On the Role of the Refinement Layer in Multiple Description Coding and Scalable Coding

Jia Wang, Jun Chen, *Member, IEEE*, Lei Zhao, Paul Cuff, *Member, IEEE*, and Haim Permuter, *Member, IEEE*

***Abstract*—We clarify the relationship among several existing achievable multiple description rate-distortion regions by investigating the role of refinement layer in multiple description coding. Specifically, we show that the refinement layer in the El Gamal-Cover (EGC) scheme and the Venkataramani–Kramer–Goyal (VKG) scheme can be removed; as a consequence, the EGC region is equivalent to the EGC* region (an antecedent version of the EGC region) while the VKG region (when specialized to the 2-description case) is equivalent to the Zhang–Berger (ZB) region. Moreover, we prove that for multiple description coding with individual and hierarchical distortion constraints, the number of layers in the VKG scheme can be significantly reduced when only certain weighted sum rates are concerned. The role of refinement layer in scalable coding (a special case of multiple description coding) is also studied.**

***Index Terms*—Contra-polymatroid, multiple description coding, rate-distortion region, scalable coding, successive refinement.**

I. Introduction

A fundamental problem of multiple description coding is to characterize the rate-distortion region, which is the set of all achievable rate-distortion tuples. El Gamal and Cover (EGC) obtained an inner bound of the 2-description rate-distortion region, which was shown to be tight for the no excess rate case by Ahlswede [1]. Zhang and Berger (ZB) [24] derived a different inner bound of the 2-description rate-distortion region and showed that it contains rate-distortion tuples not included in the EGC region. The EGC region has an antecedent version, which is sometimes referred to as the EGC* region. The EGC* region was shown to be tight for the quadratic Gaussian case by Ozarow [13]. However, the EGC* region has been largely abandoned in view of the fact that it is contained in the EGC region [24]. Other work on the 2-description problem can be found in [7], [8], [12], and [25]. Recent years have seen growth of interest in the general L-description problem [14], [15], [18], [19], [21]. In particular, Venkataramani, Kramer, and Goyal (VKG) [21] derived an inner bound of the L-description rate-distortion region. It is well understood that for the 2-description case both the EGC region and the ZB region subsume the EGC* region while all these three regions are contained in the VKG region; moreover, the reason that one region contains another is simply because more layers are used. Indeed, the ZB scheme has one more common description layer than the EGC* scheme while the EGC scheme and the VKG scheme have one more refinement layer than the EGC* scheme and the ZB scheme, respectively. Although it is known [24] that the EGC* scheme can be strictly improved via the inclusion of a common description layer, it is still unclear whether the refinement layer has the same effect. We shall show that in fact the refinement layer can be safely removed; as a consequence, the EGC region is equivalent to the EGC* region and the VKG region is equivalent to the ZB region.

An important special case of the 2-description problem is called scalable coding, also known as successive refinement[1]. The rate-distortion region of scalable coding has been characterized by Koshelev [10] [11], Equitz and Cover [5] for the no rate loss case and by Rimoldi [16] for the general case. In scalable coding, the second description is not required to reconstruct the source; instead, it serves as a refinement layer to improve the first description. However, it is clearly of interest to know whether the refinement layer itself in an optimal scalable coding scheme can be useful, i.e., whether one can achieve a nontrivial distortion using the refinement layer alone. This problem is closely related, but not identical, to multiple description coding with no excess rate.

To the end of understanding the role of refinement layer in multiple description coding as well as scalable coding, we need the following elementary result.

Lemma 1: Let U, V, and W be jointly distributed random variables taking values in finite sets $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$, respectively. There exist a random variable Z, taking values in a finite set $\mathcal{Z}$ with $|\mathcal{Z}| \leq |\mathcal{V}|(|\mathcal{W}|-1)+1$, and a function $f: \mathcal{V} \times \mathcal{Z} \rightarrow \mathcal{W}$ such that:

1) Z is independent of V;
2) $W = f(V, Z)$;
3) $U - (V, W) - Z$ form a Markov chain.

Proof: See Appendix A. ■

It is worth mentioning that Lemma 1 is not completely new. Indeed, its variants can be found in [9], [23], and even in

Manuscript received September 12, 2009; revised June 16, 2010; accepted September 01, 2010. Date of current version February 18, 2011 J. Wang was supported in part by the NSFC under Grants 60802020, 60932006, 60632040, and 863 program 2009AA01A336. J. Chen was supported in part by the Early Research Award from the Province of Ontario and in part by the Natural Science and Engineering Research Council (NSERC) of Canada under a Discovery Grant.

J. Wang is with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: jiawang@sjtu.edu.cn).

J. Chen is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1 Canada (e-mail: junchen@ece.mcmaster.ca).

L. Zhao is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: leiz@stanford.edu).

P. Cuff is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: cuff@princeton.edu).

H. Permuter is with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel (e-mail: haimp@bgu.ac.il).

Communicated by E. Ordentlich, Associate Editor for Source Coding.

Digital Object Identifier 10.1109/TIT.2011.2104650

[1]The notion of successive refinement is sometimes used in the more restrictive no rate loss scenario.

Shannon's paper [17]. Roughly speaking, this lemma states that one can remove random variable W by introducing random variable Z and deterministic function f. It will be seen in the context of multiple description coding that Z can be incorporated into other random variables due to its special property, which results in a reduction of the number of random variables.

The remainder of this paper is devoted to the applications of Lemma 1 to multiple description coding and scalable coding. In Section II, we show that the refinement layer in the EGC scheme is not needed; therefore, the EGC region is equivalent to the EGC* region and the ZB region includes the EGC region. We examine the general L-description problem in Section III. It is shown that the final refinement layer in the VKG scheme can be removed. This result implies that the VKG region, when specialized to the 2-description case, is equivalent to the ZB region. Furthermore, we prove that for multiple description coding with individual and hierarchical distortion constraints, the number of layers in the VKG scheme can be significantly reduced when only certain weighted sum rates are concerned. We study scalable coding with an emphasis on the role of refinement layer in Section IV.

II. Two-Description Case

We shall first give a formal definition of the multiple description rate-distortion region. Let $\{X(t)\}_{t=1}^{\infty}$ be an i.i.d. process with marginal distribution p_X on $\mathcal{X}$, and $d:\mathcal{X}\times\hat{\mathcal{X}}\rightarrow[0,\infty)$ be a distortion measure, where $\mathcal{X}$ and $\hat{\mathcal{X}}$ are finite sets. Define $\mathcal{I}_L=\{1,\ldots,L\}$ for any positive integer L.

Definition 1: A rate-distortion tuple $(R_1,\ldots,R_L,D_{\mathcal{K}},\emptyset\subset\mathcal{K}\subseteq\mathcal{I}_L)$ is said to be achievable if for any $\epsilon>0$, there exist encoding functions $f_k^{(n)}:\mathcal{X}^n\rightarrow\mathcal{C}_k^{(n)}$, $k\in\mathcal{I}_L$, and decoding functions $g_{\mathcal{K}}^{(n)}:\prod_{k\in\mathcal{K}}\mathcal{C}_k^{(n)}\rightarrow\hat{\mathcal{X}}^n$, $\emptyset\subset\mathcal{K}\subseteq\mathcal{I}_L$, such that

$$\frac{1}{n}\log\left|\mathcal{C}_k^{(n)}\right|\leq R_k+\epsilon,\quad k\in\mathcal{I}_L$$
$$\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}[d(X(t),\hat{X}_{\mathcal{K}}(t))]\leq D_{\mathcal{K}}+\epsilon,\quad\emptyset\subset\mathcal{K}\subseteq\mathcal{I}_L$$

for all sufficiently large n, where $\hat{X}_{\mathcal{K}}^n=g_{\mathcal{K}}^{(n)}(f_k^{(n)}(X^n),k\in\mathcal{K})$. The multiple description rate-distortion region $\mathcal{RD}_{\mathrm{MD}}$ is the set of all achievable rate-distortion tuples.

We shall focus on the 2-description case (i.e., $L=2$) in this section. The following two inner bounds of $\mathcal{RD}_{\mathrm{MD}}$ are attributed to El Gamal and Cover.

The EGC* region $\mathcal{RD}_{\mathrm{EGC}*}$ is the convex closure of the set of quintuples $(R_1,R_2,D_{\{1\}},D_{\{2\}},D_{\{1,2\}})$ for which there exist auxiliary random variables $X_{\{1\}}$ and $X_{\{2\}}$, jointly distributed with X, and functions $\phi_{\mathcal{K}},\emptyset\subset\mathcal{K}\subseteq\{1,2\}$, such that

$$R_k\geq I\left(X;X_{\{k\}}\right),\quad k\in\{1,2\}$$
$$R_1+R_2\geq I\left(X;X_{\{1\}},X_{\{2\}}\right)+I\left(X_{\{1\}};X_{\{2\}}\right)$$
$$D_{\{k\}}\geq\mathbb{E}\left[d\left(X,\phi_{\{i\}}\left(X_{\{i\}}\right)\right)\right],\quad k\in\{1,2\}$$
$$D_{\{1,2\}}\geq\mathbb{E}\left[d\left(X,\phi_{\{1,2\}}\left(X_{\{1\}},X_{\{2\}}\right)\right)\right].$$

The EGC region $\mathcal{RD}_{\mathrm{EGC}}$ is the convex closure of the set of quintuples $(R_1,R_2,D_{\{1\}},D_{\{2\}},D_{\{1,2\}})$ for which there exist auxiliary random variables $X_{\mathcal{K}}$, $\emptyset\subset\mathcal{K}\subseteq\{1,2\}$, jointly distributed with X, such that

$$R_k\geq I\left(X;X_{\{k\}}\right),\quad k\in\{1,2\}\tag{1}$$
$$R_1+R_2\geq I\left(X;X_{\{1\}},X_{\{2\}},X_{\{1,2\}}\right)+I\left(X_{\{1\}};X_{\{2\}}\right)\tag{2}$$
$$D_{\mathcal{K}}\geq\mathbb{E}[d(X,X_{\mathcal{K}})],\quad\emptyset\subset\mathcal{K}\subseteq\{1,2\}.\tag{3}$$

To see the connection between these two inner bounds, we shall write the EGC region in an alternative form. It can be verified that the EGC region is equivalent to the set of quintuples $(R_1,R_2,D_{\{1\}},D_{\{2\}},D_{\{1,2\}})$ for which there exist auxiliary random variables $X_{\mathcal{K}}$, $\emptyset\subset\mathcal{K}\subseteq\{1,2\}$, jointly distributed with X, and functions $\phi_{\mathcal{K}},\emptyset\subset\mathcal{K}\subseteq\{1,2\}$, such that

$$R_k\geq I\left(X;X_{\{k\}}\right),\quad k\in\{1,2\}$$
$$R_1+R_2\geq I\left(X;X_{\{1\}},X_{\{2\}},X_{\{1,2\}}\right)+I\left(X_{\{1\}};X_{\{2\}}\right)$$
$$D_{\{k\}}\geq\mathbb{E}\left[d\left(X,\phi_{\{k\}}\left(X_{\{k\}}\right)\right)\right],\quad k\in\{1,2\}$$
$$D_{\{1,2\}}\geq\mathbb{E}\left[d\left(X,\phi_{\{1,2\}}\left(X_{\{1\}},X_{\{2\}},X_{\{1,2\}}\right)\right)\right].$$

It is easy to see from this alternative form of the EGC region that the only difference from the EGC* region is the additional random variable $X_{\{1,2\}}$, which corresponds to a refinement layer; by setting $X_{\{1,2\}}$ to be constant (i.e., removing the refinement layer), we recover the EGC* region. Therefore, the EGC* region is contained in the EGC region. It is natural to ask whether the refinement layer leads to a strict improvement. The answer turns out to be negative as shown by the following theorem, which states that the two regions are in fact equivalent.

Theorem 1: $\mathcal{RD}_{\mathrm{EGC}*}=\mathcal{RD}_{\mathrm{EGC}}$.

Proof: In view of the fact that $\mathcal{RD}_{\mathrm{EGC}*}\subseteq\mathcal{RD}_{\mathrm{EGC}}$, it suffices to prove $\mathcal{RD}_{\mathrm{EGC}}\subseteq\mathcal{RD}_{\mathrm{EGC}*}$.

For any fixed $p_{XX_{\{1\}}X_{\{2\}}X_{\{1,2\}}}$, the region specified by (1)–(3) has two vertices

$$v_1:\left(R_1(v_1),R_2(v_1),D_{\{1\}}(v_1),D_{\{2\}}(v_1),D_{\{1,2\}}(v_1)\right)$$
$$v_2:\left(R_1(v_2),R_2(v_2),D_{\{1\}}(v_2),D_{\{2\}}(v_2),D_{\{1,2\}}(v_2)\right)$$

where

$$R_1(v_1)=I\left(X;X_{\{1\}}\right)$$
$$R_2(v_1)=I\left(X;X_{\{2\}},X_{\{1,2\}}\,\middle|X_{\{1\}}\right)+I\left(X_{\{1\}};X_{\{2\}}\right)$$
$$R_1(v_2)=I\left(X;X_{\{1\}},X_{\{1,2\}}\,\middle|X_{\{2\}}\right)+I\left(X_{\{1\}};X_{\{2\}}\right)$$
$$R_2(v_2)=I\left(X;X_{\{2\}}\right)$$
$$D_{\mathcal{K}}(v_1)=D_{\mathcal{K}}(v_2)=\mathbb{E}[d(X,X_{\mathcal{K}})]\quad\emptyset\subset\mathcal{K}\subseteq\{1,2\}.$$

We just need to show that both vertices are contained in the EGC* region. By symmetry, we shall only consider vertex v_1.

