Combinatorial Message Sharing and a New Achievable Region for Multiple Descriptions

Kumar B. Viswanatha, *Member, IEEE*, Emrah Akyol, *Member, IEEE*, and Kenneth Rose, *Fellow, IEEE*

***Abstract*— This paper presents a new achievable rate-distortion region for the general L channel multiple descriptions (MDs) problem. A well-known general region for this problem is due to Venkataramani, Kramer, and Goyal (VKG). Their encoding scheme is an extension of the El Gamal-Cover (EC) and Zhang–Berger (ZB) coding schemes to the L channel case and includes a combinatorial number of refinement codebooks, one for each subset of the descriptions. As in ZB, the scheme also allows for a single common codeword to be shared by all descriptions. This paper proposes a novel encoding technique involving combinatorial message sharing (CMS), where every subset of the descriptions may share a distinct common message. This introduces a combinatorial number of shared codebooks along with the refinement codebooks of. These shared codebooks provide a more flexible framework to tradeoff redundancy across the messages for resilience to descriptions loss. We derive an achievable rate-distortion region for the proposed technique, and show that it subsumes the VKG region for general sources and distortion measures. We further show that CMS provides a strict improvement of the achievable region for any source and distortion measures for which some two-description subset is such that ZB achieves points outside the EC region. We then show a more surprising result: CMS outperforms VKG for a general class of sources and distortion measures, including scenarios where the ZB and EC regions coincide for all two-description subsets. In particular, we show that CMS strictly improves on VKG, for the L-channel quadratic Gaussian MD problem, for all $L \geq 3$, despite the fact that the EC region is complete for the corresponding two-descriptions problem. Consequently, the correlated quantization scheme (an extreme special case of VKG) that has been proven to be optimal for several cross sections of the L-channel quadratic Gaussian MD problem is strictly suboptimal in general. Using the encoding principles derived, we show that the CMS scheme achieves the complete rate-distortion region for several asymmetric cross sections of the L-channel quadratic Gaussian MD problem.**

***Index Terms*— Multiple descriptions coding, source coding, rate-distortion theory, combinatorial message sharing.**

Manuscript received April 20, 2013; revised December 21, 2014; accepted September 30, 2015. Date of publication December 8, 2015; date of current version January 18, 2016. This work was supported by the National Science Foundation through the Division of Computing and Communication Foundations under Grants CCF-1016861, CCF-0728986, CCF-1118075, and CCF-1320599. This paper was presented at the 2011 IEEE International Symposium on Information Theory [22], the 2012 IEEE Information Theory Workshop [23], and the 2013 IEEE International Symposium on Information Theory [26].

K. B. Viswanatha was with the Department of Electrical and Computer Engineering, University of California at Santa Barbara, Santa Barbara, CA 93106 USA. He is now with the Qualcomm Research Center, San Diego, CA 92121 USA (e-mail: kumar@ece.ucsb.edu).

E. Akyol was with the Department of Electrical and Computer Engineering, University of California at Santa Barbara, Santa Barbara, CA 93106 USA. He is now with the Department of Electrical and Computer Engineering, University of Illinois Urbana–Champaign, Champaign, IL 61820 USA (e-mail: akyol@illinois.edu).

K. Rose is with the Electrical and Computer Engineering Department, University of California at Santa Barbara, Santa Barbara, CA 93106 USA (e-mail: rose@ece.ucsb.edu).

Communicated by Y. Oohama, Associate Editor for Source Coding.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2015.2506586

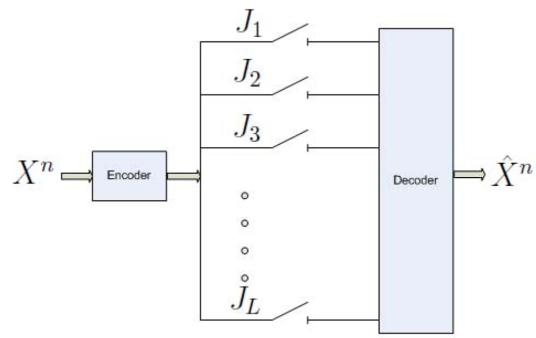

Fig. 1. $L-$Channel Multiple Descriptions Setup : Each description is received error free or is completely lost at the decoder.

I. INTRODUCTION

THE MULTIPLE Descriptions (MD) problem was proposed in the late seventies and has been studied extensively since, yielding a series of advances, ranging from the derivation of asymptotic bounds [1]–[10] to practical approaches for multiple descriptions quantizer design [11], [12]. It was originally viewed as a method to cope with channel failures, where multiple source descriptions are generated and sent over different paths. The encoder generates L descriptions for transmission over L available channels. It is assumed that the decoder receives a subset of the descriptions perfectly and the remaining are lost, as shown in Fig. 1. The objective of the MD problem is to design the encoders (for each description) and decoders (for each possible received subset of the descriptions), with respect to an overall rate-distortion (RD) trade-off. The subtlety of the problem is due to the balance between the full reconstruction quality versus quality of individual descriptions; or noise free quality versus the amount of redundancy across descriptions needed to achieve resilience to descriptions loss.

One of the first achievable regions for the 2-channel MD problem was derived by El-Gamal and Cover (EC) in 1982 [2]. It follows from Ozarow's results in [5] that the EC region is complete for the 2-channel quadratic Gaussian MD problem, i.e., when the source is Gaussian and the distortion measure is mean squared error (MSE). It was further shown by Ahlswede in [4] that the EC region is complete for a cross-section of the 2-description MD setup, called the 'no-excess rate regime', a scenario wherein the central decoder receives information at the minimum sum rate. This led to the popular belief that the EC achievable region is complete for the general 2-channel MD problem. However Zhang and Berger (ZB) in [3] proved a then surprising result, that the EC scheme is strictly sub-optimal in general. In particular, they showed that for a binary source under Hamming distortion, sending a common codeword in both descriptions can achieve points that are strictly outside the EC region. While introducing a common codeword implies explicit redundancy among the two descriptions, this codeword assists in better coordination between the descriptions leading to a strictly larger RD region.

Several researchers have since focused on extending EC and ZB to the $L-$channel MD problem [1], [7]–[10], [13], [14]. An achievable scheme, due to Venkataramani *et al.* (VGK) [1], directly builds on EC and ZB, and introduces a combinatorial number of refinement codebooks, one for each subset of the descriptions. Motivated by ZB, a *single* common codeword is also shared by all the descriptions, which assists in better coordination of the messages, improving the RD trade-off. For the $L-$channel quadratic Gaussian problem, it was shown by Wang and Viswanath in [9] that a special case of the VKG coding scheme, where no common codeword is sent, achieves the minimum sum rate when only the individual and the central distortion constraints are imposed. In particular, they showed that a 'correlated quantization' based encoding scheme, which is an extension of the Ozarow's encoding mechanism to the L-descriptions problem, achieves the minimum sum rate for the cross-section involving constraints only on the individual and central distortions. It was also shown recently by Chen in [10] that, in fact, this approach leads to the complete region for this particular cross-section.

Pradhan, Puri and Ramachandran (PPR) considered a practically interesting cross-section of the general $L-$channel MD problem in [7] and [8] called the 'symmetric MD problem' wherein it is assumed that the rates of all the descriptions are equal and the distortion is a function only of the 'number' of descriptions received rather than which particular subset is received. They proposed a new coding scheme leveraging principles from distributed source coding [15], [16], and particularly Slepian and Wolf's random binning techniques, and showed that, for this symmetric cross-section, the proposed encoding scheme improves upon the VKG region. Tian and Chen derived a new coding scheme in [17] for the symmetric MD problem which further extends the PPR region. It was also shown in [18] that this region is very close to complete for the symmetric quadratic Gaussian MD problem. Recently, Wang and Viswanath [13] derived a coding scheme based on the VKG and the PPR encoding principles and showed its sum-rate optimality for certain cross-sections of the quadratic Gaussian problem wherein only 2 layers of distortions are imposed. Moreover, in [19] , Song et.al. showed that, for the quadratic Gaussian setting, when the minimum sum rate is attained subject to two levels of distortion constraints (with the second level imposed on the complete set of descriptions), the PPR scheme also leads to the minimum achievable distortion at the intermediate levels.

In this paper we present a new encoding scheme involving "Combinatorial Message Sharing" (CMS), where a unique common codeword is sent in (shared by) each subset of the descriptions, thereby introducing a combinatorial number of *shared codebooks*, along with the refinement codebooks of [1]. The common codewords enable better coordination between descriptions, providing an improved overall RD region. We derive an achievable region for CMS and show that it subsumes VKG for general sources and distortion measures. Moreover, we show that CMS achieves a strictly larger region than VKG for all $L > 2$, if there exists a 2-description subset for which ZB achieves points strictly outside the EC region. In particular, CMS achieves strict improvement for a binary source under Hamming distortion. We then show a surprising result: CMS strictly outperforms VKG for a general class of sources and distortion measures, which includes several scenarios in which, for every 2-description subset, the ZB and the EC regions coincide. In particular, we show that for a Gaussian source under MSE, CMS achieves points strictly outside the VKG region. This result is in striking contract to the corresponding 2-descriptions setting. Optimality of EC for the 2-descriptions setup has led to a natural conjecture that common codewords do not play a necessary role in quadratic Gaussian MD coding, and all the achievable regions characterized so far neglect the common layer (see, e.g., [1], [10], [13]). In this paper, we show that the common codewords in CMS play a critical role in achieving the complete RD region for several asymmetric cross-sections of the general $L-$channel quadratic Gaussian MD problem. We note that the principles underlying CMS have been shown in precursor work to be useful in related applicational contexts of routing for networks with correlated sources [20] and data storage for selective retrieval [21]. We further note that a preliminary version of the results in this paper appeared in [22], [23], and [26]. We also note that, in this paper, we focus on generalizing the VKG coding scheme using a combinatorial number of shared messages for the MD problem with general sources and distortion measures. However, the CMS principle can be extended to incorporate random binning based encoding techniques, to utilize the underlying symmetry in the problem setup. Preliminary results in this direction have recently appeared in [24] and [25] and form part of our current research focus. The CMS scheme introduces a structured mechanism for generating random codebooks that lead to controlled redundancy across transmitted messages. The CMS scheme is only one of the several possible generalizations of the ZB scheme to L-descriptions. The fundamental idea of introducing structured codes and controlled redundancy for $L-$channel multiple descriptions problem is not new and has been explored in several publications in the past, including [7], [8], [17], and [27] and some of these schemes have been shown to outperform VKG for certain sources and distortion measures. However, to the best of our knowledge, for a Gaussian source, under MSE, CMS is the only generalization of the VKG scheme that achieves a strictly larger rate-distortion region. There are other encoding schemes that can achieve points strictly outside the VKG region, such as the PPR scheme, but are not strict generalizations VKG in the sense that their achievable region has not been shown to always subsume the entire VKG rate-distortion region. Hence there are fundamentally new insights to be gained by studying CMS.

The rest of the paper is organized as follows. In Section II, we formally state the $L-$channel MD setup and briefly describe the approaches and regions of EC [2], ZB [3] and VKG [1]. To keep the notation simple, we first describe in Section III-A, the CMS scheme for the 3 descriptions scenario and extend it to the general case in Section III-B. We then prove in section IV-A that CMS achieves strict improvement over VKG whenever there exists a 2-description subset for which ZB achieves points outside the EC region. In Section IV-B we prove that CMS outperforms VKG for a fairly general class of source-distortion measure pairs, including a Gaussian source under MSE. Finally in Section V, we derive new results for the $L-$channel quadratic Gaussian MD setup and show that CMS achieves the complete RD region for several asymmetric cross-sections of the general problem.

II. Formal Definitions and Prior Results

We use uppercase letters to denote random variables (e.g., X) and lowercase letters to denote their realizations (e.g., x). We use script letters to denote sets, alphabets and subscript indices (e.g., $\mathcal{S}, \mathcal{X}, \mathcal{K}$). A sequence of n independent and identically distributed (iid) random variables is denoted by X^n and its realization by x^n. $2^{\mathcal{S}}$ denotes the set of all subsets (power set) of any set $\mathcal{S}$ and $|\mathcal{S}|$ denotes the set cardinality. Note that $|2^{\mathcal{S}}| = 2^{|\mathcal{S}|}$. $\mathcal{S}^c$ denotes the set complement (the universal set will be explicitly specified when not obvious). For two sets $\mathcal{S}_1$ and $\mathcal{S}_2$, we denote the set difference by $\mathcal{S}_1 - \mathcal{S}_2 = \{k : k \in \mathcal{S}_1, k \notin \mathcal{S}_2\}$. All constant random variables, that take a single deterministic value, are denoted by Φ.

Throughout the paper, random variables, distortions and rates are indexed using sets, e.g., $U_{\mathcal{K}}, D_{\mathcal{K}}, R_{\mathcal{K}}$, where $\mathcal{K}$ takes values such as $\{1\}$, $\{2\}$, $\{1, 2\}$, $\{2, 3\}$ etc. These subscript indices are always subsets of the set $\{1, 2, \ldots, L\}$, where L denotes the number of descriptions, and satisfy all the standard properties of sets. For ease of notation, we often drop the braces and the comma whenever it is obvious. For example, $U_{\{1\}}$, $U_{\{2\}}$, $U_{\{1,2\}}$ and $U_{\{1,2,3\}}$ are abbreviated as U_1, U_2, U_{12} and U_{123}, respectively. Next, let $\mathcal{S}$ be a set of subscript indices. Then, we use the shorthand $\{U\}_{\mathcal{S}}$ to denote the set of variables $\{U_{\mathcal{K}} : \mathcal{K} \in \mathcal{S}\}$ (e.g., if $\mathcal{S} = \{\{1\}, \{2\}, \{1, 2\}\}$, then $\{U\}_{\mathcal{S}} = \{U_{\{1\}}, U_{\{2\}}, U_{\{1,2\}}\} = \{U_1, U_2, U_{12}\}$). Note the difference between $\{U\}_{\mathcal{S}}$ and $U_{\mathcal{K}}$. $\{U\}_{\mathcal{S}}$ is a set of variables, whereas $U_{\mathcal{K}}$ is a single variable. We use the notation in [28] to denote all standard information theoretic quantities. With a slight abuse of notation, we use $H(\cdot)$ to denote the entropy of a discrete random variable or the differential entropy of a continuous random variable.

