in (13) is achievable in the original single layer SCEC for the Gaussian source.

- The general dependency case. The right-hand side of (57), (58), and (59) are all strictly positive. By making $I(\bar{X}_{k,\mathcal{I}_{k-1}}; \bar{Y}_{k,\mathcal{I}_{k-1}})$ sufficiently small, nontrivial symmetric splitting random variables can be found.

Potentially we can now find the optimal Gaussian random variables admissible for symmetric splitting, for the Gaussian source and any symmetric Gaussian random variables $\bar{Y}_{k,\mathcal{I}_K}$, particularly for the no k description excess case where $(Y_{k,i} - X)$'s are independent. However, for the general K-description problem, the admissible conditions reduce to a set of polynomial inequalities on the variance (and the correlation coefficients for the general dependence case), which do not seem to yield a closed form optimal solution. Nevertheless this set of conditions is still useful, since numerical methods can now be applied to verify if certain random variables are admissible for symmetric splitting, and the number of inequalities to be checked is quadratic instead of being exponential (in K and k).

The choice of splitting random variables is certainly not unique, and the set of random variables being admissible for symmetric splitting is only a subset of the possible choices. We present the symmetric splitting mainly for two reasons. The first reason to favor symmetric splitting is that for the multi-layer scheme given in the next section, the symmetric splitting appears to be a more natural choice than others, because of the accumulation of lower layer information. The second reason is its smaller number of constraints. The conditions given in (66)–(68) are in fact the fundamental constraints that should be satisfied by any splitting random variables, however, the number of constraints directly derived from (66)–(68) can be exponentially large (in K and k). Other splitting may yield even better performance in the single layer setting, by allowing some combinations of less than $k - 1$ descriptions to be useful.

Before ending this section, we emphasize that in the coding scheme given above, there does not exist a completely specified coding order due to the constraints on the splitting random variables, in contrast to the schemes in [12] and [13]. Nevertheless, it can still be understood as being partially ordered: the x sublayer is first decoded if decodable, but if it is not, there is no particular sequential decoding order and all the codewords are jointly decoded.

V. MULTILAYER CODING SCHEMES

In this section we provide two general multilayer schemes for the K-description problem, by properly stacking the two single layer coding blocks given in the previous section, respectively. The schemes can be further generalized, and some of these generalizations are discussed in the final section.

A. Multilayer Scheme Based On Channel Codes and SCEC

We introduce more random variables in addition to the ones that already appear in Theorem 1: for the kth layer, $k = 1, 2, \ldots, K - 1$, k random variables $X_{k,j}$, $j = 1, 2, 3, \ldots, k$, are introduced. We further generalize the definition of the mixed symmetric distribution as follows.

Definition 6: A joint distribution $p(\bar{y}_{\mathcal{I}_{K-1},\mathcal{I}_K}, y_K, \{\bar{x}_{k,\mathcal{I}_k}, k \in \mathcal{I}_{K-1}\}|x)$ is called *multilayer mixed generalized symmetric* if for any permutations $\pi(\cdot) : \mathcal{I}_K \to \mathcal{I}_K$ and $\pi_k(\cdot) : \mathcal{I}_k \to \mathcal{I}_k$ for $k \in \mathcal{I}_{K-1}$, the joint distribution $p(\bar{y}_{\mathcal{I}_{K-1},\pi(\mathcal{I}_K)}, y_K, \{\bar{x}_{k,\pi_k(\mathcal{I}_k)}, k \in \mathcal{I}_{K-1}\}|x)$ is the same as $p(\bar{y}_{\mathcal{I}_{K-1},\mathcal{I}_K}, y_K, \{\bar{x}_{k,\mathcal{I}_k}, k \in \mathcal{I}_{K-1}\}|x)$.

The above definition essentially requires that the additional random variables $X_{k,i}$'s are symmetric according to the more stringent definition in [7], yet the random variables $Y_{k,i}$'s only need to satisfy the generalized definition of being symmetric. This new coding scheme reduces to the one in [7] without the additional $X_{k,j}$ random variables, however, with a relaxed requirement on the symmetry of the distribution.

Theorem 7: A symmetric R-D vector $(R, D_1, D_2, \ldots, D_K)$ is achievable if there exists a probability distribution

$$
\begin{aligned}
&p(x, \bar{y}_{\mathcal{I}_{K-1},\mathcal{I}_K}, y_K, \{\bar{x}_{k,\mathcal{I}_k}, k \in \mathcal{I}_{K-1}\}) \\
&= p(x)p(\bar{y}_{\mathcal{I}_{K-1},\mathcal{I}_K}, y_K, \{\bar{x}_{k,\mathcal{I}_k}, k \in \mathcal{I}_{K-1}\}|x) \quad (96)
\end{aligned}
$$

that is multilayer mixed generalized symmetric, defined over finite alphabets $\mathcal{X} \times \mathcal{Y}_1^K \times \mathcal{Y}_2^K \cdots \times \mathcal{Y}_{K-1}^K \times \mathcal{Y}_K \times \mathcal{X}_1 \times \mathcal{X}_2^2 \cdots \times \mathcal{X}_{K-1}^{K-1}$, and a set of decoding functions