It follows from Lemma 1 that there exist a random variable Z, jointly distributed with $(X,X_{\{1\}},X_{\{2\}},X_{\{1,2\}})$, and a function f such that

1) Z is independent of $(X_{\{1\}}, X_{\{2\}})$;
2) $X_{\{1,2\}} = f(X_{\{1\}}, X_{\{2\}}, Z)$;
3) $X - (X_{\{1\}}, X_{\{2\}}, X_{\{1,2\}}) - Z$ form a Markov chain.

By the fact that $X - (X_{\{1\}}, X_{\{2\}}, X_{\{1,2\}}) - Z$ form a Markov chain and that $X_{\{1,2\}}$ is a deterministic function of $(X_{\{1\}}, X_{\{2\}}, Z)$, we have

$$I\left(X; X_{\{2\}}, X_{\{1,2\}} \middle| X_{\{1\}}\right) = I\left(X; X_{\{2\}}, X_{\{1,2\}}, Z \middle| X_{\{1\}}\right) = I\left(X; X_{\{2\}}, Z \middle| X_{\{1\}}\right).$$

Moreover, since Z is independent of $(X_{\{1\}}, X_{\{2\}})$, it follows that

$$I\left(X_{\{1\}}; X_{\{2\}}\right) = I\left(X_{\{1\}}; X_{\{2\}}, Z\right).$$

By setting $X'_{\{2\}} = (X_{\{2\}}, Z)$, we can rewrite the coordinates of v_1 as

$$\begin{aligned} R_1(v_1) &= I\left(X; X_{\{1\}}\right) \\ R_2(v_1) &= I\left(X; X'_{\{2\}} \middle| X_{\{1\}}\right) + I\left(X_{\{1\}}; X'_{\{2\}}\right) \\ D_{\{1\}}(v_1) &= \mathbb{E}\left[d\left(X, \phi_{\{1\}}\left(X_{\{1\}}\right)\right)\right] \\ D_{\{2\}}(v_1) &= \mathbb{E}\left[d\left(X, \phi_{\{2\}}\left(X'_{\{2\}}\right)\right)\right] \\ D_{\{1,2\}}(v_1) &= \mathbb{E}\left[d\left(X, \phi_{\{1,2\}}\left(X_{\{1\}}, X'_{\{2\}}\right)\right)\right] \end{aligned}$$

where $\phi_{\{1\}}(X_{\{1\}}) = X_{\{1\}}$, $\phi_{\{2\}}(X'_{\{2\}}) = X_{\{2\}}$, and $\phi_{\{1,2\}}(X_{\{1\}}, X'_{\{2\}}) = f(X_{\{1\}}, X_{\{2\}}, Z) = X_{\{1,2\}}$. Therefore, it is clear that vertex v_1 is contained in the EGC* region. The proof is complete. ■

Remark: It is worth noting that the proof of Theorem 1 implicitly provides cardinality bounds for the auxiliary random variables of the EGC* region.

Now we shall proceed to discuss the ZB region, which is also an inner bound of $\mathcal{RD}_{\mathrm{MD}}$. The ZB region $\mathcal{RD}_{\mathrm{ZB}}$ is the set of quintuples $(R_1, R_2, D_{\{1\}}, D_{\{2\}}, D_{\{1,2\}})$ for which there exist auxiliary random variables $X_\emptyset$, $X_{\{1\}}$, and $X_{\{2\}}$, jointly distributed with X, and functions $\phi_{\mathcal{K}}, \emptyset \subset \mathcal{K} \subseteq \{1,2\}$, such that

$$\begin{aligned} &R_k \geq I\left(X; X_\emptyset, X_{\{k\}}\right), \quad k \in \{1,2\} \\ &R_1 + R_2 \geq 2I(X; X_\emptyset) \\ &\quad + I\left(X; X_{\{1\}}, X_{\{2\}} \middle| X_\emptyset\right) + I\left(X_{\{1\}}; X_{\{2\}} \middle| X_\emptyset\right) \\ &D_{\{k\}} \geq \mathbb{E}\left[d\left(X, \phi_{\{k\}}\left(X_\emptyset, X_{\{k\}}\right)\right)\right], \quad k \in \{1,2\} \\ &D_{\{1,2\}} \geq \mathbb{E}\left[d\left(X, \phi_{\{1,2\}}\left(X_\emptyset, X_{\{1\}}, X_{\{2\}}\right)\right)\right]. \end{aligned}$$

Note that the ZB region is a convex set. It is easy to see from the definition of the ZB region that its only difference from the EGC* region is the additional random variable $X_\emptyset$, which corresponds to a common description layer; by setting $X_\emptyset$ to be constant (i.e., removing the common description layer), we recover the EGC* region. Therefore, the EGC* region is contained in the ZB region, and the following result is an immediate consequence of Theorem 1.

Corollary 1: $\mathcal{RD}_{\mathrm{EGC}} \subseteq \mathcal{RD}_{\mathrm{ZB}}$.

Remark: Since the ZB region contains rate-distortion tuples not in the EGC region as shown in [24], the inclusion can be strict.

III. L-DESCRIPTION CASE

The general L-description problem turns out to be considerably more complex than the 2-description case. The difficulty might be attributed to the following fact. The collection of nonempty subsets of $\{1,2\}$ has a tree structure;[2] however, this is not true for subsets of $\mathcal{I}_L$ when $L > 2$. Indeed, this tree structure of distortion constraints is a fundamental feature that distinguishes the 2-description problem from the general L-description problem.

A. VKG Region

The VKG region [21], which is a natural combination and extension of the EGC region and the ZB region, is an inner bound of the L-description rate-distortion region. We shall show that the final refinement layer in the VKG scheme is dispensable, which implies that the VKG region, when specialized to the 2-description case, coincides with the ZB region. It is worth noting that the VKG scheme is not the only scheme known for the L-description problem. Indeed, there are several other schemes in the literature [14], [15], [18] which can outperform the VKG scheme in certain scenarios where the distortion constraints do not exhibit a tree structure. However, the VKG scheme remains to be the most natural one for tree-structured distortion constraints.

We shall adopt the notation in [21]. For any set $\mathcal{A}$, let $2^{\mathcal{A}}$ be the power set of $\mathcal{A}$. Given a collection of sets $\mathcal{B}$, we define $X_{(\mathcal{B})} = \{X_{\mathcal{A}} : \mathcal{A} \in \mathcal{B}\}$. Note that $X_\emptyset$ (which is a random variable) should not be confused with $X_{(\emptyset)}$ (which is interpreted as a constant). We use $R_{\mathcal{K}}$ to denote $\sum_{k\in\mathcal{K}} R_k$ for $\emptyset \subset \mathcal{K} \subseteq \mathcal{I}_L$.

The VKG region $\mathcal{RD}_{\mathrm{VKG}}$ is the set of rate-distortion tuples $(R_1, \ldots, R_L, D_{\mathcal{K}}, \emptyset \subset \mathcal{K} \subseteq \mathcal{I}_L)$ for which there exist auxiliary random variables $X_{\mathcal{K}}, \mathcal{K} \subseteq \mathcal{I}_L$, jointly distributed with X, and functions $\phi_{\mathcal{K}}, \emptyset \subset \mathcal{K} \subseteq \mathcal{I}_L$, such that

$$R_{\mathcal{K}} \geq \psi(\mathcal{K}), \quad \emptyset \subset \mathcal{K} \subseteq \mathcal{I}_L, \tag{4}$$
$$D_{\mathcal{K}} \geq \mathbb{E}\left[d_{\mathcal{K}}\left(X, \phi_{\mathcal{K}}\left(X_{(2^{\mathcal{K}})}\right)\right)\right], \quad \emptyset \subset \mathcal{K} \subseteq \mathcal{I}_L \tag{5}$$

where

$$\psi(\mathcal{K}) = (|\mathcal{K}| - 1)I(X; X_\emptyset) - H\left(X_{(2^{\mathcal{K}})} \middle| X\right) + \sum_{\mathcal{A}\subseteq\mathcal{K}} H\left(X_{\mathcal{A}} \middle| X_{(2^{\mathcal{A}}-\{\mathcal{A}\})}\right).$$

Note that the VKG region is a convex set.[3] In fact, [21] contains a weak version and a strong version of the VKG region, and the one given here is in a slightly different form from those in [21]. Specifically, one can get the weak version in [21] by replacing (5) with $D_{\mathcal{K}} \geq \mathbb{E}[d_{\mathcal{K}}(X, X_{\mathcal{K}})]$, and get the strong version in [21] by replacing (5) with $D_{\mathcal{K}} \geq \mathbb{E}[d_{\mathcal{K}}(X, \phi_{\mathcal{K}}(X_{\mathcal{K}}))]$. It is easy to verify that the strong version is equivalent to the one given here while both of them are at least as large as the weak version; moreover, all these three versions are equivalent when $L = 2$.

[2] A collection of nonempty sets is said to have a tree structure if for any two sets $\mathcal{A}$ and $\mathcal{B}$ in this collection, one of the following is true: 1) $\mathcal{A} \subseteq \mathcal{B}$, 2) $\mathcal{B} \subseteq \mathcal{A}$, 3) $\mathcal{A} \cap \mathcal{B} = \emptyset$. A collection of distortion constraints is said to have a tree structure if these distortion constraints are imposed on a collection of sets (of descriptions) with a tree structure.

[3] The convexity of the ZB region and the VKG region follows from the fact that one can incorporate a time-sharing random variable into $X_\emptyset$.

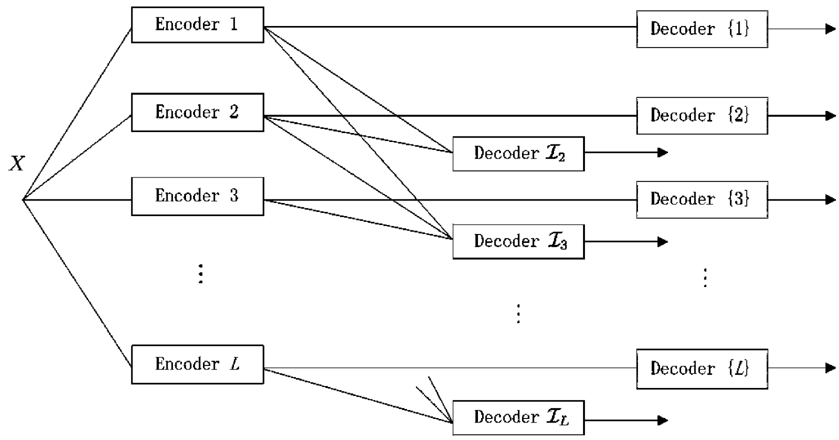

Fig. 1. Multiple description coding with individual and hierarchical distortion constraints.

We shall first give a structural characterization of the VKG region.

Lemma 2: For any fixed $p_{XX_{(2^{\mathcal{I}_L})}}$, the rate region $\{(R_1,\ldots,R_L) : R_{\mathcal{K}} \geq \psi(\mathcal{K}), \emptyset \subset \mathcal{K} \subseteq \mathcal{I}_L\}$ is a contra-polymatroid.

Proof: See Appendix B. ■

Note that the random variable $X_{\mathcal{I}_L}$ corresponds to the final refinement layer in the VKG scheme. Now we proceed to show that this refinement layer can be removed. Define the VKG* region $\mathcal{RD}_{\text{VKG}*}$ as the VKG region with $X_{\mathcal{I}_L}$ set to be a constant.

Theorem 2: $\mathcal{RD}_{\text{VKG}*} = \mathcal{RD}_{\text{VKG}}$.

Proof: The proof is given in Appendix C. ■

It is asked in [21] whether the VKG region strictly contains the ZB region. A direct consequence of Theorem 2 is that the VKG region, when specialized to the 2-description case, is equivalent to the ZB region.

Corollary 2: For the 2-description problem, $\mathcal{RD}_{\text{ZB}} = \mathcal{RD}_{\text{VKG}}$.

Remark: For the 2-description VKG region, the cardinality bound for $X_\emptyset$ can be derived by invoking Carathéodory's theorem while all the other auxiliary random variables can be assumed, with no loss of generality, to be defined on the reconstruction alphabet $\hat{\mathcal{X}}$. Therefore, one can deduce cardinality bounds for the auxiliary random variables of the ZB region by leveraging Corollary 2.

B. Multiple Description Coding With Individual and Hierarchical Distortion Constraints

We can see that for the VKG* region, the number of auxiliary random variables is exactly the same as the number of distortion constraints. Intuitively, the number of auxiliary random variables can be further reduced if we remove certain distortion constraints. We formulate the problem of multiple description coding with individual and hierarchical distortion constraints, which is a special case of tree-structured distortion constraints, and somewhat surprisingly, we show that in this setting the number of layers in the VKG scheme can be significantly reduced when only certain weighted sum rates are concerned; i.e., the number of auxiliary random variables can be significantly less than the number of distortion constraints.