We first give a formal definition of the $L-$channel MD problem. A source produces an iid sequence $X^n = \left(X^{(1)}, X^{(2)} \ldots, X^{(n)}\right)$, taking values over a finite alphabet $\mathcal{X}$. We denote $\mathcal{L} = \{1, \ldots, L\}$. There are L encoding functions, $f_l(\cdot), l \in \mathcal{L}$, which map X^n to the descriptions $J_l = f_l(X^n)$, where J_l takes values in the set $\{1, \ldots, B_l\}$. The rate of description l is defined as:

$$R_l = \frac{1}{n} \log_2(B_l) \tag{1}$$

Description l is sent over channel l and is either received at the decoder error-free or is completely lost. There are $2^L - 1$ decoding functions for each possible received combination of the descriptions $\hat{X}^n_{\mathcal{K}} = \left(\hat{X}^{(1)}_{\mathcal{K}}, \hat{X}^{(2)}_{\mathcal{K}} \ldots, \hat{X}^{(n)}_{\mathcal{K}}\right) = g_{\mathcal{K}}(J_l : l \in \mathcal{K})$, $\forall \mathcal{K} \subseteq \mathcal{L}, \mathcal{K} \neq \emptyset$, where $\hat{X}_{\mathcal{K}}$ takes values in a finite set $\hat{\mathcal{X}}_{\mathcal{K}}$, and $\emptyset$ denotes the null set. The distortion at the decoder when a subset $\mathcal{K}$ of the descriptions is received is:

$$D_{\mathcal{K}} = E\left[\frac{1}{n} \sum_{t=1}^{n} d_{\mathcal{K}}(X^{(t)}, \hat{X}^{(t)}_{\mathcal{K}})\right] \tag{2}$$

where $d_{\mathcal{K}} : \mathcal{X} \times \hat{\mathcal{X}}_{\mathcal{K}} \to \mathcal{R}$, is a well-defined bounded single letter distortion measure. We say that the RD tuple $(R_i, D_{\mathcal{K}} : i \in \mathcal{L}, \mathcal{K} \subseteq \mathcal{L}, \mathcal{K} \neq \emptyset)$ is achievable if there exist L encoding functions with rates $(R_1, \ldots, R_L)$ and $2^L - 1$ decoding functions yielding distortions $D_{\mathcal{K}}$. The closure of the set of all achievable RD tuples is defined as the '*L-channel multiple descriptions RD region*',[1] denoted hereafter by $\mathcal{RD}^L$.

A. 2 - Channel MD

1) El-Gamal-Cover Region: The first achievable region for the 2-channel MD problem was given by El-Gamal and Cover in 1982 [2]. Their region is denoted here by $\mathcal{RD}_{EC}$. It is given by the convex closure over all tuples $(R_1, R_2, D_1, D_2, D_{12})$ for which there exist auxiliary random variables U_1, U_2 and U_{12} (defined over arbitrary finite alphabets $\mathcal{U}_1$, $\mathcal{U}_2$ and $\mathcal{U}_{12}$, respectively), jointly distributed with X and functions $\psi_{\mathcal{K}}, \mathcal{K} \subseteq \{1, 2\}$, $\mathcal{K} \neq \emptyset$ such that,

$$\begin{aligned} R_1 &\geq I(X; U_1) \\ R_2 &\geq I(X; U_2) \\ R_1 + R_2 &\geq I(X; U_1, U_2, U_{12}) + I(U_1; U_2) \\ D_{\mathcal{K}} &\geq E\left[d_{\mathcal{K}}(X, \psi_{\mathcal{K}}(U_{\mathcal{K}}))\right], \quad \mathcal{K} \subseteq \{1, 2\}, \ \mathcal{K} \neq \emptyset \end{aligned} \tag{3}$$

Note that the original EC description assumes $\psi_{\mathcal{K}}$ to be identity functions (i.e., $\psi_{\mathcal{K}}(U_k) = U_{\mathcal{K}}$). However, the two characterizations are equivalent and we use the one described above as it is easier to relate to the L-channel achievable regions. We also note that the function $\psi_{12}(\cdot)$ can, in general, be made to depend on U_1, U_2 and U_{12}. However, the overall RD region remains unchanged. To see this, let us say, for some joint distribution, $\psi_{12}(\cdot)$ depends on U_1, U_2 and U_{12}. We can always construct a new joint distribution with the random variable U_{12} set equal to $\psi_{12}(U_1, U_2, U_{12})$. This new joint distribution satisfies all the rate and distortion conditions in (3) with the new $\psi_{12}(\cdot)$ being an identity function.

The order of codebook generation is shown in Fig. 2(a). As stated in the introduction, the EC region can be shown to be complete for the 2-descriptions quadratic Gaussian MD problem [5] and for the 'no excess rate' regime for general

[1] Note that this region has $L + 2^L - 1$ dimensions.

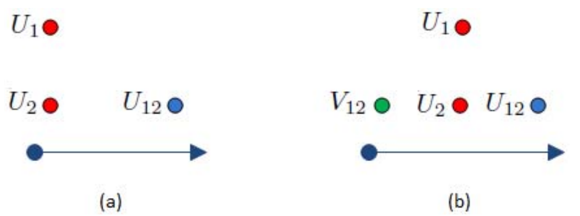

Fig. 2. Codebook generation for EC (a) and ZB (b) coding schemes. The arrow mark indicates the order in which the codebooks are generated.

sources and distortion measures [4]. The achievability of the above region is described as follows. 2^{nR_1} codewords of U_1 and 2^{nR_2} codewords of U_2 are generated, each of length n, using the marginal distributions of U_1 and U_2, respectively. For each pair (u_1^n, u_2^n) of codewords, a codeword for U_{12} is generated according to the conditional density $\prod_{t=1}^n P_{U_{12}|U_1,U_2}(u_{12}^{(t)}|u_1^{(t)}, u_2^{(t)})$. Given a sequence X^n, the encoder looks for a triplet of codewords that are jointly typical with the observed sequence. Information at rate R_1 is sent in description 1 and at rate R_2 in description 2. On receiving either of the the two descriptions, the decoder estimates the source based on the corresponding codeword of U_1 or U_2. However, if it receives both the descriptions, it estimates the source based on U_{12}. El-Gamal and Cover showed that, for joint typicality encoding, the probability of error asymptotically approaches zero if the rates satisfy conditions in (3). As the distortion measures are bounded, joint typicality assures the distortion constraints to be satisfied.

2) Zhang-Berger Region: Given the optimality of EC for the no-excess rate scenario and for the quadratic Gaussian setup, it was naturally conjectured that it is optimal in general. However, Zhang and Berger [3] proposed a new coding scheme and an associated rate-distortion region for the 2-channel MD problem and showed that it strictly subsumes $\mathcal{RD}_{EC}$. In particular, they showed that for a binary symmetric source under Hamming distortion, points outside $\mathcal{RD}_{EC}$ can be achieved by sending a common codeword in both the descriptions. This extended region is denoted here by $\mathcal{RD}_{ZB}$. They showed that, any tuple is achievable for which there exist auxiliary random variables[2] V_{12}, U_1, U_2, U_{12}, jointly distributed with X, for which there exist functions $\psi_{\mathcal{K}}$, $\mathcal{K} \subseteq \mathcal{L}, \mathcal{K} \neq \emptyset$ such that,

$$\begin{aligned} R_1 &\geq I(X; U_1, V_{12}) \\ R_2 &\geq I(X; U_2, V_{12}) \\ R_1 + R_2 &\geq 2I(X; V_{12}) + I(X; U_1, U_2, U_{12}|V_{12}) \\ &\quad + I(U_1; U_2|V_{12}) \\ D_{\mathcal{K}} &\geq E\left[d_{\mathcal{K}}(X, \psi_{\mathcal{K}}(U_{\mathcal{K}}))\right], \quad \mathcal{K} \subseteq \{1, 2\},\ \mathcal{K} \neq \emptyset \end{aligned} \tag{4}$$

The ZB region is given by the closure of all such rate-distortion tuples. Again note that the ZB region is described differently in [3], and does not include U_{12} in its characterization. However, it was recently shown in [6] that the above characterization is equivalent to the original one in [3]. Further note that the above region is convex and does not require 'convexification' as is the case for $\mathcal{RD}_{EC}$.

The codebooks are generated as follows: first, $2^{nR''_{12}}$ codewords of V_{12}, each of length n, are generated according to the marginal density of V_{12}. Conditioned on each codeword of V_{12}, $2^{nR'_1}$ codewords of U_1 and $2^{nR'_2}$ codewords of U_2 are generated according to the conditional densities $\prod_{t=1}^n P_{U_1|V_{12}}(u_1^{(t)}|v_{12}^{(t)})$ and $\prod_{t=1}^n P_{U_2|V_{12}}(u_2^{(t)}|v_{12}^{(t)})$, respectively. For each codeword tuple (v_{12}^n, u_1^n, u_2^n), a codeword for U_{12} is generated according to the conditional density $\prod_{t=1}^n P_{U_{12}|U_1,U_2,V_{12}}(u_{12}^{(t)}|u_1^{(t)}, u_2^{(t)}, v_{12}^{(t)})$. Similar to EC, descriptions 1 and 2 carry information about codewords of U_1 and U_2, respectively. However, along with these '*private*' messages, both the descriptions carry a common component, which is the codeword of V_{12}. Although this common codeword introduces explicit redundancy, it helps to co-ordinate the two messages well and therefore provides better overall efficiency. The encoding structure of the random variables is shown in Fig. 2(b).

Note the functional difference of the two random variables U_{12} and V_{12}. The codeword corresponding to V_{12} is sent in *both* the descriptions, whereas the information corresponding to the codeword of U_{12} is (implicitly) *split* between the two descriptions. We therefore call V_{12} the '*common random variable*' and U_{12} as the '*refinement random variable*'. Random variables U_1 and U_2 form the so called '*base layer random variables*'.

B. L-Channel MD

An extension of $\mathcal{RD}_{EC}$ and $\mathcal{RD}_{ZB}$ to the $L-$channel setup was proposed by Venkataramani et al. in [1]. The resulting region is denoted here by $\mathcal{RD}_{VKG}$, and described as follows. Let $(V_{\mathcal{L}}, \{U\}_{2^{\mathcal{L}}-\emptyset})$ be any set of 2^L random variables distributed jointly with X. Then, an RD tuple is said to be achievable if there exist functions $\psi_{\mathcal{K}}(\cdot)$ such that:

$$\begin{aligned} \sum_{l \in \mathcal{K}} R_l &\geq |\mathcal{K}| I(X; V_{\mathcal{L}}) - H(\{U\}_{2^{\mathcal{K}}-\emptyset}|X, V_{\mathcal{L}}) \\ &\quad + \sum_{\mathcal{A} \subseteq \mathcal{K}} H\left(U_{\mathcal{A}}|\{U\}_{2^{\mathcal{A}}-\emptyset-\mathcal{A}}\right) \end{aligned} \tag{5}$$

$$D_{\mathcal{K}} \geq E\left[d_{\mathcal{K}}\left(X, \psi_{\mathcal{K}}(U_{\mathcal{K}})\right)\right] \tag{6}$$

$\forall \mathcal{K} \subseteq \mathcal{L}$. The closure of the achievable tuples over all such 2^L random variables is $\mathcal{RD}_{VKG}$. Here, we only present an overview of the encoding scheme. The order of codebook generation of the auxiliary random variables is shown in Fig. 3. First, $2^{nR''_{\mathcal{L}}}$ codewords of $V_{\mathcal{L}}$ are generated using the marginal distribution of $V_{\mathcal{L}}$. Conditioned on each codeword of $V_{\mathcal{L}}$, $2^{nR'_l}$ codewords of U_l are generated according to their respective conditional densities. Next, for each $j \in (1, \ldots, 2^{n(R''_{\mathcal{L}} + \sum_{l \in \mathcal{K}} R'_l)})$, a single codeword is generated for $U_{\mathcal{K}}(j)$ conditioned on $(v_{\mathcal{L}}(j), \{u(j)\}_{2^{\mathcal{K}}-\emptyset-\mathcal{K}})$ $\forall \mathcal{K} \subseteq \mathcal{L}$, $|\mathcal{K}| > 1$. Note that to generate the codebook for $U_{\mathcal{K}}$, we first need the codebooks for all $\{U\}_{2^{\mathcal{K}}-\emptyset-\mathcal{K}}$ and $V_{\mathcal{L}}$.

On observing a typical sequence X^n, the encoder tries to find a jointly typical codeword tuple one from each codebook. Codeword index of U_l (at rate R'_l) is sent in description l. Along with the '*private*' message, each description also

[2]Our reason for the different notational choice for V_{12} will become evident later.

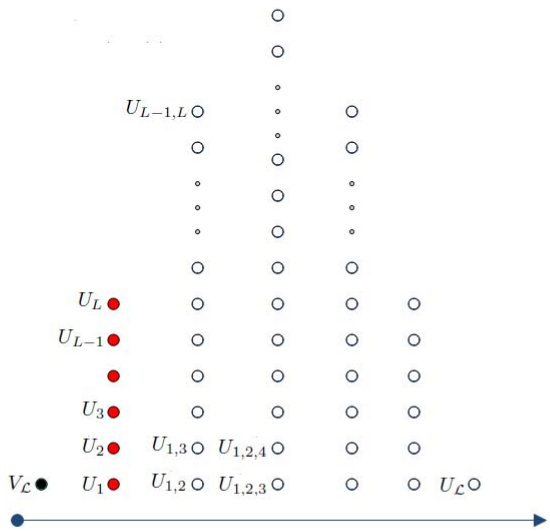

Fig. 3. Codebook generation for VKG coding scheme. Green ($V_{\mathcal{L}}$) indicates 'common random variable'. Red ($U_l l \in \mathcal{L}$) indicates 'base layer random variables' and Blue ($U_{\mathcal{S}}$ $|\mathcal{S}| > 1$) indicates 'refinement random variables'.

carries a '*shared message*' at rate $R''_{\mathcal{L}}$, which is the codeword index of $V_{\mathcal{L}}$. Hence, the rate for each description is $R_l = R'_l + R''_{\mathcal{L}}$. VKG showed that, to ensure finding a set of jointly typical codewords with the observed sequence, the rates must satisfy (5). It then follows from standard arguments (see, e.g., [29]) that, if the random variables also satisfy (6), then the distortion constraints are met. Note that $V_{\mathcal{L}}$ is the *only* shared random variable. $U_l : l \in \mathcal{L}$ form the base layer random variables and all $U_{\mathcal{K}} : |\mathcal{K}| \geq 2$ form the refinement layers. Observe that the codebook generation follows the order: shared layer → base layer → refinement layer. Also, observe that the random variable $V_{\mathcal{L}}$ plays the role of V_{12} in the ZB scheme.[3]

III. Combinatorial Message Sharing

In this section, we describe the proposed encoding scheme and derive the new achievable region. To simplify notation and understanding, we first describe the scheme for the 3-descriptions case and offer intuitive arguments to show the achievability of the new region. We then extend the arguments to the $L-$channel case and provide formal proofs as part of Theorem 1.