$$
\begin{aligned}
g_k' &: \prod_{j=1}^{k} \mathcal{Y}_j^k \times \prod_{j=1}^{k} \mathcal{X}_j^j \times \mathcal{X}_{k+1}^k \to \mathcal{X}, \quad k \in \mathcal{I}_{K-2}, \\
g_k'' &: \prod_{j=1}^{k} \mathcal{Y}_j^k \times \prod_{j=1}^{k} \mathcal{X}_j^j \to \mathcal{X}, \quad k \in \mathcal{I}_{K-2}, \\
g_{K-1} &: \prod_{j=1}^{K-1} \mathcal{Y}_j^{K-1} \times \prod_{j=1}^{K-1} \mathcal{X}_j^j \to \mathcal{X}, \\
g_K &: \prod_{j=1}^{K-1} \mathcal{Y}_j^K \times \mathcal{Y}_K \times \prod_{j=1}^{K-1} \mathcal{X}_j^j \to \mathcal{X}
\end{aligned}
$$

such that

$$
\begin{aligned}
&\lambda_k \mathbb{E}[d(X, g_k'(\bar{Y}_{\mathcal{I}_k,\mathcal{I}_k}, \{X_{j,\mathcal{I}_j}, j \in \mathcal{I}_k\}, X_{k+1,\mathcal{I}_k}))] \\
&\quad + (1 - \lambda_k)\mathbb{E}[d(X, g_k''(\bar{Y}_{\mathcal{I}_k,\mathcal{I}_k}, \{X_{j,\mathcal{I}_j}, j \in \mathcal{I}_k\}))] \le D_k, \\
&\quad k \le K - 2, \quad (97)
\end{aligned}
$$

$$
\mathbb{E}[d(X, g_{K-1}(\bar{Y}_{\mathcal{I}_{K-1},\mathcal{I}_{K-1}}, \{X_{j,\mathcal{I}_j}, j \in \mathcal{I}_{K-1}\}))] \le D_{K-1}, \quad (98)
$$

$$
\mathbb{E}[d(X, g_K(\bar{Y}_{\mathcal{I}_{K-1},\mathcal{I}_K}, Y_K, \{X_{j,\mathcal{I}_j}, j \in \mathcal{I}_{K-1}\}))] \le D_K \quad (99)
$$

where $\lambda_k = (k+1)/\binom{K}{k}$, for $k \in \mathcal{I}_{K-2}$, and

$$
\begin{aligned}
R \geq \ & I(X; X_{1,1}) \\
& + \sum_{k=1}^{K-2} \frac{1}{k} H(\bar{X}_{k+1,\mathcal{I}_k} | \bar{Y}_{\mathcal{I}_k, \mathcal{I}_k}, \{\bar{X}_{j,\mathcal{I}_j}, j \in \mathcal{I}_k\}) \\
& - \sum_{k=1}^{K-2} \frac{1}{k+1} H(\bar{X}_{k+1,\mathcal{I}_{k+1}} | X, \bar{Y}_{\mathcal{I}_k, \mathcal{I}_K}, \{\bar{X}_{j,\mathcal{I}_j}, j \in \mathcal{I}_k\}) \\
& + \sum_{k=1}^{K-1} \frac{1}{k} H(\bar{Y}_{k,\mathcal{I}_k} | \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_k}, \{\bar{X}_{j,\mathcal{I}_j}, j \in \mathcal{I}_k\}) \\
& - \frac{1}{K} \sum_{k=1}^{K-1} H(\bar{Y}_{k,\mathcal{I}_K} | X, \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_K}, \{\bar{X}_{j,\mathcal{I}_j}, j \in \mathcal{I}_k\}) \\
& + \frac{1}{K} I(X; Y_K | \bar{Y}_{\mathcal{I}_{K-1}, \mathcal{I}_K}, \{\bar{X}_{j,\mathcal{I}_j}, j \in \mathcal{I}_{K-1}\}). \quad (100)
\end{aligned}
$$

Proof: We give an outline of the coding scheme and focus only on its difference from the single layer case. Assume the distribution $p(\bar{y}_{\mathcal{I}_{K-1}, \mathcal{I}_K}, y_K, \{\bar{x}_{k,\mathcal{I}_k}, k \in \mathcal{I}_{K-1}\} | x)$ is fixed.

Random codebook generation: The codebooks are generated for each layer, and each layer consists of two sublayers; the first $K-1$ layer codebooks are generated using the method given in the single layer scheme. For the Kth layer, a conventional codebook using Y_K conditioned on $(\bar{Y}_{\mathcal{I}_{K-1}, \mathcal{I}_K}, \{\bar{X}_{j,\mathcal{I}_j}, j \in \mathcal{I}_{K-1}\})$ is used.

The encoding is performed from the lower to higher layers. The codebooks should be sufficiently large, such that we can find in each codebook at least one codeword, such that all of them are jointly typical together with any typical source sequence x^n, as well with the already found lower layer codewords with high probability. Using the properties of typical sequences, it can be shown that we need to ensure that for $k = 1, 2, \ldots, K-1$

$$
\begin{aligned}
\hat{R}_k^x \geq \ & H(X_{k,1}) \\
& - \frac{1}{k} H(\bar{X}_{k,\mathcal{I}_k} | X, \{\bar{X}_{j,\mathcal{I}_j}, j \in \mathcal{I}_{k-1}\}, \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_K}),
\end{aligned}
$$
$$(101)$$
$$
\begin{aligned}
\hat{R}_k^y \geq \ & H(Y_{k,1}) \\
& - \frac{1}{K} H(\bar{Y}_{k,\mathcal{I}_K} | X, \{\bar{X}_{j,\mathcal{I}_j}, j \in \mathcal{I}_k\}, \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_K}). (102)
\end{aligned}
$$