For any nonnegative integer k, define $\mathcal{H}_k = \emptyset$ if $k = 0$, $\mathcal{H}_k = \{\{1\}\}$ if $k = 1$, and $\mathcal{H}_k = \{\{1\},\ldots,\{k\},\mathcal{I}_2,\ldots,\mathcal{I}_k\}$ if $k \geq 2$. Multiple description coding with individual and hierarchical distortion constraints (see Fig. 1) refers to the scenario where only the following distortion constraints: $D_{\mathcal{K}}, \mathcal{K} \in \mathcal{H}_L$, are imposed. Specializing the VKG region to this setting, we can define the VKG region for multiple description coding with individual and hierarchical distortion constraints $\mathcal{RD}_{\text{IH}-\text{VKG}}$ as the set of rate-distortion tuples $(R_1,\ldots,R_L,D_{\mathcal{K}},\mathcal{K} \in \mathcal{H}_L)$ for which there exist auxiliary random variables $X_{\mathcal{K}}, \mathcal{K} \subseteq \mathcal{I}_L$, jointly distributed with X, and functions $\phi_{\mathcal{K}}$, $\mathcal{K} \in \mathcal{H}_L$, such that

$$\begin{aligned} &R_{\mathcal{K}} \geq \psi(\mathcal{K}), \quad \emptyset \subset \mathcal{K} \subseteq \mathcal{I}_L \\ &D_{\mathcal{K}} \geq \mathbb{E}\left[d_{\mathcal{K}}\left(X, \phi_{\mathcal{K}}\left(X_{(2^{\mathcal{K}})}\right)\right)\right], \quad \mathcal{K} \in \mathcal{H}_L. \end{aligned}$$

Define $\mathcal{R}_{\text{IH}-\text{VKG}}(D_{\mathcal{K}},\mathcal{K} \in \mathcal{H}_L) = \{(R_1,\ldots,R_L) : (R_1,\ldots,R_L,D_{\mathcal{K}},\mathcal{K} \in \mathcal{H}_L) \in \mathcal{RD}_{\text{IH}-\text{VKG}}\}$. It is observed in [3] that for the quadratic Gaussian case, the number of auxiliary random variables can be significantly reduced when only certain supporting hyperplanes of $\mathcal{R}_{\text{IH}-\text{VKG}}(D_{\mathcal{K}},\mathcal{K} \in \mathcal{H}_L)$ are concerned. We shall show that this phenomenon is not restricted to the quadratic Gaussian case.

Theorem 3: For any $\alpha_1 \geq \cdots \alpha_L \geq 0$, we have

$$\begin{aligned} &\min_{(R_1,\ldots,R_L)\in\mathcal{R}_{\text{IH}-\text{VKG}}(D_{\mathcal{K}},\mathcal{K}\in\mathcal{H}_L)} \sum_{k=1}^{L} \alpha_k R_k \\ &= \min_{p_{X_\emptyset X_{\{1\}}\cdots X_{\{L\}}\mid X},\phi_{\mathcal{K}},\mathcal{K}\in\mathcal{H}_L} \sum_{k=1}^{L} \alpha_k \Big[I(X;X_\emptyset) \\ &\quad + I\left(X,\{X_{\{i\}}\}_{i=1}^{k-1};X_{\{k\}}\Big| X_\emptyset\right)\Big] \end{aligned} \tag{6}$$

where the minimization in (6) is over $p_{X_\emptyset X_{\{1\}}\cdots X_{\{L\}}\mid X}$, and $\phi_{\mathcal{K}}$, $\mathcal{K} \in \mathcal{H}_L$, subject to the constraints

$$\begin{aligned} &D_{\{k\}} \geq \mathbb{E}\left[d\left(X,\phi_{\{k\}}\left(X_\emptyset,X_{\{k\}}\right)\right)\right], \quad k \in \mathcal{I}_L \\ &D_{\mathcal{I}_k} \geq \mathbb{E}\left[d\left(X,\phi_{\mathcal{I}_k}\left(X_\emptyset,X_{\{1\}},\ldots,X_{\{k\}}\right)\right)\right] \\ &\qquad\qquad k \in \mathcal{I}_L - \{1\}. \end{aligned}$$

Proof: The proof of Theorem 3 is given in Appendix D. ■

Corollary 3: For any $\alpha_1 \geq \cdots \alpha_L \geq 0$, we have

$$\min_{(R_1,\ldots,R_L)\in\mathcal{R}_{\mathrm{IH-VKG}}(D_\mathcal{K},\mathcal{K}\in\mathcal{H}_L)} \sum_{k=1}^{L} \alpha_k R_k = \min_{p_{X_\emptyset X_{(\mathcal{H}_L)}|X}} \sum_{k=1}^{L} \alpha_k \left[I(X;X_\emptyset) + I\left(X_{(\mathcal{H}_{k-1})}; X_{\{k\}} \middle| X_\emptyset\right) + I\left(X; X_{\{k\}}, X_{\mathcal{I}_k} \middle| X_\emptyset, X_{(\mathcal{H}_{k-1})}\right)\right] \tag{7}$$

where the minimization in (7) is over $p_{X_\emptyset X_{(\mathcal{H}_L)}|X}$ subject to the constraints

$$D_\mathcal{K} \geq \mathbb{E}[d(X, X_\mathcal{K})], \quad \mathcal{K} \in \mathcal{H}_L.$$

Proof: See Appendix E. ■

Remark: It should be noted that $X_\mathcal{K}$, $\mathcal{K} \in \mathcal{H}_L$, in (7) are defined on the reconstruction alphabet $\hat{\mathcal{X}}$; moreover, for $X_\emptyset$ in (7), the cardinality bound can be easily derived by invoking Carathéodory's theorem. In view of the proof of Corollary 3, one can derive cardinality bounds for the auxiliary random variables in (6) by leveraging the cardinality bounds for the auxiliary random variables in (7). This explains why "min" instead of "inf" is used in (6).

C. Multiple Description Coding With Individual and Central Distortion Constraints

A special case of multiple description coding with individual and hierarchical distortion constraints is one where only individual distortion constraints $D_{\{k\}}$, $k \in \mathcal{I}_L$, and central distortion constraint $D_{\mathcal{I}_L}$ are imposed (see [3] and [22]). Let $\mathcal{G}_L = \{\{1\},\ldots,\{L\},\mathcal{I}_L\}$. We can define the VKG region for multiple description coding with individual and central distortion constraints $\mathcal{RD}_{\mathrm{IC-VKG}}$ as the set of rate-distortion tuples $(R_1,\ldots,R_L,D_\mathcal{K},\mathcal{K} \in \mathcal{G}_L)$ for which there exist auxiliary random variables $X_\mathcal{K}$, $\mathcal{K} \subseteq \mathcal{I}_L$, jointly distributed with X, and functions $\phi_\mathcal{K}$, $\mathcal{K} \in \mathcal{G}_L$, such that

$$R_\mathcal{K} \geq \psi(\mathcal{K}), \quad \emptyset \subset \mathcal{K} \subseteq \mathcal{I}_L$$
$$D_\mathcal{K} \geq \mathbb{E}\left[d_\mathcal{K}\left(X, \phi_\mathcal{K}\left(X_{(2^\mathcal{K})}\right)\right)\right], \quad \mathcal{K} \in \mathcal{G}_L.$$

Define $\mathcal{R}_{\mathrm{IC-VKG}}(D_\mathcal{K},\mathcal{K} \in \mathcal{G}_L) = \{(R_1,\ldots,R_L) : (R_1,\ldots,R_L,D_\mathcal{K},\mathcal{K} \in \mathcal{G}_L) \in \mathcal{RD}_{\mathrm{IC-VKG}}\}$. The following result is a simple consequence of Theorem 3 and Corollary 3.

Corollary 4:

1) $\mathcal{RD}_{\mathrm{IC-VKG}}$ is equivalent to the set of rate-distortion tuples $(R_1,\ldots,R_L,D_\mathcal{K},\mathcal{K} \in \mathcal{G}_L)$ for which there exist auxiliary random variables $X_\emptyset$, $X_{\{k\}}$, $k \in \mathcal{I}_L$, jointly distributed with X, and functions $\phi_\mathcal{K}$, $\mathcal{K} \in \mathcal{G}_L$, such that

$$R_\mathcal{K} \geq |\mathcal{K}| I(X;X_\emptyset) - H\left(\{X_{\{k\}}\}_{k\in\mathcal{K}} \middle| X, X_\emptyset\right) + \sum_{k\in\mathcal{K}} H\left(X_{\{k\}} \middle| X_\emptyset\right), \quad \emptyset \subset \mathcal{K} \subseteq \mathcal{I}_L$$
$$D_{\{k\}} \geq \mathbb{E}\left[d\left(X, \phi_{\{k\}}\left(X_\emptyset, X_{\{k\}}\right)\right)\right], \quad k \in \mathcal{I}_L$$
$$D_{\mathcal{I}_L} \geq \mathbb{E}\left[d\left(X, \phi_{\mathcal{I}_L}\left(X_\emptyset, X_{\{1\}},\ldots,X_{\{L\}}\right)\right)\right].$$

2) $\mathcal{RD}_{\mathrm{IC-VKG}}$ is also equivalent to the set of rate-distortion tuples $(R_1,\ldots,R_L,D_\mathcal{K},\mathcal{K} \in \mathcal{G}_L)$ for which there exist auxiliary random variables $X_\emptyset$, $X_\mathcal{K}$, $\mathcal{K} \in \mathcal{G}_L$, jointly distributed with X, such that

$$R_\mathcal{K} \geq |\mathcal{K}| I(X;X_\emptyset) - H\left(\{X_{\{k\}}\}_{k\in\mathcal{K}} \middle| X, X_\emptyset\right) + \sum_{k\in\mathcal{K}} H\left(X_{\{k\}} \middle| X_\emptyset\right), \quad \emptyset \subset \mathcal{K} \subset \mathcal{I}_L$$
$$R_{\mathcal{I}_L} \geq L I(X;X_\emptyset) - H\left(\{X_{\{k\}}\}_{k\in\mathcal{I}_L} \middle| X, X_\emptyset\right) + \sum_{k=1}^{L} H\left(X_{\{k\}} \middle| X_\emptyset\right) + I\left(X; X_{\mathcal{I}_L} \middle| X_\emptyset, \{X_{\{k\}}\}_{k\in\mathcal{I}_L}\right)$$
$$D_\mathcal{K} \geq \mathbb{E}[d(X, X_\mathcal{K})], \quad \mathcal{K} \in \mathcal{G}_L.$$

3) For any $(\alpha_1,\ldots,\alpha_L) \in \mathbb{R}_+^L$, let π be a permutation on $\mathcal{I}_L$ such that $\alpha_{\pi(1)} \geq \cdots \geq \alpha_{\pi(L)}$; we have

$$\min_{(R_1,\ldots,R_L)\in\mathcal{R}_{\mathrm{IC-VKG}}(D_\mathcal{K},\mathcal{K}\in\mathcal{G}_L)} \sum_{k=1}^{L} \alpha_k R_k = \min_{p_{X_\emptyset X_{\{1\}}\cdots X_{\{L\}}|X},\phi_\mathcal{K},\mathcal{K}\in\mathcal{G}_L} \sum_{k=1}^{L} \alpha_{\pi(k)} \Big[I(X;X_\emptyset) + I\left(X, \{X_{\pi(i)}\}_{i=1}^{k-1}; X_{\{\pi(k)\}} \middle| X_\emptyset\right)\Big] \tag{8}$$

$$= \min_{p_{X_\emptyset X_{(\mathcal{G}_L)}|X}} \sum_{k=1}^{L} \alpha_{\pi(k)} \Big[I(X;X_\emptyset) + I\left(X, \{X_{\pi(i)}\}_{i=1}^{k-1}; X_{\{\pi(k)\}} \middle| X_\emptyset\right)\Big] + \alpha_{\pi(L)} I\left(X; X_{\mathcal{I}_L} \middle| X_\emptyset, \{X_{\{k\}}\}_{k\in\mathcal{I}_L}\right) \tag{9}$$

where the minimization in (8) is over $p_{X_\emptyset X_{\{1\}}\cdots X_{\{L\}}|X}$, and $\phi_\mathcal{K}$, $\mathcal{K} \in \mathcal{G}_L$, subject to the constraints

$$D_{\{k\}} \geq \mathbb{E}\left[d\left(X, \phi_{\{k\}}\left(X_\emptyset, X_{\{k\}}\right)\right)\right], \quad k \in \mathcal{I}_L$$
$$D_{\mathcal{I}_L} \geq \mathbb{E}\left[d\left(X, \phi_{\mathcal{I}_L}\left(X_\emptyset, X_{\{1\}},\ldots,X_{\{L\}}\right)\right)\right]$$

while the minimization in (9) is over $p_{X_\emptyset X_{(\mathcal{G}_L)}|X}$ subject to the constraints

$$D_\mathcal{K} \geq \mathbb{E}[d(X, X_\mathcal{K})], \quad \mathcal{K} \in \mathcal{G}_L.$$

IV. Scalable Coding

Scalable coding is a special case of the 2-description problem in which the distortion constraint on the second description, i.e., $D_{\{2\}}$, is not imposed. In such a setting the first description is commonly referred to as the base layer while the second description is referred to as the refinement layer.