A. 3-Descriptions Scenario

The encoding order for the 3-descriptions VKG scheme is shown in Fig. 4(a). Recall that the common codeword helps in coordinating the 3 descriptions. The VKG encoding scheme employs *one* common codeword ($V_{\mathcal{L}}$) that is sent in all the L descriptions. However, when dealing with $L > 2$ descriptions, restricting to a single shared message could be suboptimal. The CMS scheme therefore allows for 'combinatorial message sharing', i.e., a common codeword is sent in each (non-empty) subset of the descriptions.

[3]It has been recently shown in [6] that the last layer of refinement random variables ($U_{\mathcal{L}}$) can be set to be a deterministic function of all the remaining random variables without affecting the overall achievable region (for the 2-channel MD case, this corresponds to setting U_{12} to be a function of V_{12}, U_1 and U_2). A similar result can be proven even for the proposed coding scheme. However, we continue to use this last layer in our theorems to avoid additional notation.

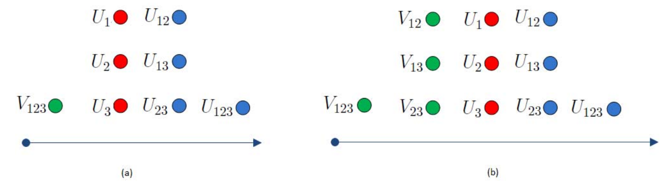

Fig. 4. (a) Denotes the codebook generation order for VKG scheme. (b) represents the codebook generation for the proposed coding scheme.

The shared random variables are denoted by 'V'. The base and the refinement layer random variables are denoted by 'U'. For the 3 descriptions scenario, we have 11 auxiliary random variables which include 3 additional variables over the VKG scheme. These variables are denoted by V_{12}, V_{13} and V_{23}. The codeword corresponding to $V_{\mathcal{K}}$ is sent in *all* the descriptions $l \in \mathcal{K}$. For example, the codeword of V_{12} is sent in both the descriptions 1 and 2. This introduces a new layer in the encoding structure as shown in Fig. 4 (b). This extra encoding layer provides an additional degree of freedom in controlling the redundancy across the messages, and hence leads to an improved rate-distortion region.

The codebook generation is done as follows. First, the codebook for V_{123} is generated containing $2^{nR''_{123}}$ independently generated codewords, each generated according to the marginal $P(V_{123})$. Then codebooks for V_{12}, V_{13} and V_{23} (each containing $2^{nR''_{12}}$, $2^{nR''_{13}}$ and $2^{nR''_{23}}$ codewords, respectively) are independently generated conditioned on each codeword of V_{123}, according to the respective conditional densities. Next, the base layer codebooks for U_l, $l \in \{1, 2, 3\}$ (each containing $2^{nR'_l}$ codewords) are generated conditioned on the codewords of all $V_{\mathcal{K}}$ such that $l \in \mathcal{K}$. For example, the codebooks for U_1 are generated conditioned on the codewords of V_{12}, V_{13} and V_{123}. Note that each codebook of U_1 contains $2^{nR'_1}$ codewords and there are such $2^{n(R''_{12}+R''_{13}+R''_{123})}$ codebooks.

The refinement layer codewords are generated similar to the VKG scheme. However, the codewords for $U_{\mathcal{K}}$ are now generated conditioned not only on the codewords of $\{U\}_{2^{\mathcal{K}}-\emptyset}$ and $V_{\mathcal{L}}$, but also on the codewords of all $V_{\mathcal{A}}$ such that $|\mathcal{A} \cap \mathcal{K}| > 0$. For example, a single codeword for U_{12} is generated conditioned on each codeword tuple of $\{U_1, U_2, V_{12}, V_{13}, V_{23}, V_{123}\}$. Note that, overall, there are $2^{n(R'_1+R'_2+R''_{12}+R''_{13}+R''_{123})}$ codewords of U_{12} that are generated. Similarly codewords for U_{13}, U_{23} and U_{123} are generated conditioned on codewords of $\{U_1, U_3, V_{12}, V_{13}, V_{23}, V_{123}\}$, $\{U_2, U_3, V_{12}, V_{13}, V_{23}, V_{123}\}$ and $\{U_1, U_2, U_3, V_{12}, V_{13}, V_{23}, V_{123}\}$, respectively.

The encoder, on observing X^n, tries to find a codeword from the codebook of V_{123} that is jointly typical with X^n. Using typicality arguments, it is possible to show that the probability of not finding such a codeword approaches zero if $R''_{123} > I(X; V_{123})$. Let us denote the selected codeword of V_{123} by v_{123}. The encoder next looks at the codebooks of V_{12}, V_{13} and V_{23}, which were generated conditioned on v_{123}, to find a triplet of codewords which are jointly typical with (X^n, v_{123}). It can be shown using arguments similar to [1] and [29] (formal proof is given in Theorem 1), that the probability of not finding such a triplet approaches zero if the following conditions are satisfied $\forall \mathcal{S} \subseteq \{\{1,2\},\{1,3\},\{2,3\}\}$:

$$\sum_{\mathcal{K}\in\mathcal{S}} R''_{\mathcal{K}} > \sum_{\mathcal{K}\in\mathcal{S}} H\left(V_{\mathcal{K}}|V_{123}\right) - H\left(\{V\}_{\mathcal{S}}|V_{123}, X\right) \quad (7)$$

Let us denote these codewords by v_{12}, v_{13} and v_{23}. The encoders next step is to find an index tuple (i_1, i_2, i_3) such that $(U_1(i_1), U_2(i_2), U_3(i_3), U_{12}(i_1,i_2), U_{13}(i_1,i_3), U_{23}(i_2,i_3), U_{123}(i_1,i_2,i_3))$ is jointly typical with $(X^n, v_{123}, v_{12}, v_{13}, v_{23})$. Here, $U_1(i_1)$ denotes the i_1th codeword from the codebook of U_1 and $U_{12}(i_1,i_2)$ denotes the codeword of U_{12} generated conditioned on $(v_{123}, v_{12}, v_{13}, v_{23}, U_1(i_1), U_2(i_2))$. Similar notation is followed for $U_2(i_2), U_3(i_3), U_{13}(i_1,i_3)$, etc. Again using arguments similar to [1], we can show that the probability of not finding such an index tuple approaches zero if $\forall \mathcal{S} \subseteq \{1,2,3\}$:

$$\sum_{\mathcal{K}\in\mathcal{S}} R'_{\mathcal{K}} > \sum_{\mathcal{K}\subseteq\mathcal{S}} H\left(U_{\mathcal{K}}|\{U\}_{2^{\mathcal{K}}-\emptyset-\{\mathcal{K}\}}, \{V\}_{\mathcal{J}(\mathcal{K})}\right) - H\left(\{U\}_{2^{\mathcal{S}}-\emptyset}|V_{12}, V_{13}, V_{23}, V_{123}, X\right) \quad (8)$$

where $\mathcal{J}(\mathcal{K}) = \{\mathcal{A} : \mathcal{A} \in \{12, 13, 23, 123\}, |\mathcal{K}\bigcap\mathcal{A}| > 0\}$. We denote the codewords corresponding to this index tuple by $(u_1, u_2, u_3, u_{12}, u_{13}, u_{23}, u_{123})$. Note that, in the above illustration, we assumed that the encoder finds jointly typical codewords in a sequential manner, i.e., it first finds a codeword of V_{123}, then finds jointly typical codewords from the codebooks of (V_{12}, V_{13}, V_{23}) and so on. This was done only for the ease of understanding. In Theorem 1, we derive the conditions on rates for the encoder to find typical sequences from all the codebooks jointly (at once). The conditions on the rates for joint encoding is generally weaker (the region is larger) than that for sequential encoding.

The encoder sends the index of the base layer codewords in the corresponding descriptions. It also sends the index of codewords corresponding to $V_{\mathcal{K}}$ in *all* the descriptions $l \in \mathcal{K}$. For example, in description 1, the encoder sends the indices corresponding to v_{123}, v_{12}, v_{13} and u_1. Similarly, descriptions 2 and 3 carry indices corresponding to $(v_{123}, v_{12}, v_{23}, u_2)$ and $(v_{123}, v_{13}, v_{23}, u_3)$, respectively. Therefore, rate for description l is:

$$R_l = R'_l + \sum_{\mathcal{K}\in\mathcal{J}(l)} R''_{\mathcal{K}} \quad (9)$$

Conditions on R_l can be obtained by substituting bounds from (7) and (8).

The decoder, on receiving a subset of descriptions, estimates X^n based on the refinement layer codeword corresponding to the received subset. For example, if the decoder receives the descriptions 1 and 2, it estimates X^n as $\psi_{12}(u_{12})$ for some function $\psi_{12}(\cdot)$. It follows from standard arguments [29] that, if there exist functions $\psi_{\mathcal{K}}(\cdot)$ satisfying the following constraints, then distortion vector $\{D_{\mathcal{K}}, \forall \mathcal{K}\in 2^{\mathcal{L}}-\emptyset\}$ is achievable.

$$D_{\mathcal{K}} \geq E\left[d_{\mathcal{K}}\left(X, \psi_{\mathcal{K}}\left(U_{\mathcal{K}}\right)\right)\right] \quad (10)$$

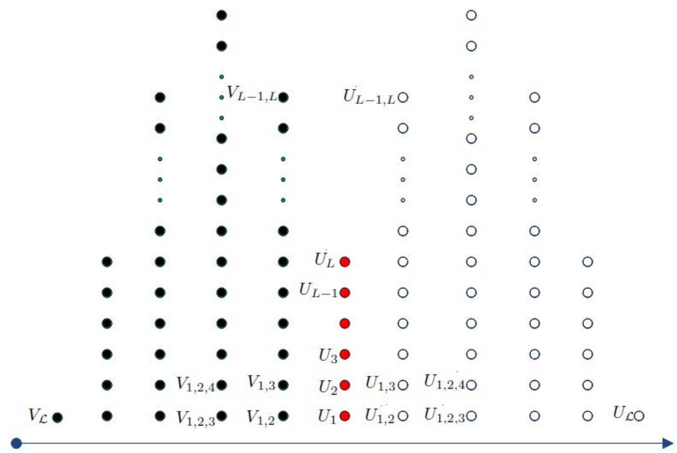

Fig. 5. Codebook generation for the proposed coding scheme. Black ($V_{\mathcal{S}}$ $|\mathcal{S}| > 1$) indicates 'shared random variable'. Red (U_l $l \in \mathcal{L}$) indicates 'base layer random variables' and White ($U_{\mathcal{S}}$ $|\mathcal{S}| > 1$) indicates 'refinement random variables'.

An achievable RD region for the 3-descriptions MD setup is obtained by taking the closure of the achievable tuples over all such 11 auxiliary random variables, jointly distributed with X.

B. L Channel Case

The fundamental idea here is same as the 3-descriptions scenario, albeit involving more complex notation. The codebook generation is done in an order as shown in Fig. 5. First, the codebook for $V_{\mathcal{L}}$ is generated. Then, the codebooks for $V_{\mathcal{K}}$, $|\mathcal{K}| = W$ are generated in the order $W = L-1, L-2 \ldots 2$. This is followed by the generation of the base layer codebooks, i.e, $U_{\mathcal{K}}$, $|\mathcal{K}| = 1$. Then, the refinement layer codebooks corresponding to $U_{\mathcal{K}}$, $|\mathcal{K}| = W$ are generated in the order $W = 2, 3 \ldots, L$. Each codebook is generated conditioned on a subset of the previously generated codewords. The specifics of codebook generation will be described as part of the proof of Theorem 1.

Before stating the theorem, we define the following subsets of $2^{\mathcal{L}}$:

$$\begin{aligned} \mathcal{I}_W &= \{\mathcal{K} : \mathcal{K} \in 2^{\mathcal{L}}, |\mathcal{K}| = W\} \\ \mathcal{I}_{W+} &= \{\mathcal{K} : \mathcal{K} \in 2^{\mathcal{L}}, |\mathcal{K}| > W\} \end{aligned} \quad (11)$$

Let $\mathcal{B}$ be any non-empty subset of $\mathcal{L}$ with $|\mathcal{B}| \leq W$. We define the following subsets of $\mathcal{I}_W$ and $\mathcal{I}_{W+}$:

$$\begin{aligned} \mathcal{I}_W(\mathcal{B}) &= \{\mathcal{K} : \mathcal{K} \in \mathcal{I}_W, \mathcal{B} \subseteq \mathcal{K}\} \\ \mathcal{I}_{W+}(\mathcal{B}) &= \{\mathcal{K} : \mathcal{K} \in \mathcal{I}_{W+}, \mathcal{B} \subseteq \mathcal{K}\} \end{aligned} \quad (12)$$

We also define:

$$\mathcal{J}(\mathcal{B}) = \{\mathcal{K} : \mathcal{K} \in 2^{\mathcal{L}}, |\mathcal{B}\bigcap\mathcal{K}| > 0\} \quad (13)$$

Note that $\mathcal{J}(\mathcal{L}) = 2^{\mathcal{L}} - \emptyset$.

Next we consider subsets of $2^{\mathcal{L}} - \emptyset$ and define some important notation. Let $\mathcal{Q}$ be any subset of $2^{\mathcal{L}} - \emptyset$. We denote by $[\mathcal{Q}]_1$ the set of all elements of $\mathcal{Q}$ of cardinality 1, i.e.,:

$$[\mathcal{Q}]_1 = \{\mathcal{K} : \mathcal{K} \in \mathcal{Q}, |\mathcal{K}| = 1\} \quad (14)$$

Finally, we say that $\mathcal{Q} \in \mathcal{Q}^*$ if it satisfies the following property.