For the last layer, a conditional codebook of the following rate suffices

$$
KR_K \geq I(X; Y_K | \bar{Y}_{\mathcal{I}_{K-1}, \mathcal{I}_K}, \{\bar{X}_{j,\mathcal{I}_j}, j \in \mathcal{I}_{K-1}\}). \quad (103)
$$

Random binning and encoding: At the kth layer, $k = 2, \ldots, K-1$, both the x and y sublayer codebooks go through an additional binning step. During encoding, the bin indices are found, and the same systematic MDS coding step is also performed as in the single layer scheme. For the first layer no binning is performed, but the x sublayer codeword indices go through the $(K, 1)$ MDS coding step, i.e., repetition in each description. For the last layer, the codeword index is split into the K descriptions.

Decoding: Decoder $T_{\mathcal{A}}$, where $|\mathcal{A}| \leq K-2$, is required to recover with high probability the y sublayer codewords in the first $|\mathcal{A}|$ layers, as well as the x sublayer codewords in the first $|\mathcal{A}|$ layers; additionally, if $\mathcal{A} \subset \mathcal{I}_{|\mathcal{A}|+1}$, then it is to recover with high probability the x sublayer codewords associated with $\bar{X}_{|\mathcal{A}|+1,\mathcal{A}}$. This additional requirement is void for the last two layers, which reduces to the one in the original PPR multilayer scheme. We thus focus on the case that $|\mathcal{A}| \leq K-2$.

For $\mathcal{A}$ such that $|\mathcal{A}| = k \leq K-2$, the decoder decodes from lower layer to higher layer sequentially, assuming the lower layer codewords are all decoded correctly; for ease of discussion, we understand implicitly that within a given layer the x sublayer is "lower" than the y sublayer. At the kth layer, the decoder tries to find a unique k-tuple of y sublayer codewords from the given bins, such that they together are jointly typical with all the lower layer codewords. The sizes of bins for the y sublayer cannot be too large. As in [6] and [7], using the symmetry to simplify the constraints, it is sufficient to have

$$
\begin{aligned}
k(\hat{R}_k^y - R_k^y) \leq \ & kH(Y_{k,1}) \\
& - H(\bar{Y}_{k,\mathcal{I}_k} | \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_k}, \{\bar{X}_{j,\mathcal{I}_j}, j \in \mathcal{I}_k\}), \quad k \in \mathcal{I}_{K-1}. \quad (104)
\end{aligned}
$$

Now consider the additional condition that if $\mathcal{A} \subset \mathcal{I}_{|\mathcal{A}|+1}$, then the decoder can recover with high probability the codewords associated with $\bar{X}_{|\mathcal{A}|+1,\mathcal{A}}$. Similar to the single layer scheme, these systematic portions of the MDS code can be used to recover the codeword bin indices, and these codewords can then be recovered using joint typicality decoding, if the binning rate is chosen properly. It suffices to control the x sublayer rate in the $(k+1)$th layer as

$$
\begin{aligned}
k(\hat{R}_{k+1}^x - R_{k+1}^x) \leq \ & kH(X_{k+1,1}) \\
& - H(\bar{X}_{k+1,\mathcal{I}_k} | \bar{Y}_{\mathcal{I}_k, \mathcal{I}_k}, \{\bar{X}_{j,\mathcal{I}_j}, j \in \mathcal{I}_k\}), \quad k \in \mathcal{I}_{K-2} \quad (105)
\end{aligned}
$$

where again the symmetry is used. Clearly due to the symmetry, if for each layer this additional condition is satisfied, the requirement is satisfied with high probability that the x sublayer codewords in the first $|\mathcal{A}|$ layers can be decoded.

Summarizing (102), (103), (104), and (105) gives the rate in the theorem. It remains to show the distortions are indeed achievable. For any $|\mathcal{A}| \leq K-2$, there are two possible decoding functions: the first is when $\mathcal{A} \subset \mathcal{I}_{|\mathcal{A}|+1}$, and the second is when $\mathcal{A} \not\subset \mathcal{I}_{|\mathcal{A}|+1}$. For the latter, the decoding function operates on the codewords associated with random variables $\{Y_{\mathcal{I}_{|\mathcal{A}|}, \mathcal{A}}, \{X_{k,\mathcal{I}_k}, k \in \mathcal{I}_{|\mathcal{A}|}\}\}$; for the former, the decoding function can additionally operate on the codewords associated with $X_{|\mathcal{A}|+1,\mathcal{A}}$. In order to make the distortions symmetric, a time-sharing step is used, and from a simple counting argument we see that the following distortion is achievable

$$
D_k = \lambda_k D_k' + (1 - \lambda_k) D_k'' \quad (106)
$$

where D_k' and D_k'' are the distortion induced by the two decoding functions, respectively. The decoding functions for the $(K-1)$th layer and the Kth layer are straightforward. The proof is thus complete. ∎

There is one critical difference between the multilayer and single layer schemes. First note that in the PPR multilayer scheme, a k-description decoder, when attempting to decode the kth layer, uses the lower $k-1$ layers in all the k available descriptions. In the above scheme, however, since the systematic MDS channel codes are introduced to boost the $(k-1)$ description reconstruction, the x sublayer cannot take full advantage of the lower $k-1$ layers in all the k descriptions, but only those in some $k-1$ descriptions. This does not cause any penalty for the single layer scheme, because the lower layers do not exist in that case; however, in the multilayer case, this effect does introduce a possible penalty. Nevertheless, it is likely that in certain regimes the penalty such introduced is compensated by the gain of the boosted performance, which is indeed true if the lower $k-1$ layers are void, as has already been shown in the previous two sections. It will be seen that the same remark is also true for the multilayer scheme based on quantization splitting. In a sense, this additional tension suggests that for certain distortion regimes, some coding components may in fact degenerate and should be removed.