A. Scalable Coding Rate-Distortion Region

The scalable coding rate-distortion region $\mathcal{RD}_{\mathrm{SC}}$ is defined as

$$\mathcal{RD}_{\mathrm{SC}} = \left\{\left(R_1, R_2, D_{\{1\}}, D_{\{1,2\}}\right) : \left(R_1, R_2, D_{\{1\}}, \infty, D_{\{1,2\}}\right) \in \mathcal{RD}_{\mathrm{MD}}\right\}.$$

It is proved in [16] that the quadruple $(R_1, R_2, D_{\{1\}}, D_{\{1,2\}}) \in \mathcal{RD}_{\text{SC}}$ if and only if there exist auxiliary random variables $X_{\{1\}}$ and $X_{\{1,2\}}$ jointly distributed with X such that

$$\begin{aligned} &R_1 \geq I\left(X; X_{\{1\}}\right) \\ &R_1 + R_2 \geq I\left(X; X_{\{1\}}, X_{\{1,2\}}\right) \\ &D_{\{1\}} \geq \mathbb{E}\left[d\left(X, X_{\{1\}}\right)\right] \\ &D_{\{1,2\}} \geq \mathbb{E}\left[d\left(X, X_{\{1,2\}}\right)\right]. \end{aligned}$$

Roughly speaking, here $X_{\{1\}}$ and $X_{\{1,2\}}$ correspond to the base layer and the refinement layer, respectively. It is clear that one can obtain this standard form of $\mathcal{RD}_{\text{SC}}$ from $\mathcal{RD}_{\text{EGC}}$ by setting $X_{\{2\}}$ to be a constant.

Since the EGC region is equivalent to the EGC* region, it is not surprising that $\mathcal{RD}_{\text{SC}}$ can be written in an alternative form which resembles the EGC* region. Indeed, by leveraging Lemma 1, one can express $\mathcal{RD}_{\text{SC}}$ as the set of quadruples $(R_1, R_2, D_{\{1\}}, D_{\{1,2\}})$ for which there exist independent random variables $X_{\{1\}}$ and $X_{\{2\}}$, jointly distributed with X, and a function f, such that

$$\begin{aligned} &R_1 \geq I\left(X; X_{\{1\}}\right) \\ &R_1 + R_2 \geq I\left(X; X_{\{1\}}, X_{\{2\}}\right) \\ &D_{\{1\}} \geq \mathbb{E}\left[d\left(X, X_{\{1\}}\right)\right] \\ &D_{\{1,2\}} \geq \mathbb{E}\left[d\left(X, f\left(X_{\{1\}}, X_{\{2\}}\right)\right)\right]. \end{aligned}$$

In contrast with the standard form of $\mathcal{RD}_{\text{SC}}$, here $X_{\{2\}}$ corresponds to the refinement layer ($X_{\{1\}}$ still corresponds to the base layer). It will be seen that this alternative form of $\mathcal{RD}_{\text{SC}}$ is useful for clarifying the role of refinement layer in scalable coding.

B. On the Role of Refinement Layer in Scalable Coding

Although $D_{\{2\}}$ is not imposed in scalable coding, it is certainly of interest to know whether the refinement layer itself can be useful, i.e., whether one can use the refinement layer alone to achieve a nontrivial reconstruction distortion. However, without further constraint, this problem is essentially the same as the multiple description problem. Therefore, we shall focus on optimal greedy scalable coding schemes (which will be defined precisely).

Let $R(D)$ denote the rate-distortion function, i.e.,

$$R(D) = \min_{p_{\hat{X} \mid X}: \mathbb{E}[d(X,\hat{X})] \leq D} I(X; \hat{X}).$$

Define the minimum scalably achievable total rate $R(R_1, D_{\{1\}}, D_{\{1,2\}})$ with respect to $(R_1, D_{\{1\}}, D_{\{1,2\}})$ as

$$R\left(R_1, D_{\{1\}}, D_{\{1,2\}}\right) = \min\left\{R_1 + R_2 : \left(R_1, R_2, D_{\{1\}}, D_{\{1,2\}}\right) \in \mathcal{RD}_{\text{SC}}\right\}.$$

It is clear that [16]

$$R\left(R_1, D_{\{1\}}, D_{\{1,2\}}\right) = \min_{\substack{I(X; X_{\{1\}}) \leq R_1 \\ \mathbb{E}[d(X, X_{\{1\}})] \leq D_{\{1\}} \\ \mathbb{E}[d(X, X_{\{1,2\}})] \leq D_{\{1,2\}}}} I\left(X; X_{\{1\}}, X_{\{1,2\}}\right).$$

Here we assume the right-hand side of the equality is greater than or equal to R_1; otherwise, $R(R_1, D_{\{1\}}, D_{\{1,2\}}) = R_1$. Now we proceed to study the minimum $D_{\{2\}}$ in the scenario where $R_1 = R(D_{\{1\}})$ and $R_1 + R_2 = R(R(D_{\{1\}}), D_{\{1\}}, D_{\{1,2\}})$. Define

$$D^*_{\{2\}}\left(D_{\{1\}}, D_{\{1,2\}}\right) = \min_{\substack{R_1 = R(D_{\{1\}}) \\ R_1 + R_2 = R(R_1, D_{\{1\}}, D_{\{1,2\}}) \\ (R_1, R_2, D_{\{1\}}, D_{\{2\}}, D_{\{1,2\}}) \in \mathcal{RD}_{\text{MD}}}} D_{\{2\}}.$$

We shall refer to $D^*_{\{2\}}(D_{\{1\}}, D_{\{1,2\}})$ as the minimum distortion achievable by the refinement layer in optimal greedy scalable coding schemes.

Let $\mathcal{Q}$ denote the convex closure of the set of quintuples $(R_1, R_2, D_{\{1\}}, D_{\{2\}}, D_{\{1,2\}})$ for which there exist auxiliary random variables $X_{\mathcal{K}}$, $\emptyset \subset \mathcal{K} \subseteq \{1,2\}$, jointly distributed with X, such that

$$\begin{aligned} &I\left(X_{\{1\}}; X_{\{2\}}\right) = 0 \\ &R_k \geq I\left(X; X_{\{k\}}\right), \quad k \in \{1,2\} \\ &R_1 + R_2 \geq I\left(X; X_{\{1\}}, X_{\{2\}}, X_{\{1,2\}}\right) \\ &D_{\mathcal{K}} \geq \mathbb{E}[d(X, X_{\mathcal{K}})], \quad \emptyset \subset \mathcal{K} \subseteq \{1,2\}. \end{aligned}$$

Note that $\mathcal{Q}$ is essentially the EGC region with an additional constraint $I(X_{\{1\}}; X_{\{2\}}) = 0$ (i.e., $X_{\{1\}}$ and $X_{\{2\}}$ are independent).

Lemma 3: The EGC region as well as $\mathcal{Q}$ is tight if $R_1 + R_2 = R(R_1, D_{\{1\}}, D_{\{1,2\}})$; more precisely

$$\begin{aligned} &\left\{\left(R_1, R_2, D_{\{1\}}, D_{\{2\}}, D_{\{1,2\}}\right) \in \mathcal{RD}_{\text{MD}} : R_1 + R_2 = R\left(R_1, D_{\{1\}}, D_{\{1,2\}}\right)\right\} \\ &= \left\{\left(R_1, R_2, D_{\{1\}}, D_{\{2\}}, D_{\{1,2\}}\right) \in \mathcal{RD}_{\text{EGC}} : R_1 + R_2 = R\left(R_1, D_{\{1\}}, D_{\{1,2\}}\right)\right\} \\ &= \left\{\left(R_1, R_2, D_{\{1\}}, D_{\{2\}}, D_{\{1,2\}}\right) \in \mathcal{Q} : R_1 + R_2 = R\left(R_1, D_{\{1\}}, D_{\{1,2\}}\right)\right\}. \end{aligned}$$

Proof: Note that this problem is closely related to multiple description coding without excess rate. Indeed, for the special case $R(R_1, D_{\{1\}}, D_{\{1,2\}}) = R(D_{\{1,2\}})$, this lemma follows immediately from Ahlswede's classical result [1]. In fact, even for the general case, Ahlswede's proof technique [1] (also cf. [20]) can be directly applied with no essential change. Note that for the no excess rate case, the two descriptions are (approximately) independent (see (3.4) in [1]). It is easy to verify that this (approximate) independence also holds under the constraint $R_1 + R_2 = R(R_1, D_{\{1\}}, D_{\{1,2\}})$. The rest of the proof is essentially the same as that in [1] and is omitted. ■

Though $D^*_{\{2\}}(D_{\{1\}}, D_{\{1,2\}})$ is in principle computable using Lemma 3, the calculation is often complicated by the convex-hull operation in the definition of the EGC region and $\mathcal{Q}$. We shall show that under mild technical conditions such a convex-hull operation is not needed for the purpose of computing $D^*_{\{2\}}(D_{\{1\}}, D_{\{1,2\}})$.

We need the following definition of weak independence from [2].

Definition 2: For jointly distributed random variables U and V, U is weakly independent of V if the rows of the stochastic matrix $[p_{U\,|\,V}(u\,|\,v)]$ are linearly dependent.

The following lemma can be found in [2].

Lemma 4: For jointly distributed random variables U and V, there exists a random variable W satisfying

1) $U - V - W$ form a Markov chain;
2) U and W are independent;
3) V and W are not independent;

if and only if U is weakly independent of V.

Theorem 4: If X is not weakly independent of $X_{\{1\}}$ for any $X_{\{1\}}$ induced by $p_{X_{\{1\}}\,|\,X}$ that achieves $R(D_{\{1\}})$, then

$$D^*_{\{2\}}\left(D_{\{1\}}, D_{\{1,2\}}\right) = \min_{p_{X_{\{1\}}X_{\{2\}}\,|\,X}, g_1, g_2} \mathbb{E}\left[d\left(X, g_1\left(X_{\{2\}}\right)\right)\right] \quad (10)$$

where the minimization is over $p_{X_{\{1\}}X_{\{2\}}\,|\,X}$, g_1, and g_2 subject to the constraints

$$\begin{aligned} &I\left(X_{\{1\}}; X_{\{2\}}\right) = 0\\ &I\left(X; X_{\{1\}}\right) = R\left(D_{\{1\}}\right)\\ &I\left(X; X_{\{1\}}, X_{\{2\}}\right) = R\left(R\left(D_{\{1\}}\right), D_{\{1\}}, D_{\{1,2\}}\right)\\ &\mathbb{E}\left[d\left(X, X_{\{1\}}\right)\right] \le D_{\{1\}}\\ &\mathbb{E}\left[d\left(X, g_2\left(X_{\{1\}}, X_{\{2\}}\right)\right)\right] \le D_{\{1,2\}}. \end{aligned}$$

Here one can assume that $X_{\{2\}}$ is defined on a finite set with cardinality no greater than $|\hat{\mathcal{X}}|^4 - |\hat{\mathcal{X}}|^3 + |\hat{\mathcal{X}}|$.

Proof: First we shall show that the right-hand side of (10) is achievable. Given any $D_{\{1\}}$ and $D_{\{1,2\}}$ for which there exist auxiliary random variables $X_{\mathcal{K}}$, $\emptyset \subset \mathcal{K} \subseteq \{1, 2\}$, jointly distributed with X, and a function g_2 such that

$$\begin{aligned} &I\left(X_{\{1\}}; X_{\{2\}}\right) = 0\\ &R\left(D_{\{1\}}\right) = I\left(X; X_{\{1\}}\right)\\ &R\left(R\left(D_{\{1\}}\right), D_{\{1\}}, D_{\{1,2\}}\right) = I\left(X; X_{\{1\}}, X_{\{2\}}\right)\\ &D_{\{1\}} \ge \mathbb{E}\left[d\left(X, X_{\{1\}}\right)\right]\\ &D_{\{1,2\}} \ge \mathbb{E}\left[d\left(X, g_2\left(X_{\{1\}}, X_{\{2\}}\right)\right)\right] \end{aligned}$$

and we have

$$\begin{aligned} &R\left(R\left(D_{\{1\}}\right), D_{\{1\}}, D_{\{1,2\}}\right)\\ &\quad = I\left(X; X_{\{1\}}, X_{\{2\}}\right) + I\left(X_{\{1\}}; X_{\{2\}}\right)\\ &R\left(R\left(D_{\{1\}}\right), D_{\{1\}}, D_{\{1,2\}}\right) - R\left(D_{\{1\}}\right)\\ &\quad = I\left(X_{\{1\}}, X; X_{\{2\}}\right) \ge I\left(X; X_{\{2\}}\right). \end{aligned}$$

Therefore, the quintuple $(R_1, R_2, D_{\{1\}}, D_{\{2\}}, D_{\{1,2\}})$, where

$$\begin{aligned} R_1 &= R\left(D_{\{1\}}\right)\\ R_2 &= R\left(R\left(D_{\{1\}}\right), D_{\{1\}}, D_{\{1,2\}}\right) - R\left(D_{\{1\}}\right)\\ D_{\{2\}} &= \mathbb{E}\left[d\left(X, g_1\left(X_{\{2\}}\right)\right)\right] \end{aligned}$$

is contained in the EGC* region for any function g_1. This proves the achievability part.