Definition 1: Let $\mathcal{Q}$ be a subset of $2^{\mathcal{L}} - \emptyset$. If, for every set $\mathcal{K}$ that belongs to $\mathcal{Q}$, all the sets $\mathcal{I}_{|\mathcal{K}|+}(\mathcal{K})$ also belong to $\mathcal{Q}$, then we say that $\mathcal{Q} \in \mathcal{Q}^*$, i.e., $\mathcal{Q} \in \mathcal{Q}^*$ if $\forall \mathcal{K} \in \mathcal{Q}$:

$$\mathcal{K} \in \mathcal{Q} \Rightarrow \mathcal{I}_{|\mathcal{K}|+}(\mathcal{K}) \subset \mathcal{Q} \tag{15}$$

Let $(\{V\}_{\mathcal{J}(\mathcal{L})}, \{U\}_{2^{\mathcal{L}}-\emptyset})$ be any set of $2^{L+1} - L - 2$ random variables jointly distributed with X. Let $\mathcal{Q} \subseteq 2^{\mathcal{L}} - \emptyset$ such that $\mathcal{Q} \in \mathcal{Q}^*$. We define:

$$\begin{aligned}\alpha(\mathcal{Q}) = &\sum_{\mathcal{K}\in\mathcal{Q}-[\mathcal{Q}]_1} H\left(V_{\mathcal{K}}|\{V\}_{\mathcal{I}_{|\mathcal{K}|+}(\mathcal{K})}\right)\\ &+ \sum_{\mathcal{K}\in 2^{[\mathcal{Q}]_1}-\emptyset} H\left(U_{\mathcal{K}}|\{V\}_{\mathcal{I}_{1+}(\mathcal{K})}, \{U\}_{2^{\mathcal{K}}-\emptyset-\mathcal{K}}\right)\\ &- H\left(\{V\}_{\mathcal{Q}-[\mathcal{Q}]_1}, \{U\}_{2^{[\mathcal{Q}]_1}-\emptyset}|X\right)\end{aligned} \tag{16}$$

We follow the convention $\alpha(\emptyset) = 0$. Next we state the rate-distortion region achievable by the CMS scheme for the $L-$descriptions framework.

Theorem 1: Let $(\{V\}_{\mathcal{J}(\mathcal{L})}, \{U\}_{2^{\mathcal{L}}-\emptyset})$ be any set of $2^{L+1} - L - 2$ random variables jointly distributed with X, where $U_{\mathcal{K}}$ and $V_{\mathcal{K}}$ take values in some finite alphabets $\mathcal{U}_{\mathcal{K}}$ and $\mathcal{V}_{\mathcal{K}}$, respectively $\forall \mathcal{K}$. Let $\mathcal{Q}^*$ be the set of all subsets of $2^{\mathcal{L}} - \emptyset$ satisfying (15) and let $R''_{\mathcal{S}}, \mathcal{S} \in \mathcal{I}_{1+}$ and $R'_l, l \in \mathcal{L}$ be $2^L - 1$ auxiliary rates satisfying:

$$\sum_{\mathcal{S}\in\mathcal{Q}-[\mathcal{Q}]_1} R''_{\mathcal{S}} + \sum_{l\in[\mathcal{Q}]_1} R'_l > \alpha(\mathcal{Q}) \forall \mathcal{Q} \in \mathcal{Q}^* \tag{17}$$

Then, the RD region for the $L-$channel MD problem contains the rates and distortions for which there exist functions $\psi_{\mathcal{K}}(\cdot)$, such that,

$$R_l = R'_l + \sum_{\mathcal{K}\in\mathcal{J}(l)} R''_{\mathcal{S}} \tag{18}$$

$$D_{\mathcal{K}} \geq E\left[d_{\mathcal{K}}\left(X, \psi_{\mathcal{K}}\left(U_{\mathcal{K}}\right)\right)\right] \tag{19}$$

The closure of the achievable tuples over all such $2^{L+1} - L - 2$ random variables is denoted by $\mathcal{RD}_{CMS}$.

Remark 1: $\mathcal{RD}_{CMS}$ can be extended to continuous random variables and well-defined distortion measures using techniques similar to [30]. We omit the details here and assume that the above region continues to hold even for well behaved continuous random variables (for example, a Gaussian source under MSE).

Remark 2: $\mathcal{RD}_{CMS}$ is convex, as a time sharing random variable can be embedded in $V_{\mathcal{L}}$.

Proof (Codebook Generation): Suppose we are given $P\left(\{v\}_{2^{\mathcal{L}}-\{(1),(2),\ldots(L)\}}, \{u\}_{2^{\mathcal{L}}}|x\right)$ and $\psi_{\mathcal{K}}(\cdot)$ satisfying (19). The codebook generation begins with $V_{\mathcal{L}}$. We independently generate $2^{nR''_{\mathcal{L}}}$ codewords of $V_{\mathcal{L}}$, denoted by $v^n_{\mathcal{L}}(j_{\mathcal{L}}) j_{\mathcal{L}} \in \{1 \ldots 2^{nR''_{\mathcal{L}}}\}$, according to the density $\prod_{t=1}^n P_{V_{\mathcal{L}}}(v^{(t)}_{\mathcal{L}})$. For each codeword $v^n_{\mathcal{L}}(j_{\mathcal{L}})$, we independently generate $2^{nR''_{\mathcal{K}}}$ codewords of $V_{\mathcal{K}}$ $\forall \mathcal{K} \in \mathcal{I}_{L-1}$, according to $\prod_{t=1}^n P_{V_{\mathcal{K}}|V_{\mathcal{L}}}(v^{(t)}_{\mathcal{K}}|v^{(t)}_{\mathcal{L}})$. We denote these codewords by $v^n_{\mathcal{K}}(j_{\mathcal{L}}, j_{\mathcal{K}})$ $j_{\mathcal{K}} \in \{1 \ldots 2^{nR''_{\mathcal{K}}}\}$. This procedure for generating the codebooks of the shared random variables continues. $2^{nR''_{\mathcal{K}}}$ codewords of $V_{\mathcal{K}}$ are independently generated for each codeword tuple of $\{V\}_{\mathcal{I}_{W+}(\mathcal{K})}$ according to $\prod_{t=1}^n P_{V_{\mathcal{K}}|\{V\}_{\mathcal{I}_{W+}(\mathcal{K})}}(v^{(t)}_{\mathcal{K}}|\{v\}^{(t)}_{\mathcal{I}_{W+}(\mathcal{K})})$. These codewords are denoted by $v^n_{\mathcal{K}}(\{j\}_{\mathcal{I}_{W+}(\mathcal{K})}, j_{\mathcal{K}}) j_{\mathcal{K}} \in \{1 \ldots 2^{nR''_{\mathcal{K}}}\}$. Note that to generate the codebooks for $V_{\mathcal{K}}$ $\forall \mathcal{K} \in \mathcal{I}_W$, we need the codebooks of $V_{\mathcal{A}} \forall \mathcal{A} \in \mathcal{I}_{W+}(\mathcal{K})$. The codebook generation follows the order indicated in Fig. 5.

Once all the codebooks of shared random variables are generated, the codebooks for the base layer random variables are generated. For each codeword tuple of $\{V\}_{\mathcal{I}_{1+}(l)}$, $2^{nR'_l}$ codewords of U_l are generated independently according to $\prod_{t=1}^n P_{U_l|(\{V\}_{\mathcal{I}_{1+}(l)})}(u^{(t)}_l|\{v\}^{(t)}_{\mathcal{I}_{1+}(l)})$ and are denoted by $u^n_l\left(\{j\}_{\mathcal{I}_{1+}(l)}, i_l\right)$ $i_l \in \{1 \ldots 2^{nR'_l}\}$. Then the codebooks for the refinement layers are formed by assigning a single codeword $u^n_{\mathcal{K}}(\{j\}_{\mathcal{J}(\mathcal{K})}, \{i\}_{\mathcal{K}})$ to each $\mathcal{K} \in 2^{\mathcal{L}} - \{\{1\}, \{2\}, \ldots \{L\}\}$ and $\forall \{j\}_{\mathcal{J}(\mathcal{K})}, \{i\}_{\mathcal{K}}$. These codewords are generated according to $\prod_{t=1}^n P_{U_{\mathcal{K}}|\{V\}_{\mathcal{J}(\mathcal{K})}, \{U\}_{2^{\mathcal{K}}-\{\mathcal{K}\}}}(u^{(t)}_{\mathcal{K}}|\{v\}^{(t)}_{\mathcal{J}(\mathcal{K})}, \{u\}^{(t)}_{2^{\mathcal{K}}-\{\mathcal{K}\}})$.

The encoder, on observing a typical sequence X^n, attempts to find a set of codewords, one for each variable, such that they are all jointly typical. If the encoder succeeds in finding such a set, from the typical average lemma [29], it follows that the average distortions are smaller than $D_{\mathcal{K}}, \forall \mathcal{K} \in 2^{\mathcal{L}}$. However, if the encoder fails to find such a set, the average distortions are upper bounded by d_{max} (as the distortion measures are assumed to be bounded). Hence, if the probability of finding a set of jointly typical codewords approaches 1, the distortion conditions (19) are satisfied. We show in Appendix A that if the rates $R''_{\mathcal{K}}$ and R'_l satisfy conditions (17), this probability approaches 1. Next, recall that the encoder sends the codewords of $V_{\mathcal{K}}$ (at rate $R''_{\mathcal{K}}$) in all the descriptions $l \in \mathcal{K}$. It also sends the codewords of U_l (at rate R'_l) in description l. Therefore the rate of description l is given by (18), proving the Theorem. □

We note that the characterization of the achievable RD region in Theorem 1 involves auxiliary rates R'_l and $R''_{\mathcal{K}}$. This makes it hard to prove general converse results. However deriving an explicit characterization only in terms of R_l is more involved and deriving such bounds will be considered as part of future work.

IV. Strict Improvement

A. Binary Symmetric Source Under Hamming Distortion Measure

In this section we show that the CMS scheme achieves points strictly outside $\mathcal{RD}_{VKG}$ whenever there exists a 2-description subset for which the ZB region achieves points outside the EC region. As a result of this theorem, it follows that for a Binary symmetric source under Hamming distortion measure, CMS strictly outperforms VKG.

Theorem 2:

(i) The rate-distortion region achievable by the CMS scheme is always at least as large as the VKG achievable region, i.e.:

$$\mathcal{RD}_{VKG} \subseteq \mathcal{RD}_{CMS} \tag{20}$$

(ii) $\forall L > 2$, the CMS scheme achieves points that are not achievable by VKG for any source and distortion measures for which there exists a 2-description subset

such that the Zhang-Berger scheme achieves points outside the El-Gamal-Cover region. i.e.:

$$\mathcal{RD}_{VKG} \subset \mathcal{RD}_{CMS} \text{ if } \mathcal{RD}_{EC} \subset \mathcal{RD}_{ZB} \tag{21}$$

for some 2-description subset. Specifically, for a binary symmetric source under Hamming distortion, CMS achieves a strictly larger rate-distortion region compared to VKG $\forall L > 2$.

Remark 3: In Section IV-B we will show a more surprising result that CMS achieve points strictly outside $\mathcal{RD}_{VKG}$ even in scenarios where $\mathcal{RD}_{ZB} = \mathcal{RD}_{EC}$ for every 2-description subset. Moreover, we will show that CMS outperforms VKG for a Gaussian source under MSE, a setting for which it is well known that the EC and ZB regions coincide.

Proof: Part (i) of the theorem is a straightforward corollary of Theorem 1. It follows directly by setting $V_{\mathcal{K}} = \Phi$, $\forall \mathcal{K}$ such that $|\mathcal{K}| < L$ in $\mathcal{RD}_{CMS}$ (where Φ denotes a constant). Substituting in (18), and setting the auxiliary rate $R''_{\mathcal{L}} = I(X; V_{\mathcal{L}})$, we get the same conditions as (6).

Part (ii) of the theorem, in fact, follows directly from Zhang and Berger's result. To see this, without loss of generality, let descriptions $\{1, 2\}$ be the 2-description subset for which ZB achieves points outside EC region. Let us consider a cross-section of the $L-$channel MD problem where we only consider R_1, R_2, D_1, D_2 and D_{12} and set the remaining rates to 0 while ignoring all the remaining distortion constraints. Clearly, the CMS scheme leads to the ZB region as the common layer codeword V_{12} is sent as part of both descriptions 1 and 2. However, observe that, VKG leads to EC as $V_{\mathcal{L}}$ must be set to constant to ensure $R_3 = \cdots = R_L = 0$. Although this example is sufficient to prove Part (ii) of the theorem, it is arguable that this particular cross-section of the $L-$channel MD setup is degenerate and hence this proof has little value in practice. We therefore provide a more general proof where all the descriptions carry non-trivial information and all rates $R_1, \ldots, R_L$ are greater than 0.

We prove (ii) for $L = 3$. Note that once we prove that the CMS scheme achieves a strictly larger region for some $L = l > 2$, then it must be true for all $L \geq l$. Consider a 3-descriptions MD problem where descriptions $\{1, 2\}$ form the 2-description subset for which ZB achieves points outside EC region. We denote the VKG and CMS achievable regions by $\mathcal{RD}^3_{VKG}$ and $\mathcal{RD}^3_{CMS}$, respectively. We now consider a particular cross-section of these regions where we only constrain D_1, D_2, D_{12} and D_{13}. We remove the constraints on all other distortions, i.e., we allow D_3, D_{23} and D_{123} to be ∞. Equivalently, we can think of a 3 descriptions MD problem with a particular channel failure pattern, wherein only one of the following sets of descriptions can reach the decoder reliably: $(1, 2, \{1, 2\}, \{1, 3\})$ as shown in Fig. 6. We denote the set of all achievable points for this setup using VKG and CMS by $\tilde{\mathcal{RD}}^3_{VKG}$ and $\tilde{\mathcal{RD}}^3_{CMS}$, respectively.

Observe that with respect to the first 2 descriptions, we have a simple 2-descriptions problem, while descriptions $\{1, 3\}$ represent a successive refinement setup. We know from the results of Zhang and Berger that a common codeword sent among the first two descriptions leads to a strictly better RD trade-off for these two descriptions. However, the VKG coding scheme cannot use this common codeword unless it is also sent as part of the third description, which is redundant for the successive refinement framework involving descriptions $\{1, 3\}$, since description 3 is never received without description 1. On the other hand, CMS allows for a unique common codeword to be sent within every subset of the transmitted descriptions. Specifically, it can send a common codeword which is only shared by the first two descriptions, achieving the ZB advantage without adding redundancy to description 3. This argument is in fact sufficient to prove the claim. We rewrite it formally below.

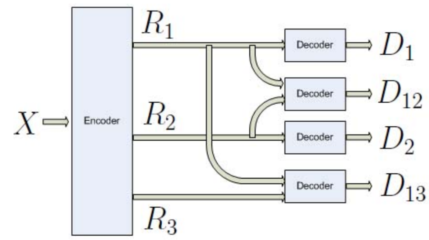

Fig. 6. (a) Equivalent model for the cross-section considered to prove that the CMS scheme achieves strict improvement over the VKG scheme for a binary symmetric source under Hamming distortion measure.