B. Multilayer Scheme Based on Quantization Splitting

In this section we provide a multilayer scheme using the single layer splitting scheme given in the previous section as basic building blocks. We first introduce a slightly generalized definition of random variables being admissible for symmetric splitting, then give the main theorem and its proof.

Definition 7: Let $p(\bar{y}_{\mathcal{I}_{K-1},\mathcal{I}_K}, y_K|x)$ be a generalized symmetric distribution. The random variables $\bar{X}_{k,\mathcal{I}_{k-1}}$, $k = 2,3,\ldots,K-1$ are *multilayer-admissible for symmetric splitting* with respect to $p(\bar{y}_{\mathcal{I}_{K-1},\mathcal{I}_K}, y_K|x)$, if they are jointly distributed with X and $(\bar{Y}_{\mathcal{I}_{K-1},\mathcal{I}_K}, Y_K)$ such that for each $k = 2,3,\ldots,K-1$, $X_{k,i} \leftrightarrow Y_{k,i} \leftrightarrow (X, \bar{Y}_{\mathcal{I}_{K-1},\mathcal{I}_K}\backslash Y_{k,i}, Y_K, \{\bar{X}_{k,\mathcal{I}_{k-1}}, k = 2,3,\ldots,K-1\}\backslash X_{k,i}, X_K)$ is a Markov string and $p(x_{k,i}|y_{k,i})$ is the same for all $i = 1,2,\ldots,k-1$; furthermore the following conditions are satisfied for $k = 2,3,\ldots,K-1$

$$\frac{l}{k-1} H(\bar{X}_{k,\mathcal{I}_{k-1}}|X, \bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_K})$$
$$- H(\bar{X}_{k,\mathcal{I}_l}|X, \bar{Y}_{k,\mathcal{I}_{l+m}\backslash\mathcal{I}_l}, \bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_K})$$
$$\leq H(\bar{Y}_{k,\mathcal{I}_m}|X, \bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_K}) - \frac{m}{K} H(\bar{Y}_{k,\mathcal{I}_K}|X, \bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_K}),$$
$$0 \le l \le k-1, \ 0 \le l+m \le K \quad (107)$$

and

$$\frac{l_2}{k-1} I(\bar{Y}_{k,\mathcal{I}_{k-1}}; \bar{X}_{k,\mathcal{I}_{k-1}}|\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_k})$$
$$- \frac{l_1}{k-1} I(\bar{Y}_{k,\mathcal{I}_{k-1}}; \bar{X}_{k,\mathcal{I}_{k-1}}|\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}})$$
$$\leq \frac{m}{k} H(\bar{Y}_{k,\mathcal{I}_k}|\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_k})$$
$$- H(\bar{Y}_{k,\mathcal{I}_m}|\bar{Y}_{k,\mathcal{I}_k\backslash\mathcal{I}_m}, \bar{X}_{k,\mathcal{I}_{l_2-l_1}}, \bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_k}),$$
$$0 \le l_1 \le l_2 \le m \le \min(k, l_2 + K - k + 1), \ l_2 < k \quad (108)$$

where l, l_1, l_2, m are nonnegative integers, and

$$\frac{1}{k-1} H(\bar{X}_{k,\mathcal{I}_{k-1}}|\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_k})$$
$$- \frac{1}{k-1} H(\bar{X}_{k,\mathcal{I}_{k-1}}|X, \bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_K})$$
$$\leq \frac{1}{k} H(\bar{Y}_{k,\mathcal{I}_k}|\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_k}) - \frac{1}{K} H(\bar{Y}_{k,\mathcal{I}_K}|X, \bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_K}).$$
$$(109)$$

We have the following theorem.

Theorem 8: Let $p(\bar{y}_{\mathcal{I}_{K-1},\mathcal{I}_K}, y_K|x)$ be generalized symmetric, and let $\bar{X}_{k,\mathcal{I}_{k-1}}$, $k = 2,3,\ldots,K-1$ be multilayer-admissible for symmetric splitting with respect to $p(\bar{y}_{\mathcal{I}_{K-1},\mathcal{I}_K}, y_K|x)$. If there exist decoding functions

$$g'_k : \prod_{j=1}^k \mathcal{Y}_j^k \times \mathcal{X}_{k+1}^k \to \mathcal{X}, \quad k = 1,2,\ldots,K-2,$$

$$g''_k : \prod_{j=1}^k \mathcal{Y}_j^k \to \mathcal{X}, \quad k = 1,2,\ldots,K-2,$$