Now we proceed to prove the converse part. Let $R_1 = R(D_{\{1\}})$ and $R_2 = R(R(D_{\{1\}}), D_{\{1\}}, D_{\{1,2\}}) - R(D_{\{1\}})$. Since the VKG region includes the EGC region, Lemma 3 implies that the VKG region is also tight when the total rate is equal to $R(R(D_{\{1\}}), D_{\{1\}}, D_{\{1,2\}})$. Therefore, if the quintuple $(R_1, R_2, D_{\{1\}}, D_{\{2\}}, D_{\{1,2\}})$ is achievable, then there exist auxiliary random variables $X_{\mathcal{K}}$, $\mathcal{K} \subseteq \{1, 2\}$, jointly distributed with X such that

$$\begin{aligned} &R_k \ge I\left(X; X_{\emptyset}, X_{\{k\}}\right), \quad k \in \{1, 2\}\\ &R_1 + R_2 \ge 2I(X; X_{\emptyset}) + I\left(X; X_{\{1\}}, X_{\{2\}}, X_{\{1,2\}} \,\middle|\, X_{\emptyset}\right)\\ &\quad + I\left(X_{\{1\}}; X_{\{2\}} \,\middle|\, X_{\emptyset}\right)\\ &D_{\mathcal{K}} \ge \mathbb{E}[d(X, X_{\mathcal{K}})], \quad \emptyset \subset \mathcal{K} \subseteq \{1, 2\}. \end{aligned}$$

By the definition of $R(D_{\{1\}})$ and $R(R(D_{\{1\}}), D_{\{1\}}, D_{\{1,2\}})$, we must have

$$\begin{aligned} &R\left(D_{\{1\}}\right) = I\left(X; X_{\emptyset}, X_{\{1\}}\right) = I\left(X; X_{\{1\}}\right)\\ &R\left(R\left(D_{\{1\}}\right), D_{\{1\}}, D_{\{1,2\}}\right) = 2I(X; X_{\emptyset})\\ &\quad + I\left(X; X_{\{1\}}, X_{\{2\}}, X_{\{1,2\}} \,\middle|\, X_{\emptyset}\right)\\ &\quad + I\left(X_{\{1\}}; X_{\{2\}} \,\middle|\, X_{\emptyset}\right)\\ &\quad = I\left(X; X_{\{1\}}, X_{\{1,2\}}\right) \end{aligned}$$

which implies that

1) X and $X_{\emptyset}$ are independent;
2) $X - X_{\{1\}} - X_{\emptyset}$ form a Markov chain;
3) $X_{\{1\}} - X_{\emptyset} - X_{\{2\}}$ form a Markov chain;
4) $X - (X_{\{1\}}, X_{\{1,2\}}) - (X_{\emptyset}, X_{\{2\}})$ form a Markov chain;
5) $p_{X_{\{1\}}\,|\,X}$ achieves $R(D_{\{1\}})$.

Since X is not weakly independent of $X_{\{1\}}$, it follows from Lemma 4 that $X_{\emptyset}$ and $X_{\{1\}}$ are independent, which further implies that $X_{\{1\}}$ and $X_{\{2\}}$ are independent. By Lemma 1, there exist a random variable Z on $\mathcal{Z}$ with $|\mathcal{Z}| \le |\hat{\mathcal{X}}|^3 - |\hat{\mathcal{X}}|^2 + 1$ and a function f such that

1) Z is independent of $(X_{\{1\}}, X_{\{2\}})$;
2) $X_{\{1,2\}} = f(X_{\{1\}}, X_{\{2\}}, Z)$;
3) $X - (X_{\{1\}}, X_{\{2\}}, X_{\{1,2\}}) - Z$ form a Markov chain.

By setting $X'_{\{2\}} = (X_{\{2\}}, Z)$, it is easy to verify that

$$\begin{aligned} &I\left(X_{\{1\}}; X'_{\{2\}}\right) = 0\\ &R\left(R\left(D_{\{1\}}\right), D_{\{1\}}, D_{\{1,2\}}\right) = I\left(X; X_{\{1\}}, X'_{\{2\}}\right)\\ &D_{\{2\}} \ge \mathbb{E}\left[d\left(X, g_1\left(X'_{\{2\}}\right)\right)\right]\\ &D_{\{1,2\}} \ge \mathbb{E}\left[d\left(X, g_2(X_{\{1\}}, X'_{\{2\}})\right)\right] \end{aligned}$$

where

$$g_1(X'_{\{2\}}) = g_1(X_{\{2\}}, Z) = X_{\{2\}}$$

and

$$g_2(X_{\{1\}}, X'_{\{2\}}) = f(X_{\{1\}}, X_{\{2\}}, Z) = X_{\{1,2\}}.$$

The proof is complete. ■

Remark: Theorem 4 is quite natural in view of the alternative form of $\mathcal{RD}_{\mathrm{SC}}$ defined in Section IV-A. Moreover, it is easy to show that (10) can be alternatively written as

$$D^*_{\{2\}}\left(D_{\{1\}}, D_{\{1,2\}}\right) = \min_{p_{X_{\{1\}}X_{\{2\}}X_{\{1,2\}}\,|\,X}} \mathbb{E}\left[d\left(X, X_{\{2\}}\right)\right]$$

where the minimization is over $p_{X_{\{1\}}X_{\{2\}}X_{\{1,2\}}\,|\,X}$ subject to the constraints

$$\begin{aligned}
&I\left(X_{\{1\}};X_{\{2\}}\right)=0\\
&I\left(X;X_{\{1\}}\right)=R\left(D_{\{1\}}\right)\\
&I\left(X;X_{\{1\}},X_{\{2\}},X_{\{1,2\}}\right)=R\left(R\left(D_{\{1\}}\right),D_{\{1\}},D_{\{1,2\}}\right)\\
&\mathbb{E}\left[d\left(X,X_{\{1\}}\right)\right]\le D_{\{1\}}\\
&\mathbb{E}\left[d\left(X,X_{\{1,2\}}\right)\right]\le D_{\{1,2\}}.
\end{aligned}$$

Now we give an example for which $D^*_{\{2\}}(D_{\{1\}},D_{\{1,2\}})$ can be calculated explicitly.

Theorem 5: For a binary symmetric source with Hamming distortion measure

$$D^*_{\{2\}}\left(D_{\{1\}},D_{\{1,2\}}\right)=\frac{1}{2}+D_{\{1,2\}}-D_{\{1\}}$$

for $0\le D_{\{1,2\}}\le D_{\{1\}}\le\frac{1}{2}$.

Proof: The proof is given in Appendix F. ■

APPENDIX A
PROOF OF LEMMA 1

Let Y be a random variable independent of V and uniformly distributed over [0, 1]. It is obvious that for each $v\in\mathcal{V}$ we can find a function f_v satisfying

$$\mathbb{P}(f_v(Y)=w)=p_{W\,|\,V}(w\,|\,v),\quad w\in\mathcal{W}.$$

Now define a function f such that

$$f(v,y)=f_v(y),\quad v\in\mathcal{V},y\in[0,1].$$

It is clear that

$$\mathbb{P}(f(V,Y)=w\,|\,V=v)=p_{W\,|\,V}(w\,|\,v),\quad v\in\mathcal{V},w\in\mathcal{W}.$$

It can be shown by invoking Carathéodory's theorem that there exist a finite set $\mathcal{Z}\subset[0,1]$ with $|\mathcal{Z}|\le|\mathcal{V}|(|\mathcal{W}|-1)+1$ and a random variable Z on $\mathcal{Z}$, independent of V, such that

$$\mathbb{P}(f(V,Z)=w\,|\,V=v)=p_{W\,|\,V}(w\,|\,v),\quad v\in\mathcal{V},w\in\mathcal{W}.$$

We can see that p_{VW} is preserved if W is set to be equal to $f(V,Z)$. Now we incorporate U into the probability space by setting $p_{U\,|\,VWZ}=p_{U\,|\,VW}$. It can be readily verified that p_{UVW} is preserved and $U-(V,W)-Z$ indeed form a Markov chain. The proof is complete.

APPENDIX B
PROOF OF LEMMA 2

By the definition of contra-polymatroid [4], it suffices to show that the set function $\psi:2^{\mathcal{I}_L}\to\mathbb{R}_+$ satisfies 1) $\psi(\emptyset)=0$ (normalized), 2) $\psi(\mathcal{S})\le\psi(\mathcal{T})$ if $\mathcal{S}\subset\mathcal{T}$ (nondecreasing), 3) $\psi(\mathcal{S})+\psi(\mathcal{T})\le\psi(\mathcal{S}\cup\mathcal{T})+\psi(\mathcal{S}\cap\mathcal{T})$ (supermodular).

1) Normalized: We have

$$\psi(\emptyset)=-I(X;X_\emptyset)-H(X_\emptyset\,|\,X)+H(X_\emptyset)=0.$$

2) Nondecreasing: If $\mathcal{S}\subset\mathcal{T}$, then

$$\begin{aligned}
&\psi(\mathcal{T})-\psi(\mathcal{S})\\
&\quad=(|\mathcal{T}|-|\mathcal{S}|)I(X;X_\emptyset)\\
&\qquad-H\left(X_{(2^{\mathcal{T}})}\,\middle|\,X\right)+H\left(X_{(2^{\mathcal{S}})}\,\middle|\,X\right)\\
&\qquad+\sum_{\mathcal{A}\in2^{\mathcal{T}}-2^{\mathcal{S}}}H\left(X_{\mathcal{A}}\,\middle|\,X_{(2^{\mathcal{A}}-\{\mathcal{A}\})}\right)\\
&\quad\ge-H\left(X_{(2^{\mathcal{T}}-2^{\mathcal{S}})}\,\middle|\,X,X_{(2^{\mathcal{S}})}\right)\\
&\qquad+\sum_{\mathcal{A}\in2^{\mathcal{T}}-2^{\mathcal{S}}}H\left(X_{\mathcal{A}}\,\middle|\,X_{(2^{\mathcal{A}}-\{\mathcal{A}\})}\right)\\
&\quad\ge-\sum_{k=1}^{|\mathcal{T}|}\sum_{\mathcal{A}\in2^{\mathcal{T}}-2^{\mathcal{S}},|\mathcal{A}|=k}H\left(X_{\mathcal{A}}\,\middle|\,X,\{X_{\mathcal{B}}\}_{\mathcal{B}\in2^{\mathcal{T}},|\mathcal{B}|<k}\right)\\
&\qquad+\sum_{\mathcal{A}\in2^{\mathcal{T}}-2^{\mathcal{S}}}H\left(X_{\mathcal{A}}\,\middle|\,X_{(2^{\mathcal{A}}-\{\mathcal{A}\})}\right)\\
&\quad\ge-\sum_{k=1}^{|\mathcal{T}|}\sum_{\mathcal{A}\in2^{\mathcal{T}}-2^{\mathcal{S}},|\mathcal{A}|=k}H\left(X_{\mathcal{A}}\,\middle|\,X_{(2^{\mathcal{A}}-\{\mathcal{A}\})}\right)\\
&\qquad+\sum_{\mathcal{A}\in2^{\mathcal{T}}-2^{\mathcal{S}}}H\left(X_{\mathcal{A}}\,\middle|\,X_{(2^{\mathcal{A}}-\{\mathcal{A}\})}\right)=0.
\end{aligned}$$

3) Supermodular: We have

$$\begin{aligned}
&(\psi(\mathcal{S}\cup\mathcal{T})-\psi(\mathcal{T}))-(\psi(\mathcal{S})-\psi(\mathcal{S}\cap\mathcal{T}))\\
&=(|\mathcal{S}\cup\mathcal{T}|-|\mathcal{T}|)I(X;X_\emptyset)\\
&\quad-H\left(X_{(2^{\mathcal{S}\cup\mathcal{T}}-2^{\mathcal{T}})}\,\middle|\,X,X_{(2^{\mathcal{T}})}\right)\\
&\quad+\sum_{\mathcal{A}\in2^{\mathcal{S}\cup\mathcal{T}}-2^{\mathcal{T}}}H\left(X_{\mathcal{A}}\,\middle|\,X_{(2^{\mathcal{A}}-\{\mathcal{A}\})}\right)\\
&\quad-(|\mathcal{S}|-|\mathcal{S}\cap\mathcal{T}|)I(X;X_\emptyset)\\
&\quad+H\left(X_{(2^{\mathcal{S}}-2^{\mathcal{S}\cap\mathcal{T}})}\,\middle|\,X,X_{(2^{\mathcal{S}\cap\mathcal{T}})}\right)\\
&\quad-\sum_{\mathcal{A}\in2^{\mathcal{S}}-2^{\mathcal{S}\cap\mathcal{T}}}H\left(X_{\mathcal{A}}\,\middle|\,X_{(2^{\mathcal{A}}-\{\mathcal{A}\})}\right)\\
&=-H\left(X_{(2^{\mathcal{S}\cup\mathcal{T}}-2^{\mathcal{T}})}\,\middle|\,X,X_{(2^{\mathcal{T}})}\right)\\
&\quad+H\left(X_{(2^{\mathcal{S}}-2^{\mathcal{S}\cap\mathcal{T}})}\,\middle|\,X,X_{(2^{\mathcal{S}\cap\mathcal{T}})}\right)\\
&\quad+\sum_{\mathcal{A}\in2^{\mathcal{S}\cup\mathcal{T}}-\mathcal{M}}H\left(X_{\mathcal{A}}\,\middle|\,X_{(2^{\mathcal{A}}-\{\mathcal{A}\})}\right)\\
&\ge-H\left(X_{(2^{\mathcal{S}\cup\mathcal{T}}-\mathcal{M})}\,\middle|\,X,X_{(\mathcal{M})}\right)\\
&\quad+\sum_{\mathcal{A}\in2^{\mathcal{S}\cup\mathcal{T}}-\mathcal{M}}H\left(X_{\mathcal{A}}\,\middle|\,X_{(2^{\mathcal{A}}-\{\mathcal{A}\})}\right)\\
&\ge-\sum_{k=1}^{|\mathcal{S}\cup\mathcal{T}|}\sum_{\mathcal{A}\in2^{\mathcal{S}\cup\mathcal{T}}-\mathcal{M},|\mathcal{A}|=k}H\left(X_{\mathcal{A}}\,\middle|\,X,\{X_{\mathcal{B}}\}_{\mathcal{B}\in2^{\mathcal{S}\cup\mathcal{T}},|\mathcal{B}|<k}\right)\\
&\quad+\sum_{\mathcal{A}\in2^{\mathcal{S}\cup\mathcal{T}}-\mathcal{M}}H\left(X_{\mathcal{A}}\,\middle|\,X_{(2^{\mathcal{A}}-\{\mathcal{A}\})}\right)\\
&\ge-\sum_{k=1}^{|\mathcal{S}\cup\mathcal{T}|}\sum_{\mathcal{A}\in2^{\mathcal{S}\cup\mathcal{T}}-\mathcal{M},|\mathcal{A}|=k}H\left(X_{\mathcal{A}}\,\middle|\,X_{(2^{\mathcal{A}}-\{\mathcal{A}\})}\right)\\
&\quad+\sum_{\mathcal{A}\in2^{\mathcal{S}\cup\mathcal{T}}-\mathcal{M}}H\left(X_{\mathcal{A}}\,\middle|\,X_{(2^{\mathcal{A}}-\{\mathcal{A}\})}\right)=0
\end{aligned}$$

where $\mathcal{M}=2^{\mathcal{S}}\cup2^{\mathcal{T}}$. The proof is complete.