Consider a point $P^* = \{R_1, R_2, D_1, D_2, D_{12}\}$ in the ZB region that is not in the EC region. Let the rate of the common codeword that is sent as part of both the descriptions at P^* be denoted by R^*_{12}. Denote by $\bar{R}_{VKG}$ the following cross-section of $\tilde{\mathcal{RD}}^3_{VKG}$:

$$\bar{R}_{VKG} = \inf\Big\{R_3 : \{R_1, R_2, D_1, D_2, D_{12}\} = P^*, \\ (R_1, R_2, R_3, D_1, D_2, D_{12}, D_{13}) \in \tilde{\mathcal{RD}}^3_{VKG}\Big\}$$

and for the corresponding cross-section of $\tilde{\mathcal{RD}}^3_{CMS}$, the infimum R_3 rate by $\overline{R}_{CMS}(P^*)$. As P^* is achievable by ZB and not by EC, the constraint $\{R_1, R_2, D_1, D_2, D_{12}\} = P^*$ forces a common codeword to be sent in descriptions 1 and 2. To use this common codeword, VKG requires it to be sent as part of description 3 as well leading to decoder $\{1, 3\}$ receiving the same codeword twice. However, CMS sends this common codeword only as part of descriptions 1 and 2 leading to a smaller rate for the third description. Hence it follows that $\overline{R}_{CMS}$ is smaller than $\overline{R}_{VKG}$ by at least the rate of this common codeword which is equal to R^*_{12}. It can be verified from [3] that for a Binary symmetric source under Hamming distortion measure:

$$\begin{aligned}\overline{R}_{VKG} - \overline{R}_{CMS} &\geq 1 - H_b(0.30585) \\ &= 0.1117\text{bits}\end{aligned} \tag{22}$$

where $H_b(\cdot)$ denotes the binary entropy function. Hence, we have shown that the CMS region is strictly larger than VKG region whenever there exists a 2-descriptions subset for which ZB outperforms EC. Moreover, the gap between CMS and VKG regions is at least as large as the gap between the EC and the ZB regions for the corresponding 2-descriptions subset. □

B. General Sources and Distortion Measures

It is clear from Theorem 2 that for any given distribution for X and distortion measures, if there exists a 2-description subset such that $\mathcal{RD}_{EC} \subset \mathcal{RD}_{ZB}$, then CMS strictly outperforms VKG. In this section we show a more surprising result that the common layer codewords in CMS play a critical role in achieving a strictly larger region for a fairly general (and larger) class of source distributions and distortion measures. What makes it particularly interesting is the fact that under MSE, a Gaussian source ($X \sim \mathcal{N}(0, 1)$) belongs to this class, i.e., we will show that CMS achieves strictly larger RD region compared to VKG for the $L-$channel quadratic Gaussian MD problem $\forall L \geq 3$. This result is in striking contrast to the corresponding results for the 2-descriptions setting. It follows from Ozarow's results in [5] that for a Gaussian source under MSE, the EC coding scheme achieves the complete RD region, i.e., sending a common codeword among the two descriptions does not provide any improvement in the RD region ($\mathcal{RD}_{ZB}(\mathcal{N}(0,1)) = \mathcal{RD}_{EC}(\mathcal{N}(0,1))$). This result has lead to a natural belief that common codewords do not play a role in the $L-$channel quadratic Gaussian MD problem and all the explicit characterizations for the achievable regions in the past make no use of common random variables (see for example, [1], [9], [10], [14], [17]). Surprisingly, in the $L-$descriptions framework, the common codewords in CMS can be used constructively for better coordination between the descriptions leading to achievable points outside the VKG region.

We begin by defining a set of random variables, denoted by $\mathcal{Z}_{ZB}$, that plays a critical role in the subsequent theorems. Let us consider a two descriptions MD setup. For any given distortion measures, d_1, d_2, d_{12}, we say that $X \in \mathcal{Z}_{ZB}(d_1, d_2, d_{12})$, if there exists an operating point $(R_1, R_2, D_1, D_2, D_{12})$ that belongs to $\mathcal{RD}_{ZB}$, but *cannot* be achieved by an 'independent quantization' mechanism using the ZB coding scheme. A formal definition of $\mathcal{Z}_{ZB}$ is given below.

Definition 2 ($\mathcal{Z}_{ZB}$): Let $\epsilon \geq 0$. Let us denote $\mathcal{RD}_{ZB}^{IQ}(\epsilon)$ to be the rate-distortion region achievable by the ZB coding scheme when the closure in (4) if defined only over joint distributions for the auxiliary random variables satisfying the following conditions[4]:

$$\begin{aligned} I(U_1; U_2|X, V_{12}) &< \epsilon \\ E\left[d_{\mathcal{K}}(X, \psi_{\mathcal{K}}(U_{\mathcal{K}}))\right] &\leq D_{\mathcal{K}}, \quad \mathcal{K} \in \{1, 2, 12\} \\ I(U_{12}; X|U_1, U_2, V_{12}) &< \epsilon \end{aligned} \tag{23}$$

We say that $X \in \mathcal{Z}_{ZB}(d_1, d_2, d_{12})$, if:

$$\lim_{\epsilon \to 0} \mathcal{RD}_{ZB}^{IQ}(\epsilon) \subset \mathcal{RD}_{ZB}$$

i.e., if $\lim_{\epsilon \to 0} \mathcal{RD}_{ZB}^{IQ}(\epsilon)$ is strictly subsumed in $\mathcal{RD}_{ZB}$.

Remark 4: Note that if $\mathcal{RD}_{ZB}^{IQ}(0) \subset \mathcal{RD}_{ZB}$, then clearly $X \in \mathcal{Z}_{ZB}(d_1, d_2, d_{12})$, i.e., if there is a strict suboptimality in the ZB region when the closure in (4) is defined only over joint distributions for the auxiliary random variables satisfying:

$$\begin{aligned} U_1 &\leftrightarrow (X, V_{12}) \leftrightarrow U_2 \\ E\left[d_{\mathcal{K}}(X, \psi_{\mathcal{K}}(U_{\mathcal{K}}))\right] &\leq D_{\mathcal{K}}, \quad \mathcal{K} \in \{1, 2, 12\} \\ I(U_{12}; X|U_1, U_2, V_{12}) &= 0 \end{aligned} \tag{24}$$

then $X \in \mathcal{Z}_{ZB}(d_1, d_2, d_{12})$. Observe that the last constraint $I(U_{12}; X|U_1, U_2, V_{12}) = 0$ can be replaced with $U_{12} = f(U_1, U_2, V_{12})$, without any loss of optimality. This follows from the fact that for any given joint distribution $P(X, U_1, U_2, V_{12}, U_{12})$ that satisfies all the conditions in (24), it is possible to construct a conditional distribution as follows:

$$Q(U_{12}|U_1, U_2, V_{12}) = \begin{cases} 1 & \text{if } U_{12} = \arg\min_{u_{12}} E\left[d_{12}(X, u_{12})\middle| U_1, U_2, V_{12}\right] \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

The joint distribution $P(X, U_1, U_2, V_{12}) Q(U_{12}|U_1, U_2, V_{12})$ achieves the smallest distortion for D_{12}. Therefore, it follows that the constraint $I(U_{12}; X|U_1, U_2, V_{12}) = 0$ can be replaced with $U_{12} = f(U_1, U_2, V_{12})$, without any loss of optimality. This simplified definition of $\mathcal{Z}_{ZB}$ will be used in the proof of Theorem 4.

We will show in Theorem 3 that $\forall X \in \mathcal{Z}_{ZB}(d_1, d_2, d_{12})$, $\mathcal{RD}_{VKG} \subset \mathcal{RD}_{CMS}$. We will later provide more intuition on the underlying relation between $\mathcal{Z}_{ZB}(d_1, d_2, d_{12})$ and the reason for strict improvement of CMS over VKG. Hereafter, we will often drop the parenthesis and abbreviate $\mathcal{Z}_{ZB}(d_1, d_2, d_{12})$ by $\mathcal{Z}_{ZB}$, whenever the distortion measures are obvious.

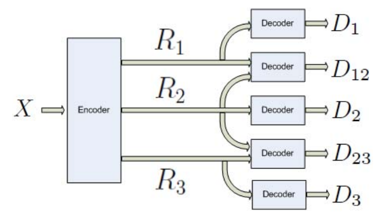

Fig. 7. The cross-section that we consider to prove that CMS achieves points outside the VKG region for a general class of source and distortion measures. CMS achieves the the complete RD region for this setup for several distortion regimes for the quadratic Gaussian MD problem.

Before stating the result we describe the particular cross-section of the RD region that we will use to prove strict improvement in Theorem 3. Consider a 3-descriptions MD setup for a source X wherein we impose constraints only on distortions $(D_1, D_2, D_3, D_{12}, D_{23})$ and set the limit on the rest of the distortions, (D_{13}, D_{123}) to ∞. This cross-section is schematically shown in Fig. 7. To illustrate the gains underlying CMS, let us further restrict to the setting where we impose $D_1 = D_3$ and $D_{12} = D_{23}$. We also assume that the distortion measures $d_1(\cdot, \cdot)$ and $d_{12}(\cdot, \cdot)$ are same as the distortion measures $d_3(\cdot, \cdot)$ and $d_{23}(\cdot, \cdot)$, respectively. The points in this cross-section, achievable by VKG and CMS, are denoted by $\overline{\mathcal{RD}}_{VKG}(X)$ and $\overline{\mathcal{RD}}_{CMS}(X)$, respectively. We note that the symmetric setting is considered *only* for ease of understanding the proof. The arguments can be easily extended to the asymmetric framework.

[4]The superscript of 'IQ' in $\mathcal{RD}_{ZB}^{IQ}(\epsilon)$ refers to 'independent quantization'.

This particular symmetric cross-section of the 3-descriptions MD problem is equivalent to the corresponding 2-descriptions problem, in the sense that, one could use any coding scheme to generate bit-streams for descriptions 1 and 2, respectively. Description 3 would then carry a replica (exact copy) of the bits sent in description 1. Due to the underlying symmetry in the problem setup, the distortion constraints at all the decoders are satisfied. Hence an achievable region based on the ZB coding scheme can be derived as follows. Let $(G_{12}, F_1, F_2, F_{12})$ be any random variables jointly distributed with X and taking values over arbitrary finite alphabets. Then the following RD-region is achievable for which there exist functions $(\psi_1, \psi_2, \psi_{12})$ such that $R_1 = R_3$, $D_1 = D_3$, $D_{12} = D_{23}$ and:

$$\begin{aligned} R_1 &\geq I(X; F_1, G_{12}) \\ R_2 &\geq I(X; F_2, G_{12}) \\ R_1 + R_2 &\geq 2I(X; G_{12}) + H(F_1|G_{12}) + H(F_2|G_{12}) \\ &\quad + H(F_{12}|F_1, F_2, G_{12}) \\ &\quad - H(F_1, F_2, F_{12}|X, G_{12}) \\ D_{\mathcal{K}} &\geq E\left[d_{\mathcal{K}}(X, \psi_{\mathcal{K}}(F_{\mathcal{K}}))\right], \quad \mathcal{K} \subseteq \{1, 2\}, \ \mathcal{K} \neq \emptyset \end{aligned} \tag{26}$$

The closure of achievable RD-tuples over all random variables $(G_{12}, F_1, F_2, F_{12})$ is denoted by $\overline{\mathcal{RD}}(X)$. In the following theorem, we will show that $\overline{\mathcal{RD}}(X) \subseteq \overline{\mathcal{RD}}_{CMS}(X)$. We also show that the VKG coding scheme *cannot* achieve the above RD region, i.e., $\overline{\mathcal{RD}}_{VKG}(X) \subset \overline{\mathcal{RD}}(X)$, if $X \in \mathcal{Z}_{ZB}(d_1, d_2, d_{12})$. Throughout proof of Theorem 3, we use the notations $\overline{\mathcal{R}}_{VKG}(X, D_1, D_2, D_{12})$, $\overline{\mathcal{R}}_{CMS}(X, D_1, D_2, D_{12})$ and $\overline{\mathcal{R}}(X, D_1, D_2, D_{12})$ to denote the rate-region cross-sections of $\overline{\mathcal{RD}}_{VKG}(X)$, $\overline{\mathcal{RD}}_{CMS}(X)$ and $\overline{\mathcal{RD}}(X)$ at distortions $(D_1, D_2, D_1, D_{12}, D_{12})$, respectively. We note that in Theorem 3, we focus only on the 3-descriptions setting. However, the results can be easily extended to the general $L-$descriptions scenario. Also note that $\overline{\mathcal{RD}}_{CMS}(X)$ could be strictly larger than $\overline{\mathcal{RD}}(X)$, in general.

Theorem 3:

(i) *For the setup shown in Fig. 7 the CMS scheme achieves* $\overline{\mathcal{RD}}(X)$*, i.e.,* $\overline{\mathcal{RD}}(X) \subseteq \overline{\mathcal{RD}}_{CMS}(X)$.

(ii) *If* $X \in \mathcal{Z}_{ZB}(d_1, d_2, d_{12})$*, then there exists points in* $\overline{\mathcal{RD}}(X)$ *that* ***cannot*** *be achieved by the VKG encoding scheme, i.e.,* $\overline{\mathcal{RD}}_{VKG} \subset \overline{\mathcal{RD}}(X)$,

Remark 5: It directly follows from (i) and (ii) that $\mathcal{RD}_{VKG} \subset \mathcal{RD}_{CMS}$ for the $L-$channel MD problem $\forall L \geq 3$, if $X \in \mathcal{Z}_{ZB}$.

Intuition: We first provide an intuitive argument to justify the claim. Due to the underlying symmetry in the setup the CMS scheme introduces common layer random variables $V_{123} = G_{12}$ and $V_{13} = F_1$. It then sends the codeword of V_{13} is both descriptions 1 and 3 (i.e., $U_1 = U_3 = V_{13}$). Hence it is sufficient for the encoder to generate enough codewords of $U_2 = F_2$ (conditioned on V_{123}) to maintain joint typicality with the codewords of V_{13}. However, the VKG scheme is forced to set the common layer random variable V_{13} to a constant. Thus, in this case, the encoder needs to generate enough number of codewords of U_2 (and U_{12}) so as to maintain joint typicality individually with the codewords of U_1 and U_3, which are now generated independently conditioned on V_{123}. It is possible to show that this entails some excess rate on U_2, unless (U_1, U_2, U_3) are pairwise independent conditioned on V_{123}. However, if $X \in \mathcal{Z}_{ZB}$, then enforcing independence of U_1 and U_2 conditioned on V_{123} leads to a strictly smaller rate-distortion region. Therefore, $\forall X \in \mathcal{Z}_{ZB}$, the VKG scheme leads to a strictly smaller rate-distortion region compared to the CMS scheme.