$$g_{K-1} : \prod_{j=1}^{K-1} \mathcal{Y}_j^{K-1} \to \mathcal{X},$$

$$g_K : \prod_{j=1}^{K-1} \mathcal{Y}_j^K \times \mathcal{Y}_K \to \mathcal{X}$$

such that

$$\lambda_k^* \mathbb{E}d(X, g'_k(\bar{Y}_{\mathcal{I}_k,\mathcal{I}_k}, \bar{X}_{k+1,\mathcal{I}_k}))$$
$$+ (1-\lambda_k^*) \mathbb{E}d(X, g''_k(\bar{Y}_{\mathcal{I}_k,\mathcal{I}_k})) \leq D_k,$$
$$k = 1,2,\ldots,K-2, \quad (110)$$

$$\mathbb{E}d(X, g_{K-1}(\bar{Y}_{\mathcal{I}_{K-1},\mathcal{I}_{K-1}})) \leq D_{K-1}, \quad (111)$$

$$\mathbb{E}d(X, g_K(\bar{Y}_{\mathcal{I}_{K-1},\mathcal{I}_K}, Y_K)) \leq D_K \quad (112)$$

where $\lambda_k^* = \binom{K}{k}^{-1}$, then the symmetric R-D tuple $(R, D_1, D_2, \ldots, D_K)$ is achievable where

$$R \ge \sum_{k=1}^{K-1} \frac{1}{k} H(\bar{Y}_{k,\mathcal{I}_k}|\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_k}) + \frac{1}{K} H(Y_K|\bar{Y}_{\mathcal{I}_{K-1},\mathcal{I}_K})$$
$$- \frac{1}{K} H(\bar{Y}_{\mathcal{I}_{K-1},\mathcal{I}_K}, Y_K|X)$$
$$+ \frac{1}{K} \sum_{k=2}^{K-1} I(\bar{X}_{k,\mathcal{I}_{k-1}}; \bar{Y}_{\mathcal{I}_{k-1},k}|\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}). \quad (113)$$

Remark: By letting the random variables $\{\bar{X}_{k,\mathcal{I}_{k-1}}, k = 2,3,\ldots,K-1\}$ be constants, we recover Theorem 1. Different from the single layer case where is no rate penalty from the original single layer SCEC in general, there exists a possible rate penalty, i.e., the last term in (113), comparing to the rate expression in Theorem 1 in the multilayer case. Nevertheless, in certain regimes, such penalty is likely to be compensated by the boosted performance, similar to the multilayer scheme based on channel codes and SCEC. Also note that due to the Markov strings given in the admissible condition, $X_{k,i}$ is useless when

$Y_{k,i}$ is given, and thus the decoding functions do not need to include those x sublayer random variables, unless $Y_{k,i}$ is not available. We next give an outline for the coding scheme. The majority of the proof is similar to the single layer case, and we thus only focus on the difference.

Proof: First fix the joint distribution. For the first layer $k = 1$, a codebook of size $2^{n\hat{R}^y_{1,i}}$ is generated using the marginal distribution of $Y_{1,i}$ for $i \in \mathcal{I}_K$. Then for each $k = 2, 3, \ldots K - 1$, the x and y sublayer codebooks are generated using the same mechanism as in the single layer scheme, including the binning step, but the rates are to be specified later. The Kth layer is based on a conditional codebook.

During encoding in each layer, the encoder finds in each codebook a codeword such that they are jointly typical with a typical source sequence as well as with each other and the lower layer codewords already found. Then the encoding is performed similarly as the single layer scheme, except the first and the last layer. The first layer codewords' indices are used directly as the description information without binning; the last layer is straightforward.

For decoder $T_{\mathcal{A}}$, such that $|\mathcal{A}| = k \leq K - 2$, two kinds of decoders exist. If $\mathcal{A} = \mathcal{I}_k$, then the decoder uses the first k layer codewords associated with $\bar{Y}_{\mathcal{I}_k, \mathcal{I}_k}$, and in addition the x sublayer codewords associated with the random variables $\bar{X}_{k+1, \mathcal{I}_k}$; otherwise, only the codewords associated with $\bar{Y}_{\mathcal{I}_k, \mathcal{A}}$ are used. The last two layer decoders are of course straightforward.

Similar constraints exist on the codebook rates and the binning rates as in the single layer scheme. We have for each $k = 2, 3, \ldots, K - 1$, the following conditions should hold for $(\hat{R}^x_{k,1}, \ldots, \hat{R}^x_{k,k-1}, \hat{R}^y_{k,1}, \ldots, \hat{R}^y_{k,K}) \in \mathbb{R}^{K+k-1}_+$ and $(R^x_{k,1}, \ldots, R^x_{k,k-1}, R^y_{k,1}, \ldots, R^y_{k,K}) \in \mathbb{R}^{K+k-1}_+$:

1) Codebook constraints:

$$
\begin{aligned}
&\sum_{i \in \mathcal{A}} \hat{R}^y_{k,i} + \sum_{i \in \mathcal{B}} \hat{R}^x_{k,i} \\
&\geq \sum_{i \in \mathcal{A}} H(Y_{k,i}) + \sum_{i \in \mathcal{B}} H(X_{k,i}) \\
&\quad - H(\bar{Y}_{k,\mathcal{A}}, \bar{X}_{k,\mathcal{B}} | X, \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_K}), \\
&\quad\quad \mathcal{A} \subseteq \mathcal{I}_K, \mathcal{B} \subseteq \mathcal{I}_{k-1}.
\end{aligned} \tag{114}
$$