APPENDIX C
PROOF OF THEOREM 2

It is clear that $\mathcal{RD}_{\text{VKG}*} \subseteq \mathcal{RD}_{\text{VKG}}$. Therefore, we just need to show that $\mathcal{RD}_{\text{VKG}} \subseteq \mathcal{RD}_{\text{VKG}*}$.

In view of Lemma 2 and the property of contra-polymatroid [4], for fixed $p_{XX_{(2^{\mathcal{I}_L})}}$ and $\phi_{\mathcal{K}}$, $\emptyset \subset \mathcal{K} \subseteq \mathcal{I}_L$, the region specified by (4) and (5) has $L!$ vertices: $(R_1(\pi), \dots, R_L(\pi), D_{\mathcal{K}}(\pi), \emptyset \subset \mathcal{K} \subseteq \mathcal{I}_L)$ is a vertex for each permutation π on $\mathcal{I}_L$, where

$$\begin{aligned}
R_{\pi(1)}(\pi) &= \psi(\{\pi(1)\}) \\
R_{\pi(k)}(\pi) &= \psi(\{\pi(1), \dots, \pi(k)\}) \\
&\quad - \psi(\{\pi(1), \dots, \pi(k-1)\}), \quad k \in \mathcal{I}_L - \{1\} \\
D_{\mathcal{K}}(\pi) &= \mathbb{E}\left[d\left(X, \phi_{\mathcal{K}}\left(X_{(2^{\mathcal{K}})}\right)\right)\right], \quad \emptyset \subset \mathcal{K} \subseteq \mathcal{I}_L.
\end{aligned}$$

Since the VKG* region is a convex set, it suffices to show that these $L!$ vertices are contained in the VKG* region.

Without loss of generality, we shall assume that $\pi(k) = k$, $k \in \mathcal{I}_L$. In this case, we have

$$R_L(\pi) = \psi(\mathcal{I}_L) - \psi(\mathcal{I}_{L-1}).$$

Now we proceed to write $R_L(\pi)$ as a sum of certain mutual information quantities. Define

$$\begin{aligned}
\mathcal{S}_1(k) &= \{\mathcal{A} : \mathcal{A} \in \mathcal{I}_L, |\mathcal{A}| = k, L \in \mathcal{A}\} \\
\mathcal{S}_2(k) &= \{\mathcal{A} : \mathcal{A} \in \mathcal{I}_L, |\mathcal{A}| < k, L \in \mathcal{A}\}.
\end{aligned}$$

Note that

$$\begin{aligned}
R_L(\pi) &= I(X; X_\emptyset) + H\left(X_{(2^{\mathcal{I}_{L-1}})}\middle| X\right) \\
&\quad - H\left(X_{(2^{\mathcal{I}_L})}\middle| X\right) \\
&\quad + \sum_{k=1}^{L} \sum_{\mathcal{A} \in \mathcal{S}_1(k)} H\left(X_{\mathcal{A}} \middle| X_{(2^{\mathcal{A}} - \{\mathcal{A}\})}\right) \\
&= I(X; X_\emptyset) - H\left(X_{(2^{\mathcal{I}_L})} \middle| X, X_{(2^{\mathcal{I}_{L-1}})}\right) \\
&\quad + H(X_{\{L\}} | X_\emptyset) + \sum_{k=2}^{L} \sum_{\mathcal{A} \in \mathcal{S}_1(k)} H\left(X_{\mathcal{A}} \middle| X_{(2^{\mathcal{A}} - \{\mathcal{A}\})}\right) \\
&= I(X; X_\emptyset) + I\left(X; X_{\{L\}} \middle| X_{(2^{\mathcal{I}_{L-1}})}\right) \\
&\quad + I\left(X_{(2^{\mathcal{I}_{L-1}})}; X_{\{L\}} \middle| X_\emptyset\right) \\
&\quad - H\left(X_{(2^{\mathcal{I}_L})} \middle| X, X_{(2^{\mathcal{I}_{L-1}})}, X_{\{L\}}\right) \\
&\quad + \sum_{k=2}^{L} \sum_{\mathcal{A} \in \mathcal{S}_1(k)} H\left(X_{\mathcal{A}} \middle| X_{(2^{\mathcal{A}} - \{\mathcal{A}\})}\right) \\
&= I(X; X_\emptyset) + I\left(X; X_{\{L\}} \middle| X_{(2^{\mathcal{I}_{L-1}})}\right) \\
&\quad + I\left(X_{(2^{\mathcal{I}_{L-1}})}; X_{\{L\}} \middle| X_\emptyset\right) \\
&\quad + \sum_{k=2}^{L} \left[\sum_{\mathcal{A} \in \mathcal{S}_1(k)} H\left(X_{\mathcal{A}} \middle| X_{(2^{\mathcal{A}} - \{\mathcal{A}\})}\right) \right. \\
&\quad \left. - H\left(X_{(\mathcal{S}_1(k))} \middle| X, X_{(2^{\mathcal{I}_{L-1}})}, X_{(\mathcal{S}_2(k))}\right) \right].
\end{aligned}$$

We arrange the sets in $\mathcal{S}_1(k)$ in some arbitrary order and denote them by $\mathcal{S}_{k,1}, \dots, \mathcal{S}_{k,N(k)}$, respectively, where $N(k) = \binom{L}{k} - \binom{L-1}{k}$. Then for each k

$$\begin{aligned}
&\sum_{\mathcal{A} \in \mathcal{S}_1(k)} H\left(X_{\mathcal{A}} \middle| X_{(2^{\mathcal{A}} - \{\mathcal{A}\})}\right) \\
&\quad - H\left(X_{(\mathcal{S}_1(k))} \middle| X, X_{(2^{\mathcal{I}_{L-1}})}, X_{(\mathcal{S}_2(k))}\right) \\
&= \sum_{i=1}^{N(k)} \left[H\left(X_{\mathcal{S}_{k,i}} \middle| X_{(2^{\mathcal{S}_{k,i}} - \{\mathcal{S}_{k,i}\})}\right) \right. \\
&\qquad \left. - H\left(X_{\mathcal{S}_{k,i}} \middle| X, X_{(2^{\mathcal{I}_{L-1}})}, X_{(\mathcal{S}_2(k))}, X_{(\{\mathcal{S}_{k,j}\}_{j=1}^{i-1})}\right) \right] \\
&= \sum_{i=1}^{N(k)} I\left(X, X_{(2^{\mathcal{I}_{L-1}})}, X_{(\mathcal{S}_2(k))}, X_{(\{\mathcal{S}_{k,j}\}_{j=1}^{i-1})}; \right. \\
&\qquad \left. X_{\mathcal{S}_{k,i}} \middle| X_{(2^{\mathcal{S}_{k,i}} - \{\mathcal{S}_{k,i}\})}\right) \\
&= \sum_{i=1}^{N(k)} I\left(X_{(2^{\mathcal{I}_{L-1}})}, X_{(\mathcal{S}_2(k))}, X_{(\{\mathcal{S}_{k,j}\}_{j=1}^{i-1})}; \right. \\
&\qquad \left. X_{\mathcal{S}_{k,i}} \middle| X_{(2^{\mathcal{S}_{k,i}} - \{\mathcal{S}_{k,i}\})}\right) \\
&\quad + \sum_{i=1}^{N(k)} I\left(X; X_{\mathcal{S}_{k,i}} \middle| X_{(2^{\mathcal{I}_{L-1}})}, X_{(\mathcal{S}_2(k))}, X_{(\{\mathcal{S}_{k,j}\}_{j=1}^{i-1})}\right) \\
&= \sum_{i=1}^{N(k)} I\left(X_{(2^{\mathcal{I}_{L-1}})}, X_{(\mathcal{S}_2(k))}, X_{(\{\mathcal{S}_{k,j}\}_{j=1}^{i-1})}; \right. \\
&\qquad \left. X_{\mathcal{S}_{k,i}} \middle| X_{(2^{\mathcal{S}_{k,i}} - \{\mathcal{S}_{k,i}\})}\right) \\
&\quad + I\left(X; X_{(\mathcal{S}_1(k))} \middle| X_{(2^{\mathcal{I}_{L-1}})}, X_{(\mathcal{S}_2(k))}\right).
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
R_L(\pi) &= I(X; X_\emptyset) + I\left(X; X_{\{L\}} \middle| X_{(2^{\mathcal{I}_{L-1}})}\right) \\
&\quad + I\left(X_{(2^{\mathcal{I}_{L-1}})}; X_{\{L\}} \middle| X_\emptyset\right) \\
&\quad + \sum_{k=2}^{L} \left[\sum_{i=1}^{N(k)} I\left(X_{\mathcal{S}_{k,i}}; X_{(\{\mathcal{S}_{k,j}\}_{j=1}^{i-1})}, X_{(2^{\mathcal{I}_{L-1}})}, \right.\right. \\
&\qquad \left. X_{(\mathcal{S}_2(k))} \middle| X_{(2^{\mathcal{S}_{k,i}} - \{\mathcal{S}_{k,i}\})}\right) \\
&\qquad \left. + I\left(X; X_{(\mathcal{S}_1(k))} \middle| X_{(2^{\mathcal{I}_{L-1}})}, X_{(\mathcal{S}_2(k))}\right) \right] \\
&= I(X; X_\emptyset) + I\left(X; X_{(\mathcal{S}_2(L))}, X_{\mathcal{I}_L} \middle| X_{(2^{\mathcal{I}_{L-1}})}\right) \\
&\quad + I\left(X_{(2^{\mathcal{I}_{L-1}})}; X_{\{L\}} \middle| X_\emptyset\right) \\
&\quad + \sum_{k=2}^{L-1} \left[\sum_{i=1}^{N(k)} I\left(X_{(2^{\mathcal{I}_{L-1}})}, X_{(\mathcal{S}_2(k))}, X_{(\{\mathcal{S}_{k,j}\}_{j=1}^{i-1})}; \right.\right. \\
&\qquad \left.\left. X_{\mathcal{S}_{k,i}} \middle| X_{(2^{\mathcal{S}_{k,i}} - \{\mathcal{S}_{k,i}\})}\right) \right]. \qquad (11)
\end{aligned}$$

It follows from Lemma 1 that there exist an auxiliary random variable Z and a function f such that

1) Z is independent of $(X_{(2^{\mathcal{I}_{L-1}})}, X_{(\mathcal{S}_2(L))})$;

2) $X_{\mathcal{I}_L} = f(X_{(2^{\mathcal{I}_{L-1}})}, X_{(\mathcal{S}_2(L))}, Z)$;
3) $X - (X_{(2^{\mathcal{I}_{L-1}})}, X_{(\mathcal{S}_2(L))}, X_{\mathcal{I}_L}) - Z$ form a Markov chain.
Therefore, we have

$$I\left(X; X_{(\mathcal{S}_2(L))}, X_{\mathcal{I}_L} \middle| X_{(2^{\mathcal{I}_{L-1}})}\right) = I\left(X; X_{(\mathcal{S}_2(L))}, Z \middle| X_{(2^{\mathcal{I}_{L-1}})}\right)$$

$$I\left(X_{(2^{\mathcal{I}_{L-1}})}; X_{\{L\}} \middle| X_\emptyset\right) = I\left(X_{(2^{\mathcal{I}_{L-1}})}; X_{\{L\}}, Z \middle| X_\emptyset\right)$$

and

$$I\left(X_{(2^{\mathcal{I}_{L-1}})}, X_{(\mathcal{S}_2(k))}, X_{(\{\mathcal{S}_{k,j}\}_{j=1}^{i-1})}; X_{\mathcal{S}_{k,i}} \middle| X_{(2^{\mathcal{S}_{k,i}} - \{\mathcal{S}_{k,i}\})}\right) = I\left(X_{(2^{\mathcal{I}_{L-1}})}, X_{(\mathcal{S}_2(k))}, X_{(\{\mathcal{S}_{k,j}\}_{j=1}^{i-1})}, Z; X_{\mathcal{S}_{k,i}} \middle| X_{(2^{\mathcal{S}_{k,i}} - \{\mathcal{S}_{k,i}\})}\right)$$

for $1 \le i \le N(k)$ and $2 \le k \le L-1$. Now it can be easily verified that $(R_1(\pi), \ldots, R_L(\pi), D_\mathcal{K}(\pi), \emptyset \subset \mathcal{K} \subseteq \mathcal{I}_L)$ is preserved if we substitute $X_{\{L\}}$ with $(X_{\{L\}}, Z)$, set $X_{\mathcal{I}_L}$ to be a constant, and modify $\phi_\mathcal{K}$, $\emptyset \subset \mathcal{K} \subseteq \mathcal{I}_L$, accordingly. By the definition of the VKG* region, it is clear that $(R_1(\pi), \ldots, R_L(\pi), D_\mathcal{K}(\pi), \emptyset \subset \mathcal{K} \subseteq \mathcal{I}_L) \in \mathcal{RD}_{\mathrm{VKG*}}$. The proof is complete.