Proof: Part (i) of the theorem is straightforward to prove. We set,

$$\begin{aligned} V_{123} &= G_{12} \\ V_{13} &= F_1 \\ U_2 &= F_2 \\ U_{12} &= U_{23} = F_{12} \\ U_1 &= U_3 = V_{13} \end{aligned} \tag{27}$$

and the rest of the random variables in Theorem 1 to constants. It is easy to verify that the rate-distortion region given in (27) is achievable by the CMS scheme using the above joint distribution for the auxiliary random variables and by choosing the auxiliary rate $R''_{123} = I(X; V_{123})$.

Next, let $X \in \mathcal{Z}_{ZB}(d_1, d_2, d_{12})$ with respect to the given distortion measures, i.e., there exists a strict suboptimality in the ZB region when the closure of rates is defined only over joint densities for the auxiliary random variables satisfying (23). For ease of understanding the proof, we make a simplifying assumption that there exists at least one 'corner point' in the ZB region that is not achievable using joint densities satisfying (23). Specifically, let the distortions be (D_1, D_2, D_{12}), respectively, at the three decoders. We assume that:

$$\begin{aligned} &\inf_{R_2:\{R_1 = R_X(D_1)\}} R_{ZB}(D_1, D_2, D_{12}) \\ &\qquad < \inf_{R_2:\{R_1 = R_X(D_1)\}} R_{ZB}^{IQ}(D_1, D_2, D_{12}) \end{aligned} \tag{28}$$

i.e., if $R_1 = R_X(D_1)$, then the infimum over all R_2 required to achieve (D_1, D_2, D_{12}), using joint densities satisfying (23), is strictly larger than the infimum over all R_2 achievable using ZB. However, it is important to note that the proof can be easily extended to the general case of strict suboptimality occurring at any intermediate point. We will briefly discuss the extension towards the end of the proof.

Towards proving (ii), we consider one particular boundary point of (26) and show that this cannot be achieved by the VKG encoding scheme. Let $\epsilon > 0$ and D_1, D_2 and D_{12} be fixed. Define the following quantity:

$$\begin{aligned} &R^*_{VKG}(D_1, D_2, D_{12}, \epsilon) \\ &\quad = \inf\Big\{R_2 : R_1 < R_X(D_1) + \epsilon, R_3 < R_X(D_1) + \epsilon, \\ &\qquad\qquad (R_1, R_2, R_3) \in \overline{\mathcal{R}}_{VKG}(X, D_1, D_2, D_{12})\Big\} \end{aligned} \tag{29}$$

The corresponding quantity defined for the CMS region is given by:

$$R^*_{CMS}(D_1, D_2, D_{12}, \epsilon) = \inf\Big\{R_2 : R_1 < R_X(D_1) + \epsilon, R_3 < R_X(D_1) + \epsilon, (R_1, R_2, R_3) \in \overline{\mathcal{R}}_{CMS}(X, D_1, D_2, D_{12})\Big\} \quad (30)$$

We will show that:

$$\lim_{\epsilon\to 0} R^*_{CMS}(\epsilon) < \lim_{\epsilon\to 0} R^*_{VKG}(\epsilon)$$

Similarly, we use the notation $R^*(\epsilon)$ to denote the same quantity defined over $\overline{\mathcal{R}}(X, D_1, D_2, D_{12})$, i.e.,

$$R^*(D_1, D_2, D_{12}, \epsilon) = \inf R_2 : \Big\{R_1 < R_X(D_1) + \epsilon, R_3 < R_X(D_1) + \epsilon, (R_1, R_2, R_3) \in \overline{\mathcal{R}}(X, D_1, D_2, D_{12})\Big\} \quad (31)$$

From CMS region characterization in Theorem 1, $R^*_{CMS}(\epsilon)$ is given by the solution to the following optimization problem:

$$R^*_{CMS}(\epsilon) = \inf\Big\{I(X; V_{123}) + I(U_2; X, U_1|V_{123}) + I(U_{12}; X|V_{123}, U_1, U_2)\Big\} \quad (32)$$

where the infimum is over all joint distributions $P(V_{123}, U_1, U_2, U_{12}|X)$, and conditional distributions $P(V_{123}, U_1|X)$ for which there exists a function $\psi_1(\cdot)$ such that:

$$\begin{aligned} I(X; V_{123}, U_1) &< R_X(D_1) + \epsilon \\ E\left[d_1(X, \psi_1(U_1))\right] &\le D_1 \end{aligned} \quad (33)$$

i.e., (V_{123}, U_1) achieves a close-to-optimal reconstruction of X at D_1 and $P(U_{12}, U_2|X, U_1, V_{123})$ is any distribution for which there exists functions $\psi_2(\cdot)$ and $\psi_{12}(\cdot)$ satisfying the distortion constraints for D_2 and D_{12}, respectively.

We next specialize and restate $\overline{\mathcal{RD}}_{VKG}(X)$ for the considered cross-section. Let $(V_{123}, U_1, U_2, U_3, U_{12}, U_{23}, U_{13}, U_{123})$ be any random variables jointly distributed with X taking values on arbitrary alphabets. Then, the following rate-distortion tuples are achievable by the VKG scheme for which there exist functions $\psi_1(\cdot), \psi_2(\cdot), \psi_3(\cdot), \psi_{12}(\cdot), \psi_{23}(\cdot)$, such that:

$$\begin{aligned} R_i &\ge I(X; U_i, V_{123}) i \in \{1, 2, 3\} \\ R_1 + R_2 &\ge 2I(X; V_{123}) + I(U_1; U_2|V_{123}) + I(X; U_1, U_2, U_{12}|V_{123}) \\ R_2 + R_3 &\ge 2I(X; V_{123}) + I(U_2; U_3|V_{123}) + I(X; U_2, U_3, U_{23}|V_{123}) \\ R_1 + R_3 &\ge 2I(X; V_{123}) + I(U_1; U_3|V_{123}) + I(X; U_1, U_3, U_{13}|V_{123}) \\ R_1 + R_2 + R_3 &\ge 3I(X; V_{123}) + \sum_{i=1}^{3} H(U_i|V_{123}) + \sum_{\mathcal{K}\in\{12,23,13\}} H(U_{\mathcal{K}}|\{U\}_{\{k\in\mathcal{K}\}}, V_{123}) \\ &\quad + H(U_{123}|\{U\}_{\{1,2,3,12,13,23\}}, V_{123}) - H(\{U\}_{\{1,2,3,12,13,23,123\}}|X, V_{123}) \end{aligned} \quad (34)$$

$$\begin{aligned} E\left(d_1(X, \psi_1(U_1))\right) &\le D_1 \\ E\left(d_2(X, \psi_2(U_2))\right) &\le D_2 \\ E\left(d_1(X, \psi_3(U_3))\right) &\le D_1 \\ E\left(d_{12}(X, \psi_{12}(U_{12}))\right) &\le D_{12} \\ E\left(d_{12}(X, \psi_{23}(U_{23}))\right) &\le D_{12} \end{aligned} \quad (35)$$

The closure of the above rate-distortion region over all joint distributions leads to $\overline{\mathcal{RD}}_{VKG}$. Observe that there are no distortion constraints imposed on D_{13} and D_{123}. This allows us to simplify the region further, without any loss of optimality. First, the random variables U_{13} and U_{123} can be set to constants. This is because they do not appear in any of the distortion constraints and setting them to constants leads to a larger rate-region for the given distribution over all other random variables. This step simplifies the constraints on $R_1 + R_3$ and $R_1 + R_2 + R_3$ to:

$$\begin{aligned} R_1 + R_3 &\ge 2I(X; V_{123}) + H(U_1|V_{123}) + H(U_3|V_{123}) - H(U_1, U_3|X, V_{123}) \\ R_1 + R_2 + R_3 &\ge 3I(X; V_{123}) + \sum_{i=1}^{3} H(U_i|V_{123}) + \sum_{\mathcal{K}\in\{12,23\}} H(U_{\mathcal{K}}|\{U\}_{\{k\in\mathcal{K}\}}, V_{123}) \\ &\quad - H(\{U\}_{\{1,2,3,12,23\}}|X, V_{123}) \end{aligned} \quad (36)$$

As a next step of simplification, we restrict the closure of the region to be taken only over joint densities, $P(V_{123}, U_1, U_2, U_3, U_{12}, U_{23}|X)$, that satisfy the following constraint:

$$P(U_{12}, U_{23}|X, V_{123}, U_1, U_2, U_3) = P(U_{12}|X, V_{123}, U_1, U_2)P(U_{23}|X, V_{123}, U_2, U_3) \quad (37)$$

We note that this restriction does not lead to any loss in $\overline{\mathcal{RD}}_{VKG}(X)$ for this particular cross-section. This is because, for any given joint distribution $Q(X, V_{123}, U_1, U_2, U_3, U_{12}, U_{23})$, we can construct another joint distribution that satisfies (37) and leads to a larger rate-distortion region. To see this, consider the joint distribution constructed as:

$$Q(X, V_{123}, U_1, U_2, U_3) \times Q(U_{12}|X, V_{123}, U_1, U_2) \times Q(U_{23}|X, V_{123}, U_2, U_3) \quad (38)$$

which satisfies (37). Observe that this joint distribution satisfies the same distortion constraints as $Q(\cdot)$. Moreover, it leads to the same rate constraints as $Q(\cdot)$, except for the constraint on $R_1+R_2+R_3$. However, the constraint it imposes on $R_1 + R_2 + R_3$ is weaker than that imposed by $Q(\cdot)$ and hence the rate-distortion region is larger than that achievable by $Q(\cdot)$. Therefore, it is sufficient to consider only those joint distributions that satisfy (37) for $\overline{\mathcal{RD}}_{VKG}(X)$.

We next impose the constraints $R_1 < R_X(D_1) + \epsilon$ and $R_3 < R_X(D_1) + \epsilon$ in (36). This enforces the conditional density $P(V_{123}, U_1, U_3|X)$ to satisfy the following constraints:

$$\begin{aligned} I(X; V_{123}, U_1) &< R_X(D_1) + \epsilon \\ I(X; V_{123}, U_3) &< R_X(D_1) + \epsilon \\ E\left[d_1(X, \psi_1(V_{123}, U_1))\right] &\le D_1 \\ E\left[d_1(X, \psi_3(V_{123}, U_3))\right] &\le D_1 \\ I(U_1; U_3|X, V_{123}) &< \epsilon \end{aligned} \tag{39}$$

where the last condition is required to satisfy the constraint on $R_1 + R_3$ in (36). Therefore, using (36) and (37), $R^*_{VKG}(\epsilon)$ can be written as the solution to the following optimization problem:

$$\begin{aligned} R^*_{VKG}(\epsilon) = \inf\Big\{ & I(X; V_{123}) + I(U_2; U_1, U_3, X|V_{123}) \\ & + I(X; U_{12}|U_1, U_2, V_{123}) \\ & + I(X; U_{23}|U_2, U_3, V_{123})\Big\} \end{aligned} \tag{40}$$

where the infimum is over all conditional densities $P(V_{123}, U_1, U_2, U_3, U_{12}, U_{23}|X)$ satisfying (39) for which there exist functions $\psi_2(\cdot), \psi_{12}(\cdot), \psi_{23}(\cdot)$ satisfying the distortion constraints in (35).

Summarizing what we have so far:

$$\begin{aligned} \lim_{\epsilon\to 0} R^*_{VKG}(\epsilon) &\le R^*_{VKG}(\epsilon) \\ &= \inf\Big\{ I(X; V_{123}) + I(U_2; U_1, U_3, X|V_{123}) \\ &\quad + I(X; U_{12}|V_{123}, U_1, U_2) \\ &\quad + I(X; U_{23}|V_{123}, U_2, U_3)\Big\} \end{aligned} \tag{41}$$

$$\begin{aligned} \lim_{\epsilon\to 0} R^*_{CMS}(\epsilon) &\le R^*_{CMS}(\epsilon) \\ &= \inf\Big\{ I(X; V_{123}) + I(U_2; U_1, X|V_{123}) \\ &\quad + I(X; U_{12}|V_{123}, U_1, U_2)\Big\} \end{aligned} \tag{42}$$

where the infimum for $R^*_{VKG}(\epsilon)$ is subject to (39) and the infimum for $R^*_{CMS}(\epsilon)$ is subject to (33).

We first note that the optimization objective for $R^*_{VKG}(\epsilon)$ is always greater than or equal to the optimization objective for $R^*_{CMS}(\epsilon)$, for a given joint distribution over the random variables. This follows from the following standard information theoretic inequalities:

$$\begin{aligned} I(U_2; U_1, U_3, X|V_{123}) &\ge I(U_2; U_1, X|V_{123}) \\ I(X; U_{23}|V_{123}, U_2, U_3) &\ge 0 \end{aligned}$$

Moreover, the constraints for $R^*_{CMS}(\epsilon)$ are a subset of the constraints for $R^*_{VKG}(\epsilon)$. Therefore, for $\lim_{\epsilon\to 0} R^*_{VKG}(\epsilon)$ to be equal to $\lim_{\epsilon\to 0} R^*_{CMS}(\epsilon)$, there should exist a small enough $\epsilon > 0$ for every $\lambda > 0$, such that:

$$R^*_{VKG}(\epsilon) - R^*_{CMS}(\epsilon) < \lambda \tag{43}$$

We will next show that this cannot happen if $X \in \mathcal{Z}_{ZB}(d_1, d_2, d_{12})$. i.e., we will show that $R^*_{VKG}(\epsilon) - R^*_{CMS}(\epsilon)$ is lower bounded by a value that is strictly greater than 0.