2) Binning constraints for k description decoders at the k-layer: for any $\mathcal{A} \in 2^{\mathcal{I}_K}_k$, let $\mathcal{B} = \mathcal{A} \cap \mathcal{I}_{k-1}$, then

$$
\begin{aligned}
0 &\leq \sum_{i \in \mathcal{A}'} (\hat{R}^y_{k,i} - R^y_{k,i}) + \sum_{i \in \mathcal{B}'} (\hat{R}^x_{k,i} - R^x_{k,i}) \\
&\leq \sum_{i \in \mathcal{A}'} H(Y_{k,i}) + \sum_{i \in \mathcal{B}'} H(X_{k,i}) \\
&\quad - H(\bar{Y}_{k,\mathcal{A}'}, \bar{X}_{k,\mathcal{B}'} | \bar{Y}_{k,\mathcal{A} \backslash \mathcal{A}'}, \bar{X}_{k,\mathcal{B} \backslash \mathcal{B}'}, \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_k}), \\
&\quad\quad \text{any } \mathcal{A}' \subseteq \mathcal{A}, \mathcal{B}' \subseteq \mathcal{B}.
\end{aligned} \tag{115}
$$

3) Binning constraints for the $\mathcal{I}_{k-1}$ description decoder at the kth layer:

$$
\begin{aligned}
0 &\leq \sum_{i \in \mathcal{B}} (\hat{R}^x_{k,i} - R^x_{k,i}) \\
&\leq \sum_{i \in \mathcal{B}} H(X_{k,i}) \\
&\quad - H(\bar{X}_{k,\mathcal{B}} | \bar{X}_{k,\mathcal{I}_{k-1} \backslash \mathcal{B}}, \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_{k-1}}), \\
&\quad\quad \text{any } \mathcal{B} \subseteq \mathcal{I}_{k-1}.
\end{aligned} \tag{116}
$$

Similarly as in the single layer case, we choose certain specific rates. For each $k = 2, 3, \ldots, K - 1$, the codebook rates are set as

$$
\begin{aligned}
\hat{R}^x_{k,i} = \hat{R}^x_k &\triangleq H(X_{k,1}) - \frac{1}{k-1} H(\bar{X}_{k,\mathcal{I}_{k-1}} | X, \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_K}), \\
&\quad i \in \mathcal{I}_{k-1}, \tag{117}
\end{aligned}
$$

$$
\begin{aligned}
\hat{R}^y_{k,i} = \hat{R}^{y*}_k &\triangleq H(Y_{k,1}) - \frac{1}{K} H(\bar{Y}_{k,\mathcal{I}_K} | X, \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_K}) \\
&\quad + \frac{1}{k-1} I(\bar{X}_{k,\mathcal{I}_{k-1}}; \bar{Y}_{k,\mathcal{I}_{k-1}} | X, \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_K}), \\
&\quad i \in \mathcal{I}_{k-1}, \tag{118}
\end{aligned}
$$

$$
\begin{aligned}
\hat{R}^y_{k,i} = \hat{R}^y_k &\triangleq H(Y_{k,1}) - \frac{1}{K} H(\bar{Y}_{k,\mathcal{I}_K} | X, \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_K}), \\
&\quad i \in \mathcal{I}_K \backslash \mathcal{I}_{k-1} \tag{119}
\end{aligned}
$$

and the binning rates are set as

$$
\begin{aligned}
R^x_{k,i} = R^x_k &\triangleq \frac{1}{k-1} H(\bar{X}_{k,\mathcal{I}_{k-1}} | \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_{k-1}}) \\
&\quad - \frac{1}{k-1} H(\bar{X}_{k,\mathcal{I}_{k-1}} | X, \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_K}), \quad i \in \mathcal{I}_{k-1}, \tag{120}
\end{aligned}
$$

$$
\begin{aligned}
R^y_{k,i} = R^{y*}_k &\triangleq \frac{1}{k} H(\bar{Y}_{k,\mathcal{I}_k} | \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_k}) \\
&\quad - \frac{1}{K} H(\bar{Y}_{k,\mathcal{I}_K} | X, \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_K}) \\
&\quad - \frac{1}{k-1} H(\bar{X}_{k,\mathcal{I}_{k-1}} | \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_k}) \\
&\quad + \frac{1}{k-1} H(\bar{X}_{k,\mathcal{I}_{k-1}} | X, \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_K}), \\
&\quad i \in \mathcal{I}_{k-1}, \tag{121}
\end{aligned}
$$

$$
\begin{aligned}
R^y_{k,i} = R^y_k &\triangleq \frac{1}{k} H(\bar{Y}_{k,\mathcal{I}_k} | \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_k}) \\
&\quad - \frac{1}{K} H(\bar{Y}_{k,\mathcal{I}_K} | X, \bar{Y}_{\mathcal{I}_{k-1}, \mathcal{I}_K}), \\
&\quad i \in \mathcal{I}_K \backslash \mathcal{I}_{k-1}. \tag{122}
\end{aligned}
$$

We only need to check the above choice indeed satisfies (114)–(116). The cookbook constraint (114) can be reduced similarly as in the single layer case, which eventually can be shown to be exactly the multilayer admissible splitting condition (107); the binning constraint (116) is also straightforward to verify. Thus, we only need to check if (115) is satisfied.