Appendix D
Proof of Theorem 3

Let $R_1^* = \psi(\{1\})$, and $R_k^* = \psi(\mathcal{I}_k) - \psi(\mathcal{I}_{k-1})$, $k \in \mathcal{I}_L - \{1\}$. By Lemma 2 and the property of contra-polymatroid [4], $(R_1^*, \ldots, R_L^*)$ is a vertex of the rate region $\{(R_1, \ldots, R_L) : R_\mathcal{K} \ge \psi(\mathcal{K}), \emptyset \subset \mathcal{K} \subseteq \mathcal{I}_L\}$; moreover, we have

$$\min_{(R_1,\ldots,R_L) \in \mathcal{R}_{\mathrm{IH-VKG}}(D_\mathcal{K}, \mathcal{K} \in \mathcal{H}_L)} \sum_{k=1}^{L} \alpha_k R_k = \min_{p_{X_\emptyset X_{\{1\}} \cdots X_{\{L\}} \mid X}, \phi_\mathcal{K}, \mathcal{K} \in \mathcal{H}_L} \sum_{k=1}^{L} \alpha_k R_k^* \quad (12)$$

where the minimization in (12) is over $p_{X_\emptyset X_{\{1\}} \cdots X_{\{L\}} \mid X}$, and $\phi_\mathcal{K}$, $\mathcal{K} \in \mathcal{H}_L$, subject to the constraints

$$D_\mathcal{K} \ge \mathbb{E}\left[d\left(X, \phi_\mathcal{K}\left(X_{(2^\mathcal{K})}\right)\right)\right], \quad \mathcal{K} \in \mathcal{H}_L.$$

It follows from Theorem 2 that $X_{\mathcal{I}_L}$ can be eliminated. Inspecting (11) reveals that the same method can be used to eliminate $X_\mathcal{K}$, $\mathcal{K} \in \mathcal{S}_2(L) - \{L\}$, successively in the reverse order (i.e., the bottom-to-top and right-to-left order in Fig. 2). For k from $L-1$ to 2, we write R_k^* in a form analogous to (11) and execute this elimination procedure. In this way, all the auxiliary random variables, except $X_\emptyset, X_{\{1\}}, \ldots, X_{\{L\}}$, are eliminated. It can be verified that the resulting expression for $(R_1^*, \ldots, R_L^*)$ is

$$R_k^* = I(X; X_\emptyset) + I\left(X, \{X_{\{i\}}\}_{i=1}^{k-1}; X_{\{k\}} \middle| X_\emptyset\right), \ k \in \mathcal{I}_L.$$

The proof is complete.

Appendix E
Proof of Corollary 3

First, we shall show that (6) is greater than or equal to (7). Let $X'_{\{k\}} = \phi_{\{k\}}(X_\emptyset, X_{\{k\}})$, $k \in \mathcal{I}_L$, and $X'_{\mathcal{I}_k} = \phi_{\mathcal{I}_k}(X_\emptyset, X_{\{1\}}, \ldots, X_{\{k\}})$, $k \in \mathcal{I}_L - \{1\}$. It can be verified that

$$\begin{aligned}
&\sum_{k=1}^{L} \alpha_k \left[I(X; X_\emptyset) + I\left(X, \{X_{\{i\}}\}_{i=1}^{k-1}; X_{\{k\}} \middle| X_\emptyset\right)\right] \\
&= \sum_{k=1}^{L} \alpha_k \left[I(X; X_\emptyset) + I\left(\{X_{\{i\}}\}_{i=1}^{k-1}; X_{\{k\}} \middle| X_\emptyset\right) + I\left(X; X_{\{k\}} \middle| X_\emptyset, \{X_{\{i\}}\}_{i=1}^{k-1}\right)\right] \\
&= \sum_{k=2}^{L} \alpha_k \left[I(X; X_\emptyset) + I\left(\{X_{\{i\}}\}_{i=1}^{k-1}; X_{\{k\}} \middle| X_\emptyset\right)\right] + \sum_{k=1}^{L} (\alpha_k - \alpha_{k+1}) I\left(X; X_\emptyset, \{X_{\{i\}}\}_{i=1}^{k}\right) \\
&\ge \sum_{k=2}^{L} \alpha_k \left[I(X; X_\emptyset) + I\left(X'_{(\mathcal{H}_{k-1})}; X'_{\{k\}} \middle| X_\emptyset\right)\right] + \sum_{k=1}^{L} (\alpha_k - \alpha_{k+1}) I\left(X; X_\emptyset, X'_{(\mathcal{H}_k)}\right) \\
&= \sum_{k=1}^{L} \alpha_k \left[I(X; X_\emptyset) + I\left(X'_{(\mathcal{H}_{k-1})}; X'_{\{k\}} \middle| X_\emptyset\right) - I\left(X; X_\emptyset, X'_{(\mathcal{H}_{k-1})}\right) + I(X; X_\emptyset, X'_{(\mathcal{H}_k)})\right] \\
&= \sum_{k=1}^{L} \alpha_k \left[I(X; X_\emptyset) + I\left(X'_{(\mathcal{H}_{k-1})}; X'_{\{k\}} \middle| X_\emptyset\right) + I\left(X; X'_{\{k\}}, X'_{\mathcal{I}_k} \middle| X_\emptyset, X'_{(\mathcal{H}_{k-1})}\right)\right]
\end{aligned}$$

where $\alpha_{L+1} \triangleq 0$.

Now we proceed to show that (7) is greater than or equal to (6). It follows from Lemma 1 that there exist a random variable Z and a function f such that
1) Z is independent of $(X_\emptyset, X_{(\mathcal{H}_{L-1})}, X_{\{L\}})$;
2) $X_{\mathcal{I}_L} = f(X_\emptyset, X_{(\mathcal{H}_{L-1})}, X_{\{L\}}, Z)$;
3) $X - (X_\emptyset, X_{(\mathcal{H}_L)}) - Z$ form a Markov chain.
Note that

$$I\left(X_{(\mathcal{H}_{L-1})}; X_{\{L\}} \middle| X_\emptyset\right) = I\left(X_{(\mathcal{H}_{L-1})}; X_{\{L\}}, Z \middle| X_\emptyset\right)$$

$$I\left(X; X_{\{L\}}, X_{\mathcal{I}_L} \middle| X_\emptyset, X_{(\mathcal{H}_{L-1})}\right) = I\left(X; X_{\{L\}}, Z \middle| X_\emptyset, X_{(\mathcal{H}_{L-1})}\right).$$

Therefore, we can substitute $X_{\{L\}}$ with $(X_{\{L\}}, Z)$ and eliminate $X_{\mathcal{I}_L}$. It is clear that one can successively eliminate $X_{\mathcal{I}_{L-1}}, \ldots, X_{\mathcal{I}_2}$ in a similar manner. The proof is complete.

Appendix F
Proof of Theorem 5

It is obvious that $D^*_{\{2\}}(D_{\{1\}}, D_{\{1,2\}}) = D_{\{1,2\}}$ if $D_{\{1\}} = \frac{1}{2}$. Therefore, we shall only consider the case $D_{\{1\}} < \frac{1}{2}$.

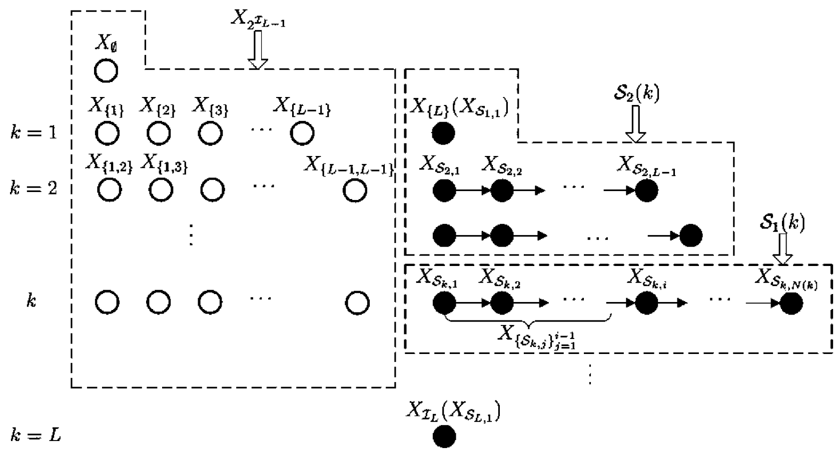

Fig. 2. Structure of auxiliary random variables for the VKG region.

Since binary symmetric sources are successively refinable, it follows that

$$R\left(D_{\{1\}}\right) = 1 - H_b\left(D_{\{1\}}\right)$$
$$R\left(R_1\left(D_{\{1\}}\right), D_{\{1\}}, D_{\{1,2\}}\right) = 1 - H_b\left(D_{\{1,2\}}\right)$$

where $H_b(\cdot)$ is the binary entropy function. If $D_{\{1\}} < \frac{1}{2}$, then $R(D_{\{1\}})$ is achieved if and only if $p_{X_{\{1\}}|X}$ is a binary symmetric channel with crossover probability $D_{\{1\}}$; it is clear that X is not weakly independent with the resulting $X_{\{1\}}$. Therefore, Theorem 4 is applicable here.

Define $X_{\{1,2\}} = g_2(X_{\{1\}}, X_{\{2\}})$. Note that we must have $\mathbb{E}[d(X, X_{\{1,2\}})] \leq D_{\{1,2\}}$ and

$$\begin{aligned} I\left(X; X_{\{1\}}, X_{\{2\}}\right) &= I\left(X; X_{\{1\}}, X_{\{2\}}, X_{\{1,2\}}\right) \\ &= I\left(X; X_{\{1,2\}}\right) \\ &= 1 - H_b\left(D_{\{1,2\}}\right) \end{aligned}$$

which implies that $X - X_{\{1,2\}} - (X_{\{1\}}, X_{\{2\}})$ form a Markov chain and $p_{X_{\{1,2\}}|X}$ is a binary symmetric channel with crossover probability $D_{\{1,2\}}$. Therefore, $p_{XX_{\{1\}}X_{\{1,2\}}}$ is completely specified by the backward test channels shown in Fig. 3. Now it is clear that one can obtain $D^*(D_{\{1\}}, D_{\{1,2\}})$ by solving the following optimization problem:

$$D^*\left(D_{\{1\}}, D_{\{1,2\}}\right) = \min_{p_{X_{\{2\}}|XX_{\{1\}}X_{\{1,2\}}}, g_1} \mathbb{E}\left[d\left(X, g_1\left(X_{\{2\}}\right)\right)\right]$$

subject to the constraints

1) $X_{\{1\}}$ and $X_{\{2\}}$ are independent;
2) $X_{\{1,2\}}$ is a deterministic function of $X_{\{1\}}$ and $X_{\{2\}}$;
3) $X - X_{\{1,2\}} - (X_{\{1\}}, X_{\{2\}})$ form a Markov chain.

Assume that $X_{\{2\}}$ takes values in $\{0, 1, \ldots, n-1\}$ for some finite n. We tabulate $p_{XX_{\{1\}}X_{\{2\}}X_{\{1,2\}}}$, $p_{X_{\{1\}}X_{\{2\}}}$, $p_{XX_{\{2\}}}$, and $p_{X_{\{1\}}X_{\{2\}}X_{\{1,2\}}}$ for ease of reading.

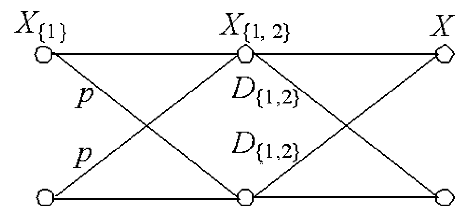

Fig. 3. Backward channels for successive refinement of a binary symmetric source: $p = \frac{D_{\{1\}} - D_{\{1,2\}}}{1 - 2D_{\{1,2\}}}$.

According to $p_{XX_{\{1\}}X_{\{1,2\}}}$ (cf. Fig. 3), it is easy to see that

$$\begin{aligned} \sum_{i=0}^{n-1} a_{0,i} &= \sum_{i=0}^{n-1} a_{7,i} = \frac{1}{2}(1-p)\left(1 - D_{\{1,2\}}\right) \\ \sum_{i=0}^{n-1} a_{1,i} &= \sum_{i=0}^{n-1} a_{6,i} = \frac{1}{2} p D_{\{1,2\}} \\ \sum_{i=0}^{n-1} a_{2,i} &= \sum_{i=0}^{n-1} a_{5,i} = \frac{1}{2} p \left(1 - D_{\{1,2\}}\right) \\ \sum_{i=0}^{n-1} a_{3,i} &= \sum_{i=0}^{n-1} a_{4,i} = \frac{1}{2}(1-p) D_{\{1,2\}}. \end{aligned} \quad (13)$$

Furthermore, one can verify the following statements.