We prove this claim by contradiction. We assume that it is possible to find a small enough $\epsilon > 0$ for every $\lambda > 0$ such that (43) holds and then arrive at a contradiction. Let $\epsilon_1 > 0$. Let $P(V_{123}, U_1, U_2, U_3, U_{12}, U_{23}|X)$ be any conditional distribution that satisfies the constraints in (39) and satisfies:

$$\begin{aligned} & I^P(X; V_{123}) + I^P(U_2; U_1, U_3, X|V_{123}) \\ &\quad + I^P(X; U_{12}|V_{123}, U_1, U_2) \\ &\quad + I^P(X; U_{23}|V_{123}, U_2, U_3) - R^*_{VKG}(\epsilon) < \epsilon_1 \end{aligned} \tag{44}$$

where the superscript P has been added to emphasize that the mutual information is with respect to the distribution P. The distribution P achieves the optimization objective in (41), within ϵ_1 of $R^*_{VKG}(\epsilon)$. For (43) to hold, the following condition must be satisfied by all conditional distributions $\tilde{P}(V_{123}, U_1, U_2, U_{12}|X)$ that satisfy the constraints in (33):

$$\begin{aligned} & I^P(X; V_{123}) - I^{\tilde{P}}(X; V_{123}) \\ &\quad + I^P(U_2; U_1, U_3, X|V_{123}) - I^{\tilde{P}}(U_2; U_1, X|V_{123}) \\ &\quad + I^P(X; U_{12}|V_{123}, U_1, U_2) - I^{\tilde{P}}(X; U_{12}|V_{123}, U_1, U_2) \\ &\quad + I^P(X; U_{23}|V_{123}, U_2, U_3) < \epsilon_1 + \lambda \end{aligned} \tag{45}$$

We will next show that this leads to a contradiction if $X \in \mathcal{Z}_{ZB}(d_1, d_2, d_{12})$. Towards proving that, we set $\tilde{P}(V_{123}, U_1, U_2, U_{12}|X) = P(V_{123}, U_1, U_2, U_{12}|X)$. Observe that P is a valid candidate for $\tilde{P}$ as the constraints in (33) are a subset of (39). With $\tilde{P} = P$, for (45) to hold, we need:

$$\begin{aligned} I^P(U_2; U_3|V_{123}, U_1, X) &< \epsilon_1 + \lambda \\ I^P(X; U_{23}|V_{123}, U_2, U_3) &< \epsilon_1 + \lambda \end{aligned} \tag{46}$$

First, observe that the constraint $I^P(U_2; U_3|V_{123}, U_1, X) < \epsilon_1 + \lambda$ implies that $H^P(U_3|V_{123}, U_1, X) < H^P(U_3|V_{123}, U_1, U_2, X) + \epsilon_1 + \lambda$. However, as P satisfies (39), we have $H^P(U_3|V_{123}, X) < H^P(U_3|V_{123}, X, U_1) + \epsilon$. On substituting, we get:

$$H^P(U_3|V_{123}, X) < H^P(U_3|V_{123}, U_1, U_2, X) + \epsilon_1 + \lambda + \epsilon \tag{47}$$

i.e.,

$$I^P(U_3; U_2, U_1|V_{123}, X) < \epsilon_1 + \lambda + \epsilon \tag{48}$$

which implies that:

$$I^P(U_3; U_2|V_{123}, X) < \epsilon_1 + \lambda + \epsilon \tag{49}$$

Hence, from (46) and (49), for (43) to hold, we need the following constraints to be satisfied for the joint distribution P that achieves close to optimality in (41):

$$\begin{aligned} I^P(U_3; U_2|V_{123}, X) &< \epsilon_1 + \lambda + \epsilon \\ I^P(X; U_{23}|V_{123}, U_2, U_3) &< \epsilon_1 + \lambda + \epsilon \\ E^P\left[d_{\mathcal{K}}(X, \psi_{\mathcal{K}}(U_{\mathcal{K}}))\right] &\le D_{\mathcal{K}}, \quad \mathcal{K} \in \{2, 3, 23\} \end{aligned} \tag{50}$$

Under the limits of $\epsilon \to 0$, $\epsilon_1 \to 0$ and $\lambda \to 0$, if $X \in \mathcal{Z}_{ZB}(d_1, d_2, d_{12})$, the above constraints imply that there is a strict sub-optimality in the ZB region. From our assumption in (28), it immediately follows that:

$$\lim_{\epsilon\to 0} R^*_{VKG}(\epsilon) > \lim_{\epsilon\to 0} R^*(\epsilon) \tag{51}$$

where $R^*(\epsilon)$ is defined in (31). However, from (i) of the theorem, it follows that:

$$\lim_{\epsilon \to 0} R^*(\epsilon) \geq \lim_{\epsilon \to 0} R^*_{CMS}(\epsilon) \quad (52)$$

This leads to a contradiction as it implies that (43) is not satisfied. It therefore follows that if $X \in \mathcal{Z}_{ZB}(d_1, d_2, d_{12})$, CMS achieves a strictly larger RD region compared to VKG, proving the theorem.

We note that if the strict sub-optimality in (28) exists at any other boundary point of the ZB region, the above proof holds by changing the definitions of $R^*_{VKG}(\epsilon)$, $R^*_{CMS}(\epsilon)$ and $R^*(\epsilon)$ accordingly. We only consider this particular corner point in this proof for ease of understanding. □

Discussion: A direct consequence of the above theorem is that, if $X \in \mathcal{Z}_{ZB}$, then the common layer codewords of CMS are needed to achieve strict improvement in the region, i.e., if $X \in \mathcal{Z}_{ZB}$, $\mathcal{RD}_{VKG}\Big|_{V_{\mathcal{L}}=\Phi} \subseteq \mathcal{RD}_{VKG} \subset \mathcal{RD}_{CMS}$, where $\mathcal{RD}\Big|_{V_{\mathcal{L}}=\Phi}$ denotes the VKG region when the common layer random variable (denoted by $V_{\mathcal{L}}$) is set to a constant.[5] In fact, it is possible to show that, whenever $X \in \mathcal{Z}_{EC}$, $\mathcal{RD}_{VKG}\Big|_{V_{\mathcal{L}}=\Phi} \subset \mathcal{RD}_{CMS}$, where $\mathcal{Z}_{EC}$ is similarly defined as the set of all random variables X for which there exists an operating point, with respect to the given distortion measures, that *cannot* be achieved by an 'independent quantization' mechanism using the EC coding scheme, i.e., if there exists an operating point in the EC region that *cannot* be achieved by the EC coding scheme using a joint density for the auxiliary random variables satisfying:

$$\begin{aligned} P(U_1, U_2|X) &= P(U_1|X)P(U_2|X) \\ E\left[d_{\mathcal{K}}(X, \psi_{\mathcal{K}}(U_{\mathcal{K}}))\right] &\leq D_{\mathcal{K}} \mathcal{K} \in \{1, 2, 12\} \\ U_{12} &= f(U_1, U_2) \end{aligned} \quad (53)$$

where f is any deterministic function.[6] Note that the set $\mathcal{Z}_{ZB}$ is a subset of $\mathcal{Z}_{EC}$. It is possible to construct distributions and distortion measures such that $\mathcal{Z}_{ZB}$ is a strict subset of $\mathcal{Z}_{EC}$, but as the construction does not provide any useful insights into the MD problem, we choose to omit the details. Also observe that if $X \notin \mathcal{Z}_{EC}$, the concatenation of two independent optimal quantizers is optimal in achieving a joint reconstruction. While this condition could be satisfied for specific values of D_1, D_2 and D_{12}, it is seldom achieved *for all* values of (D_1, D_2, D_{12}). Though such sources are of some theoretical interest, the multiple descriptions encoding for such sources is degenerate. Hence with some trivial exceptions, it can be asserted that the common layer codewords in CMS can be used to achieve a strictly larger region (compared to not using any common codewords) for all sources and distortion measures, $\forall L \geq 3$.

We note in passing that, for the two descriptions setting, Zhang and Berger studied an important cross-section of the problem in [31] called the "no excess marginal rate" setting, where $R_1 = R_X(D_1)$ and $R_2 = R_X(D_2)$. They derived upper and lower bounds on the achievable D_{12} and showed that the gap is negligible for a binary source under Hamming distortion measure. The constraints $R_1 = R_3 = R_X(D_1)$ imposed in the proof of Theorem 3 resemble the constraints imposed in [31] and hence the results in [31] may seem relevant to the setting considered in this paper. However, it is important to note that in the proof of Theorem 3, R_2 can be greater than $R_X(D_2)$. As the interaction is only between descriptions $\{1, 2\}$ and $\{2, 3\}$, the problem considered here is not directly related to the 'no excess marginal rate' case and requires new tools to prove the results.

At a first glance, it is tempting to conclude from Ozarow's results in [5] that under MSE a Gaussian random variable belongs to both $\mathcal{Z}_{EC}$ and $\mathcal{Z}_{ZB}$. However, the formal proof is non-trivial. We will formally show in the next section that this is indeed the case and an independent quantization scheme (with or without a common codeword) leads to strict suboptimality for the 2-descriptions quadratic Gaussian MD problem.

V. Gaussian MSE Setting

In this section we present a series of new results for the $L-$descriptions quadratic Gaussian MD problem. We will first show that CMS achieves a strictly larger RD region, by proving that under MSE, a Gaussian source belongs to $\mathcal{Z}_{ZB}$. We then use similar encoding principles to derive the complete rate region in several asymmetric distortion regimes. Throughout this section, we will assume that $X \sim \mathcal{N}(0, 1)$ and the distortion at all the decoders is the squared error, i.e., $d(x_1, x_2) = (x_1 - x_2)^2$.

Before stating the results formally, we review Ozarow's result for the 2-descriptions MD setting. Ozarow showed that the complete region for the 2-descriptions Gaussian MD problem can be achieved using a 'correlated quantization' scheme which imposes the following joint distribution for (U_1, U_2, U_{12}) in the EC scheme:

$$\begin{aligned} U_1 &= X + W_1 \\ U_2 &= X + W_2 \end{aligned} \quad (54)$$

$U_{12} = E(X|U_1, U_2)$, where W_1 and W_2 are zero mean Gaussian random variables independent of X with covariance matrix $K_{W_1W_2}$, and the functions $\psi_{\mathcal{K}}(U_{\mathcal{K}})$ are given by the respective MSE optimal estimators, e.g., $\psi_1(U_1) = E[X|U_1]$. The covariance matrix $K_{W_1W_2}$ is set to satisfy all the distortion constraints. Specifically, the optimal $K_{W_1W_2}$ is given by:

$$K_{W_1W_2} = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \quad (55)$$

where $\sigma_i^2 = \frac{D_i}{1-D_i} i \in \{1, 2\}$ and the optimal ρ_{12}, denoted by ρ_{12}^*, is given by (see [32]):

$$\rho_{12}^* = \begin{cases} -\frac{\sqrt{\pi D_{12}^2 + \gamma} - \sqrt{\pi D_{12}^2}}{(1-D_{12})\sqrt{D_1 D_2}} & D_{12} \leq D_{12}^{max} \\ 0 & D_{12} \geq D_{12}^{max} \end{cases}$$

$$\gamma = (1 - D_{12})\Big[(D_1 - D_{12})(D_2 - D_{12}) + D_{12}D_1D_2 - D_{12}^2\Big]$$

[5] Note that setting $V_{\mathcal{L}}$ to a constant in VKG is equivalent to setting all the common layer random variables to constants in CMS.

[6] A formal definition of $\mathcal{Z}_{EC}$ would be in lines of the definition of $\mathcal{Z}_{ZB}$ in Definition 2. We state the simpler version here for brevity.

$$D_{12}^{max} = D_1D_2/(D_1 + D_2 - D_1D_2)$$
$$\pi = (1 - D_1)(1 - D_2) \quad (56)$$

We denote the complete Gaussian-MSE L-descriptions region by $\mathcal{RD}_G^L$. The characterization of $\mathcal{RD}_G^2$ is given in [2] (see also [32]) and we omit restating it explicitly here for brevity.

A. Strict Improvement for the Quadratic Gaussian Case

Equipped with these results, we next show that CMS achieves points outside the VKG region, by proving that a Gaussian source under MSE belongs to $\mathcal{Z}_{ZB}$.

Theorem 4:

(i) *CMS achieves the* **complete** *RD region for the symmetric 3-descriptions quadratic Gaussian setup shown in Fig. 7.*

$$\begin{aligned} R_1 &\geq \frac{1}{2}\log\frac{1}{D_1} \\ R_2 &\geq \frac{1}{2}\log\frac{1}{D_2} \\ R_1 + R_2 &\geq \frac{1}{2}\log\frac{1}{D_1D_2} + \delta \\ R_1 &= R_3 \\ D_{12} &= D_{23} \quad D_3 = D_1 \end{aligned} \quad (57)$$

where $\delta = \delta(D_{12}, D_1, D_2)$ is defined by:

$$\delta = \frac{1}{2}\log\left(\frac{1}{1-(\rho_{12}^*)^2}\right) \quad (58)$$

where ρ_{12}^* is defined in (56).

(ii) *The VKG encoding scheme cannot achieve all the points in the region, i.e., $\overline{\mathcal{RD}}_{VKG} \subset \overline{\mathcal{RD}}_{CMS}$.*

Remark 6: It follows from (i) and (ii) that $\mathcal{RD}_{VKG} \subset \mathcal{RD}_{CMS}$ for the $L-$channel quadratic Gaussian MD problem $\forall L > 2$.

Proof: To prove achievability using CMS, we set,

$$\begin{aligned} V_{13} &= X + W_1 \\ U_2 &= X + W_2 \\ U_1 &= U_3 = V_{13} \\ U_{12} &= U_{23} = E[X|V_{12}, U_2] \end{aligned} \quad (59)$$

where W_1 and W_2 are zero mean Gaussian random variables independent of X with covariance matrix $K_{W_1W_2}$ and the functions $\psi_{\mathcal{K}}(U_{\mathcal{K}})$ are given by the respective MSE optimal estimators, e.g., $\psi_1(U_1) = E[X|U_1]$. We set all the remaining auxiliary random variables to constants. It follows directly that the rate-distortion region given in (57) is achievable by the CMS scheme. Following the footsteps of Ozarow in [5], it is straightforward to show that the above region is also complete for the symmetric setup considered.

Note that Ozarow's results suggest that if $D_{12} \leq D_{12}^{max}$, then an 'independent quantization' scheme does not achieve the smallest D_{12}. It might be tempting to conclude that (ii) follows directly from this observation and Theorem 3. However, a closer inspection reveals that the above argument holds only after we prove the optimality of Gaussian codebooks under 'independent quantization' mechanism. We relegate the proof of (ii) to Appendix B as the underlying principles are quite orthogonal to the rest of the paper. □

Remark 7: Note that, as $\mathcal{Z}_{ZB} \subseteq \mathcal{Z}_{EC}$, a Gaussian source under MSE belongs to $\mathcal{Z}_{EC}$. Hence, the 'correlated quantization' scheme (an extreme special case of VKG) which has been proven to be complete for several cross-sections of the $L-$descriptions quadratic Gaussian MD problem [10], is strictly suboptimal in general.