Again let $\mathcal{S} \subseteq \mathcal{I}_{k-1}$ and $\mathcal{T} \subseteq \mathcal{I}_K \backslash \mathcal{I}_{k-1}$, such that $|\mathcal{S} \cup \mathcal{T}| = k$. With the given rates, the condition (115) is equivalent to that for any such $\mathcal{S}$ and $\mathcal{T}$, and their subsets $\mathcal{S}_x, \mathcal{S}_y, \mathcal{T}'$ such that $\mathcal{S}_x \subseteq \mathcal{S}$, $\mathcal{S}_y \subseteq \mathcal{S}$ and $\mathcal{T}' \subseteq \mathcal{T}$, as given in (123), shown at

$$\frac{|\mathcal{S}_x|}{k-1}H(\bar{X}_{k,\mathcal{I}_{k-1}}|\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}) + \frac{|\mathcal{S}_y| + |\mathcal{T}'|}{k}H(\bar{Y}_{k,\mathcal{I}_k}|\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_k}) - \frac{|\mathcal{S}_y|}{k-1}I(\bar{X}_{k,\mathcal{I}_{k-1}};\bar{Y}_{k,\mathcal{I}_{k-1}}|\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_k})$$
$$\geq H(\bar{X}_{k,\mathcal{S}_x},\bar{Y}_{k,\mathcal{S}_y},\bar{Y}_{k,\mathcal{T}'}|\bar{X}_{k,\mathcal{S}\setminus\mathcal{S}_x},\bar{Y}_{k,\mathcal{S}\setminus\mathcal{S}_y},\bar{Y}_{k,\mathcal{T}\setminus\mathcal{T}'},\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_k})$$
$$= |\mathcal{S}_x|H(X_{k,1}|Y_{k,1}) + H(\bar{Y}_{k,\mathcal{S}_y},\bar{Y}_{k,\mathcal{T}'}|\bar{X}_{k,\mathcal{S}\setminus\mathcal{S}_x},\bar{Y}_{k,\mathcal{S}\setminus\mathcal{S}_y},\bar{Y}_{k,\mathcal{T}\setminus\mathcal{T}'},\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_k}). \tag{123}$$

the bottom of the page. Similarly to the single layer case, (123) is always satisfied when $|\mathcal{S}_x| > |\mathcal{S}_y|$, and thus only the case $|\mathcal{S}_x| \leq |\mathcal{S}_y|$ needs to be considered. However, for this case, the dominant inequality is also given when $\mathcal{S}_x \subseteq \mathcal{S}_y$. By defining $l_1 \triangleq |\mathcal{S}_x|$, $l_2 \triangleq |\mathcal{S}_y|$, and $m \triangleq |\mathcal{S}_y| + |\mathcal{T}'|$, it is seen that the most stringent conditions in (123) are exactly those defined in the multilayer admissible condition of (108).

For the kth layer, $k = 2, 3, \ldots, K-1$, the rate $R_{k,i}$ is actually not the same for $i \leq k-1$ and $i > k-1$

$$R_{k,i} = \frac{1}{k}H(\bar{Y}_{k,\mathcal{I}_k}|\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_k}) - \frac{1}{K}H(\bar{Y}_{k,\mathcal{I}_K}|\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_K})$$
$$+ \frac{1}{k-1}I(\bar{X}_{k,\mathcal{I}_{k-1}};\bar{Y}_{\mathcal{I}_{k-1},k}|X,\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}),$$
$$i \in \mathcal{I}_{k-1}, \tag{124}$$
$$R_{k,i} = \frac{1}{k}H(\bar{Y}_{k,\mathcal{I}_k}|\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_k}) - \frac{1}{K}H(\bar{Y}_{k,\mathcal{I}_K}|\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_K}),$$
$$i \in \mathcal{I}_K\setminus\mathcal{I}_{k-1}. \tag{125}$$

We can thus time-share the scheme in each layer to arrive at symmetric rates and symmetric distortions, for which we have that for $k = 2, 3, \ldots, K-1$

$$R_k = \frac{1}{k}H(\bar{Y}_{k,\mathcal{I}_k}|\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_k}) - \frac{1}{K}H(\bar{Y}_{k,\mathcal{I}_K}|\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_K})$$
$$+ \frac{1}{K}I(\bar{X}_{k,\mathcal{I}_{k-1}};\bar{Y}_{\mathcal{I}_{k-1},k}|X,\bar{Y}_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}). \tag{126}$$

The rates in the first and the last layer are conventional. Adding these rates together completes the proof. ∎

VI. CONCLUDING REMARKS

We considered the symmetric multiple description problem and provided novel coding schemes. These schemes yield rate-distortion tuples outside the PPR multilayer achievable region. The two new ingredients are systematic MDS codes on some smaller scale multiple description codes, and the quantization splitting technique. General multilayer schemes are also provided. Though only symmetric problem is considered, the proposed schemes actually benefit from time-sharing asymmetric distortion tuples.

The given multilayer schemes can be further generalized and improved, and several directions for doing so are as follows.

• Time-sharing between the systematic MDS code and SCEC based scheme, and quantization splitting based scheme. As can be seen in the Gaussian source example, even for the three description case, such an approach will lead to improvements in certain distortion regime.

• Combination of the systematic MDS code and SECE based scheme, and quantization splitting based scheme. Instead

of time-sharing, we can use the single layer scheme based on systematic MDS channel code and SCEC in certain layers, and the quantization splitting scheme in the others. Quantization splitting can also be applied to the systematic part of the MDS codes.