1) The fact that $X_{\{1\}}$ and $X_{\{2\}}$ are independent and that $X_{\{1\}}$ is uniformly distributed over $\{0, 1\}$ implies

$$a_{0,i} + a_{1,i} + a_{4,i} + a_{5,i} = a_{2,i} + a_{3,i} + a_{6,i} + a_{7,i}, \quad i = 0, \ldots, n-1. \quad (14)$$

2) The fact that $X_{\{1,2\}}$ is a deterministic function of $(X_{\{1\}}, X_{\{2\}})$ implies

$$\begin{aligned} (a_{0,i} + a_{4,i})(a_{1,i} + a_{5,i}) &= (a_{2,i} + a_{6,i})(a_{3,i} + a_{7,i}) = 0 \\ & i = 0, \ldots, n-1. \end{aligned} \quad (15)$$

$x, x_{\{1\}}, x_{\{1,2\}}$ \ $x_{\{2\}}$	0	1	2	$\cdots$	$n-1$
0,0,0	$a_{0,0}$	$a_{0,1}$	$a_{0,2}$	$\cdots$	$a_{0,n-1}$
0,0,1	$a_{1,0}$	$a_{1,1}$	$a_{1,2}$	$\cdots$	$a_{1,n-1}$
0,1,0	$a_{2,0}$	$a_{2,1}$	$a_{2,2}$	$\cdots$	$a_{2,n-1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
1,1,1	$a_{7,0}$	$a_{7,1}$	$a_{7,2}$	$\cdots$	$a_{7,n-1}$

$x_{\{1\}}$ \ $x_{\{2\}}$	0	$\cdots$	$n-1$
0	$a_{0,0}+a_{1,0}+a_{4,0}+a_{5,0}$	$\cdots$	$a_{0,n-1}+a_{1,n-1}+a_{4,n-1}+a_{5,n-1}$
1	$a_{2,0}+a_{3,0}+a_{6,0}+a_{7,0}$	$\cdots$	$a_{2,n-1}+a_{3,n-1}+a_{6,n-1}+a_{7,n-1}$

x \ $x_{\{2\}}$	0	$\cdots$	$n-1$
0	$a_{0,0}+a_{1,0}+a_{2,0}+a_{3,0}$	$\cdots$	$a_{0,n-1}+a_{1,n-1}+a_{2,n-1}+a_{3,n-1}$
1	$a_{4,0}+a_{5,0}+a_{6,0}+a_{7,0}$	$\cdots$	$a_{4,n-1}+a_{5,n-1}+a_{6,n-1}+a_{7,n-1}$

$x_{\{1,2\}}$ \ $x_{\{1\}}, x_{\{2\}}$	0,0	$\cdots$	$0, n-1$	1,0	$\cdots$	$1, n-1$
0	$a_{0,0}+a_{4,0}$	$\cdots$	$a_{0,n-1}+a_{4,n-1}$	$a_{2,0}+a_{6,0}$	$\cdots$	$a_{2,n-1}+a_{6,n-1}$
1	$a_{1,0}+a_{5,0}$	$\cdots$	$a_{1,n-1}+a_{5,n-1}$	$a_{3,0}+a_{7,0}$	$\cdots$	$a_{3,n-1}+a_{7,n-1}$

3) The fact that $X - X_{\{1,2\}} - (X_{\{1\}}, X_{\{2\}})$ form a Markov chain implies

$$a_{0,i} = \frac{1-D_{\{1,2\}}}{D_{\{1,2\}}} a_{4,i}, \quad a_{5,i} = \frac{1-D_{\{1,2\}}}{D_{\{1,2\}}} a_{1,i}$$
$$a_{2,i} = \frac{1-D_{\{1,2\}}}{D_{\{1,2\}}} a_{6,i}, \quad a_{7,i} = \frac{1-D_{\{1,2\}}}{D_{\{1,2\}}} a_{3,i}$$
$$i = 0, \dots, n-1. \quad (16)$$

According to (15), there are four possibilities for each i

$$a_{0,i} = a_{2,i} = a_{4,i} = a_{6,i} = 0$$
$$\text{or} \quad a_{0,i} = a_{3,i} = a_{4,i} = a_{7,i} = 0$$
$$\text{or} \quad a_{1,i} = a_{2,i} = a_{5,i} = a_{6,i} = 0$$
$$\text{or} \quad a_{1,i} = a_{3,i} = a_{5,i} = a_{7,i} = 0, \quad i = 0, \dots, n-1.$$

Moreover, in view of (14), we can partition $\{0, 1, \dots, n-1\}$ into four disjoint sets $\mathcal{S}_j$, $j = 1, 2, 3, 4$, such that

$$a_{1,i} + a_{5,i} = a_{3,i} + a_{7,i}, \quad i \in \mathcal{S}_1$$
$$a_{1,i} + a_{5,i} = a_{2,i} + a_{6,i}, \quad i \in \mathcal{S}_2$$
$$a_{0,i} + a_{4,i} = a_{3,i} + a_{7,i}, \quad i \in \mathcal{S}_3$$
$$a_{0,i} + a_{4,i} = a_{2,i} + a_{6,i}, \quad i \in \mathcal{S}_4. \quad (17)$$

Combining (16) and (17) yields

$$a_{1,i} = a_{3,i}, \ a_{5,i} = a_{7,i}, \quad i \in \mathcal{S}_1$$
$$a_{1,i} = a_{6,i}, \ a_{2,i} = a_{5,i}, \quad i \in \mathcal{S}_2$$
$$a_{0,i} = a_{7,i}, \ a_{3,i} = a_{4,i}, \quad i \in \mathcal{S}_3$$
$$a_{0,i} = a_{2,i}, \ a_{4,i} = a_{6,i}, \quad i \in \mathcal{S}_4.$$

It is easy to see that different values in each $\mathcal{S}_j$, $j = 1, 2, 3, 4$, can be combined. That is to say, we can assume that $X_{\{2\}}$ takes values in $\{0, 1, 2, 3\}$ with no loss of generality. As a consequence, $p_{XX_{\{1\}}X_{\{2\}}X_{\{1,2\}}}$ and $p_{XX_{\{2\}}}$ can be re-tabulated shown at the top of the page.

Note that α_i and β_i satisfy

$$\beta_i = \frac{1-D_{\{1,2\}}}{D_{\{1,2\}}} \alpha_i, \quad i = 1, 2, 3, 4$$
$$\alpha_1 + \alpha_2 = \alpha_4 + \alpha_2 = \frac{1}{2} p D_{\{1,2\}}$$
$$\alpha_1 + \alpha_3 = \frac{1}{2}(1-p) D_{\{1,2\}}$$

$x, x_{\{1\}}, x_{\{1,2\}}$ \ $x_{\{2\}}$	0	1	2	3
0,0,0	0	0	β_3	β_4
0,0,1	α_1	α_2	0	0
0,1,0	0	β_2	0	β_4
0,1,1	α_1	0	α_3	0
1,0,0	0	0	α_3	α_4
1,0,1	β_1	β_2	0	0
1,1,0	0	α_2	0	α_4
1,1,1	β_1	0	β_3	0

x \ $x_{\{2\}}$	0	1	2	3
0	$2\alpha_1$	$\alpha_2 + \beta_2$	$\alpha_3 + \beta_3$	$2\beta_4$
1	$2\beta_1$	$\alpha_2 + \beta_2$	$\alpha_3 + \beta_3$	$2\alpha_4$

where the first four equalities follow from (16) while the others follow from (13). Using $X_{\{2\}}$ to reconstruct X, one can achieve

$$\begin{aligned} D_{\{2\}} &= 2\alpha_1 + \alpha_2 + \beta_2 + \alpha_3 + \beta_3 + 2\alpha_4 \\ &= \frac{1}{2} - (\beta_1 - \alpha_1 + \beta_4 - \alpha_4) \\ &= \frac{1}{2} - \frac{1 - 2D_{\{1,2\}}}{D_{\{1,2\}}}\alpha_1 - \frac{1 - 2D_{\{1,2\}}}{D_{\{1,2\}}}\alpha_4. \end{aligned}$$

It can be easily verified that $D_{\{2\}}$ is minimized when $\alpha_1 = \alpha_4 = \frac{1}{2}pD_{\{1,2\}}$. Therefore, we have

$$\begin{aligned} D^*_{\{2\}}\left(D_{\{1\}}, D_{\{1,2\}}\right) &= 2pD_{\{1,2\}} + \frac{1}{2}(1 - 2p) \\ &= \frac{1}{2} + D_{\{1,2\}} - D_{\{1\}}. \end{aligned}$$

Acknowledgment

The authors would like to thank Y.-H. Kim, the associate editor, and the anonymous reviewers for their valuable comments.

References

[1] R. Ahlswede, "The rate-distortion region for multiple descriptions without excess rate," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 6, pp. 721–726, Nov. 1985.
[2] T. Berger and R. Yeung, "Multiterminal source encoding with encoder breakdown," *IEEE Trans. Inf. Theory*, vol. IT-35, no. 2, pp. 237–244, Mar. 1989.
[3] J. Chen, "Rate region of Gaussian multiple description coding with individual and central distortion constraints," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 3991–4005, Sep. 2009.
[4] J. Edmonds, , R. Guy, H. Hanani, N. Sauer, and J. Schonheim, Eds., "Submodular functions, matroids and certain polyhedra," in *Combinatorial Structures and Their Applications*. New York: Gordon and Breach, 1970, p. 6987.
[5] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 269–275, Mar. 1991.
[6] A. El Gamal and T. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 6, pp. 851–857, Nov. 1982.
[7] H. Feng and M. Effros, "On the rate loss of multiple description source codes," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 671–683, Feb. 2005.
[8] F. Fu and R. W. Yeung, "On the rate-distortion region for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 2012–2021, Jul. 2002.
[9] Y.-H. Kim, A. Sutivong, and T. M. Cover, "State amplification," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1850–1859, May 2008.
[10] V. Koshelev, "Hierarchical coding of discrete sources," *Probl. Pered. Inf.*, vol. 16, no. 3, pp. 31–49, 1980.
[11] V. Koshelev, "An evaluation of the average distortion for discrete schemes of sequential approximation," *Probl. Pered. Inf.*, vol. 17, no. 3, pp. 20–33, 1981.
[12] L. Lastras-Montano and V. Castelli, "Near sufficiency of random coding for two descriptions," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 681–695, Feb. 2006.
[13] L. Ozarow, "On a source-coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909–1921, Dec. 1980.
[14] 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.
[15] 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, no. 4, pp. 1377–1392, Apr. 2005.
[16] B. Rimoldi, "Successive refinement of information: Characterization of the achievable rates," *IEEE Trans. Inf. Theory*, vol. 40, no. 1, pp. 253–259, Jan. 1994.
[17] C. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Devel.*, vol. 2, pp. 289–293, 1958.
[18] C. Tian and J. Chen, "New coding schemes for the symmetric K-description problem," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5344–5365, Oct. 2010.
[19] 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.
[20] E. Tuncel and K. Rose, "Additive successive refinement," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1983–1991, Aug. 2003.
[21] R. Venkataramani, G. Kramer, and V. K. Goyal, "Multiple description coding with many channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 9, pp. 2106–2114, Sep. 2003.
[22] 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.
[23] F. M. J. Willems and E. C. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 3, pp. 313–327, May 1985.
[24] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 33, no. 4, pp. 50–521, Jul. 1987.
[25] R. Zamir, "Gaussian codes and Shannon bounds for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 45, no. 11, pp. 2629–2636, Nov. 1999.

Jia Wang received the B.Sc. degree in electrical engineering, the M.S. degree in pattern recognition and intelligence control, and the Ph.D. degree in electrical engineering from Shanghai Jiao Tong University, China, in 1997, 1999, and 2002, respectively.

Since 2002, he has been with the Institute of Image Communication and Information Processing, where he is currently an Associate Professor. His research interests include multiuser information theory and its application in video coding.

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.

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

Lei Zhao received the B.Eng. degree from Tsinghua University, China, in 2003, and the M.S. degree in electrical and computer engineering from Iowa State University, Ames, in 2006. He is now a Ph.D. candidate in the Department of Electrical Engineering, Stanford University, Stanford, CA.

Paul Cuff (S'08–M'10) received the B.S. degree in electrical engineering from Brigham Young University, Provo, UT, in 2004, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 2006 and 2009, respectively.

He was awarded the ISIT 2008 Student Paper Award for his work titled "Communication Requirements for Generating Correlated Random Variables" and was a recipient of the National Defense Science and Engineering Graduate Fellowship and the Numerical Technologies Fellowship. He is an Assistant Professor of electrical engineering at Princeton University, Princeton, NJ.

Haim Permuter (M'08) received the B.Sc. (summa cum laude) and M.Sc. (summa cum laude) degrees in electrical and computer engineering from the Ben-Gurion University, Israel, in 1997 and 2003, respectively, and the Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 2008.

Between 1997 and 2004, he was an officer at a research and development unit of the Israeli Defense Forces. He is currently a lecturer at Ben-Gurion university. He is a recipient of the Fulbright Fellowship, the Stanford Graduate Fellowship (SGF), the Allon Fellowship, and the Bergman award.