• Other possible quantization splitting choices. In the coding scheme we only considered symmetric splittings, formalized by the admissible conditions for symmetric splitting. This is by no means necessary and asymmetric splitting may yield even better performance.

• Other schemes using asymmetric rates. Recall the multilayer scheme based on rate splitting in fact has both asymmetric rates as well as asymmetric distortions before time-sharing. It is conceivable that more general schemes using the possible components without the symmetric structure can provide even better performance.

We note that though the above directions are promising for new rate-distortion tuples, the complexity of such general schemes can quickly grow out of control. As such, it is practically more important to find schemes with simple coding structure and competitive performance, even if they are not optimal.

REFERENCES

[1] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 74–93, Sep. 2001.

[2] A. E. Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 851–857, Nov. 1982.

[3] L. Ozarow, "On a source-coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, pp. 1909–1921, Dec. 1980.

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

[5] R. Venkataramani, G. Kramer, and V. K. Goyal, "Multiple description coding with many channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2106–2114, Sep. 2003.

[6] S. S. Pradhan, R. Puri, and K. Ramchandran, "n-channel symmetric multiple descriptions—Part I (n, k) source-channel erasure codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 47–61, Jan. 2004.

[7] R. Puri, S. S. Pradhan, and K. Ramchandran, "n-channel symmetric multiple descriptions—Part II: An achievable rate-distortion region," *IEEE Trans. Inf. Theory*, vol. 51, pp. 1377–1392, Apr. 2005.

[8] H. Wang and P. Viswanath, "Vector Gaussian multiple description with individual and central receivers," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2133–2153, Jun. 2007.

[9] H. Wang and P. Viswanath, "Vector Gaussian multiple description with two levels of receivers," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 401–410, Jan. 2009.

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

[11] R. Ahlswede, C. Ning, R. Li S.-Y, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[12] J. Chen and T. Berger, "Successive Wyner–Ziv coding scheme and its application to the quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 54, no. 4, pp. 1586–1603, Apr. 2008.

[13] J. Chen, C. Tian, S. Hemami, and T. Berger, "Multiple description quantization via gram-schmidt orthogonalization," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5197–5217, Dec. 2006.

[14] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder II: General sources," *Inf. Control*, vol. 38, pp. 60–80, 1978.

[15] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, pp. 364–375, Mar. 1996.

[16] A. J. Grant, B. Rimoldi, R. L. Urbanke, and P. A. Whiting, "Rate-splitting multiple access for discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 47, pp. 873–890, Mar. 2001.

[17] B. Rimoldi, "Generalized time sharing: A low-complexity capacity-achieving multiple-access technique," *IEEE Trans. Inf. Theory*, vol. 47, pp. 2432–2442, Sep. 2001.

[18] B. Rimoldi and R. Urbanke, "Asynchronous slepian-wolf coding via source-splitting," in *Proc. IEEE Int. Symp. Inf. Theory*, Ulm, Germany, Jun. 4, 1997, p. 271.

[19] Y. Oohama, "Rate-distortion theory for Gaussian multiterminal source codings systems with several side informations at the decoder," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2577–2593, Jul. 2005.

[20] V. Prabhakaran, D. Tse, and K. Ramchandran, "Rate region of the quadratic Gaussian CEO problem," in *Proc. Int. Symp. Inf. Theory*, Chicago, IL, Jul. 2004, p. 119.

[21] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, ser. CISM Courses and Lectures, G. Longo, Ed. New York: Springer-Verlag, 1978, pp. 171–231.

[22] S. Y. Tung, "Multiterminal source coding," Ph.D dissertation, Sch. Electr. Eng., Cornell Univ, Ithaca, NY, 1978.

[23] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.

[24] J. Chen and T. Berger, "Robust distributed source coding," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3385–3398, Aug. 2008.

[25] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[26] C. Tian and J. Chen, "A novel coding scheme for symmetric multiple description coding," in *Proc. Int. Symp. Inf. Theory*, Toronto, ON, Canada, Jul. 2008, pp. 1418–1422.

Chao Tian (S'00–M'05) received the B.E. degree in electronic engineering from Tsinghua University, Beijing, China, in 2000 and the M.S. and Ph.D. degrees in electrical and computer engineering from Cornell University, Ithaca, NY, in 2003 and 2005, respectively.

He was a Postdoctoral Researcher at Ecole Polytechnique Federale de Lausanne (EPFL) from 2005 to 2007. He joined AT&T Labs-Research, Florham Park, NJ, in 2007, where he is now a Senior Member of Technical Staff. His research interests include multiuser information theory, joint source-channel coding, quantization design and analysis, as well as image/video coding and processing.

Jun Chen (S'03–M'06) received the B.E. degree with honors in communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2001 and the M.S. and Ph.D. degrees in electrical and computer engineering from Cornell University, Ithaca, NY, in 2004 and 2006, respectively.

He was a Postdoctoral Research Associate in the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign, Urbana, IL, from 2005 to 2006, and a Postdoctoral Fellow at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, from 2006 to 2007. He is currently an Assistant Professor of Electrical and Computer Engineering at McMaster University, Hamilton, ON, Canada. He holds the Barber-Gennum Chair in Information Technology. His research interests include information theory, wireless communications, and signal processing.

Prof. Chen received the Josef Raviv Memorial Postdoctoral Fellowship in 2006 and the Early Research Award from the Province of Ontario in 2010.