B. Points on the Boundary - 3-Descriptions Setting

In this section we show that CMS achieves the complete RD region for several cross-sections of the general quadratic Gaussian $L-$channel MD problem. We again begin with the 3-descriptions case and then extend the results to the L channel framework. Recall the setup shown in Fig. 6, i.e, a cross-section of the general 3-descriptions rate-distortion region wherein we impose constraints only on distortions $(D_1, D_2, D_3, D_{12}, D_{23})$ and set the rest of the distortions, (D_{13}, D_{123}) to 1. Here we consider the general asymmetric case, i.e. $D_1 \neq D_3$ and $D_{12} \neq D_{23}$ and show that the CMS scheme achieves the complete rate region in several distortion regimes.

In the following theorem, without loss of generality we assume that $D_1 \leq D_3$. If $D_3 \leq D_1$, then the theorem holds by interchanging '1' and '3' everywhere. Let D_{12} be any distortion such that $D_{12} \leq \min\{D_1, D_2\}$. We define $D_{23}^* = D_{23}^*(D_1, D_2, D_3, D_{12})$ as:

$$D_{23}^* = \frac{\sigma_2^2\sigma_3^2\left(1-\rho^2\right)}{\sigma_2^2\sigma_3^2\left(1-\rho^2\right) + \sigma_2^2 + \sigma_3^2 - 2\sigma_2\sigma_3\rho} \quad (60)$$

where $\sigma_i^2 = \frac{D_i}{1-D_i}$ $i \in \{2, 3\}$ and

$$\rho = \rho_{12}^*\frac{\sigma_1}{\sigma_3} \quad (61)$$

where ρ_{12}^* is defined in (56). In the following theorem, we will show that CMS achieves the complete rate-region if $D_{23} = D_{23}^*$.

Theorem 5: For the setup shown in Fig. 6, let $D_1 \leq D_3$. Then,

(i) *CMS achieves the complete rate-region if:*

$$D_{23} = D_{23}^*(D_1, D_2, D_3, D_{12}) \quad (62)$$

where D_{23}^* is defined in (60). The rate region is given by:

$$\begin{aligned} R_i &\geq \frac{1}{2}\log\frac{1}{D_i} i \in \{1, 2, 3\} \\ R_1 + R_2 &\geq \frac{1}{2}\log\frac{1}{D_1D_2} + \delta(D_1, D_2, D_{12}) \\ R_2 + R_3 &\geq \frac{1}{2}\log\frac{1}{D_2D_3} + \delta(D_2, D_3, D_{23}) \end{aligned} \quad (63)$$

where $\delta(\cdot)$ is defined in (58).

(ii) *Moreover, CMS achieves the minimum sum-rate if one of the following hold:*

(a) *For a fixed D_{12}, $D_{23} \geq D_{23}^*(D_1, D_2, D_3, D_{12})$*

(b) For a fixed D_{23}, $D_{12} \in \{D_{12} : \delta(D_2, D_3, D_{23}) \geq \delta(D_1, D_2, D_{12})\}$

Remark 8: We note that the above rate region *cannot* be achieved by VKG. We omit the details of the proof here as it can be proved in same lines as the proof of Theorem 4.

Remark 9: An achievable rate-distortion region can be derived for general distortions using the encoding principles we derive as part of this proof. However, it is hard to prove outer bounds if the conditions in (62) are not satisfied and hence we omit stating the results explicitly here.

Remark 10: Both CMS and VKG achieve the complete rate region when $D_{12} \geq D_{12}^{max}$ and $D_{23} \geq D_{23}^{max}$, where D_{12}^{max} and D_{23}^{max} are defined in (56). It can be easily verified that in this case an independent quantization scheme is optimal and the complete achievable rate-region is given by $R_i \geq \frac{1}{2}\log\frac{1}{D_i}$, $i \in \{1, 2, 3\}$.

Remark 11: It follows from the above theorem that CMS achieves the minimum sum-rate whenever $D_{12} = D_{23}$ for any D_1, D_3.

Proof: We begin with the proof of (i). The proof of (ii) then follows almost directly from (i). First we show the converse, which is quite obvious. Conditions on R_i follow from the converse to the source coding theorem. Conditions on $R_1 + R_2$ and $R_2 + R_3$ follow from Ozarow's results in [5], to achieve (D_1, D_2, D_{12}) using descriptions $\{1, 2\}$ and to achieve (D_2, D_3, D_{23}) using descriptions $\{2, 3\}$ at the respective decoders.

We next prove that CMS achieves the rate region in (63) if (62) holds. We first give an intuitive argument to explain the encoding scheme. Description 3 carries an RD-optimal quantized version of X (which achieves distortion D_3). Description 1 carries all the bits embedded in description 3 along with 'refinement bits' which assist in achieving distortion $D_1 \leq D_3$. This entails no loss in optimality as a Gaussian source is successively refinable under MSE [33]. Description 2 then carries a quantized version of the source which is correlated with the information in descriptions 1 and 3. We will show that if $D_{23} = D_{23}^*(D_1, D_2, D_3, D_{12})$, then the correlations can be set such that description 2 is optimal with respect to both descriptions 1 and 3.

Formally, to achieve the rate region in (63), we set the auxiliary random variables in the CMS coding scheme as follows:

$$\begin{aligned} V_{13} &= X + W_1 + W_3 \\ U_3 &= V_{13} \\ U_1 &= X + W_1 \\ U_2 &= X + W_2 \\ U_{12} &= \Phi \quad U_{23} = \Phi \end{aligned} \tag{64}$$

and the functions $\psi(\cdot)$ as the respective MSE optimal estimators, where W_1, W_2, W_3 are zero mean Gaussian random variables independent of X with a covariance matrix:

$$K_{W_1W_2W_3} = \begin{bmatrix} \tilde{\sigma}_1^2 & \rho_{12}\tilde{\sigma}_1\tilde{\sigma}_2 & 0 \\ \rho_{12}\tilde{\sigma}_1\tilde{\sigma}_2 & \tilde{\sigma}_2^2 & 0 \\ 0 & 0 & \tilde{\sigma}_3^2 \end{bmatrix} \tag{65}$$

where $\tilde{\sigma}_1^2 = \sigma_1^2 = \frac{D_1}{1-D_1}$, $\tilde{\sigma}_2^2 = \sigma_2^2 = \frac{D_2}{1-D_2}$, $\tilde{\sigma}_3^2 = \sigma_3^2 - \sigma_1^2 = \frac{D_3}{1-D_3} - \frac{D_1}{1-D_1}$. The correlation coefficient ρ_{12} is set to achieve distortion D_{12}, i.e. $\rho_{12} = \rho_{12}^*$ defined in (56). Let us denote by $W_{13} = W_1 + W_3$. Observe that the encoding for descriptions 2 and 3 resembles Ozarow's correlated quantization scheme with $U_2 = X + W_2$ and $U_3 = X + W_{13}$. Let us denote the correlation coefficient between W_2 and W_{13} be ρ. We have the following equation relating ρ_{12} and ρ (which is equivalent to (61)):

$$\rho_{12}^*\tilde{\sigma}_1 = \rho\sqrt{\tilde{\sigma}_1^2 + \tilde{\sigma}_3^2} \tag{66}$$

Note that the above relation is derived using the independence of W_3 and (W_1, W_2), which follows from our choice of $K_{W_1W_2W_3}$. Hence the minimum distortion D_{23} achievable using the above choice for the joint density of the auxiliary random variables is given by:

$$\begin{aligned} D_{23} &= \text{Var}(X|U_2, U_3, V_{13}) \\ &= \text{Var}(X|U_2, V_{13}) \\ &= D_{23}^* \end{aligned} \tag{67}$$

We next derive the rates required by this choice of $K_{W_1W_2W_3}$. Application of Theorem 1 using the above joint density leads to the following achievable rate region for any given distortions $D_1, D_2, D_3, D_{12}, D_{23}$:

$$\begin{aligned} R''_{13} &\geq \frac{1}{2}\log\frac{1}{D_3} \\ R'_2 &\geq \frac{1}{2}\log\frac{1}{D_2} \\ R'_1 + R''_{13} &\geq \frac{1}{2}\log\frac{1}{D_1} \\ R'_2 + R''_{13} &\geq H(V_{13}) + H(U_2) - H(V_{13}, U_2|X) \\ &= H(U_3) + H(U_2) - H(U_3, U_2|X) \\ &= \frac{1}{2}\log\frac{1}{D_3D_2} + \frac{1}{2}\log\left(\frac{1}{1-\rho^2}\right) \\ &= \frac{1}{2}\log\frac{1}{D_3D_2} + \delta(D_2, D_3, D_{23}^*) \\ R'_1 + R'_2 + R''_{13} &\geq H(V_{13}) + H(U_1|V_{13}) + H(U_2) \\ &\quad - H(U_1, V_{13}, U_2|X) \\ &= I(X; U_1, V_{13}) + I(U_2; X, U_1, V_{13}) \\ &=^{(a)} I(X; U_1) + I(X; U_2) \\ &\quad + I(U_2; U_1, V_{13}|X) \\ &= I(X; U_1) + I(X; U_2) \\ &\quad + I(U_2; U_1, U_3|X) \\ &=^{(b)} I(X; U_1) + I(X; U_2) \\ &\quad + I(W_2; W_1, W_1 + W_3) \\ &= I(X; U_1) + I(X; U_2) + I(W_2; W_1) \\ &\quad + I(W_2; W_3|W_1) \\ &=^{(c)} I(X; U_1) + I(X; U_2) + I(W_2; W_1) \\ &= \frac{1}{2}\log\frac{1}{D_1D_2} + \frac{1}{2}\log\left(\frac{1}{1-(\rho_{12}^*)^2}\right) \\ &= \frac{1}{2}\log\frac{1}{D_1D_2} + \delta(D_1, D_2, D_{12}) \end{aligned}$$

$$\begin{aligned} R_1 &= R''_{13} + R'_1 \\ R_2 &= R'_2 \\ R_3 &= R''_{13} \end{aligned} \tag{68}$$

where (a) follows from the Markov chain $X \leftrightarrow U_1 \leftrightarrow V_{13}$, (b) from the independence of X and (W_1, W_2, W_3) and (c) from the independence of W_3 and (W_1, W_2).

At a first glance, it might be tempting to conclude that the region for the tuple (R_1, R_2, R_3) in (68) is equivalent to the region given by (63). This is not the case in general as the equations in (68) have an implicit constraint on the auxiliary rates $R''_{13}, R'_1, R'_2 \geq 0$. However, we will show that if $D_3 \geq D_1$, then the two regions are indeed equivalent. We denote the rate region given in (63) by $\mathcal{R}$ and the region in (68) by $\mathcal{R}^*$. Clearly, $\mathcal{R}^* \subseteq \mathcal{R}$, as any (R_1, R_2, R_3) that satisfies (68) also satisfies (63). We need to show that $\mathcal{R}^* \supseteq \mathcal{R}$. Towards proving this claim, note that both $\mathcal{R}$ and $\mathcal{R}^*$ are convex regions bounded by hyper-planes. Hence, it is sufficient for us to show that all the corner points of $\mathcal{R}$ lie in $\mathcal{R}^*$. Clearly, $\mathcal{R}$ has 6 corner points denoted by P_{ijk} $i, j, k \in \{1, 2, 3\}$ defined as:

$$\begin{aligned} P_{ijk} &= \{r_i, r_j, r_k\} \\ r_i &= \min R_i \\ r_j &= \min_{R_i = r_i} R_j \\ r_k &= \min_{R_i = r_i, R_j = r_j} R_k \end{aligned} \tag{69}$$

To prove $\mathcal{R}^* \supseteq \mathcal{R}$, we need to prove that every corner point $(r_1, r_2, r_3) \in \mathcal{R}$ is achieved by some non-negative $(R''_{13}, R'_1, R'_2, R_1, R_2, R_3) \in \mathcal{R}^*$ such that $R_i = r_i, i \in \{1, 2, 3\}$. We set $R''_{13} = R_3 = r_3$ and $R'_2 = R_2 = r_2$ and show that we can always find $R'_1 \geq 0$ satisfying (68) such that $R_1 = R'_1 + R'_{13} = r_1$. Let us first consider the points $P_{213} = P_{231}$ given by:

$$\begin{aligned} r_1 &= \frac{1}{2}\log\frac{1}{D_1} + \delta(D_1, D_2, D_{12}) \\ r_2 &= \frac{1}{2}\log\frac{1}{D_2} \\ r_3 &= \frac{1}{2}\log\frac{1}{D_3} + \delta(D_2, D_3, D_{23}) \end{aligned} \tag{70}$$

This can be achieved by using the following auxiliary rates, $R'_2 = r_2$, $R''_{13} = r_3$ and

$$\begin{aligned} R'_1 &= \frac{1}{2}\log\frac{D_3}{D_1} + \delta(D_1, D_2, D_{12}) \\ &\quad - \delta(D_2, D_3, D_{23}) \\ &= \frac{1}{2}\log\frac{(1-D_1)D_3 - (\rho^*_{12})^2 D_1(1-D_3)}{(1-D_1)D_1(1-(\rho^*_{12})^2)} \end{aligned} \tag{71}$$

It is easy to verify that $R'_1 \geq 0$ if $D_3 \geq D_1$. Hence $P_{213} = P_{231} \in \mathcal{R}^*$. Let us next consider the points $P_{132} = P_{312}$ given by:

$$\begin{aligned} r_1 &= \frac{1}{2}\log\frac{1}{D_1} \\ r_2 &= \frac{1}{2}\log\frac{1}{D_2} \\ &\quad + \max\{\delta(D_1, D_2, D_{12}), \delta(D_2, D_3, D_{23})\} \\ r_3 &= \frac{1}{2}\log\frac{1}{D_3} \end{aligned} \tag{72}$$

Again it is straightforward to show that $(R''_{13}, R'_1, R'_2) = (r_3, r_1 - r_3, r_2)$ belongs to $\mathcal{R}^*$. Finally, we consider the remaining two points P_{123} and P_{321}. P_{123} is given by:

$$\begin{aligned} r_1 &= \frac{1}{2}\log\frac{1}{D_1} \\ r_2 &= \frac{1}{2}\log\frac{1}{D_2} + \delta(D_1, D_2, D_{12}) \\ r_3 &= \frac{1}{2}\log\frac{1}{D_3} \\ &\quad + (\delta(D_2, D_3, D_{23}) - \delta(D_1, D_2, D_{12}))^+ \end{aligned} \tag{73}$$

where $x^+ = \max\{x, 0\}$. Consider the following auxiliary rates: $R''_{13} = r_3$, $R'_2 = r_2$ and $R'_1 = \frac{1}{2}\log\frac{D_3}{D_1}$. Clearly the first three constraints in (68) are satisfied by these auxiliary rates. The following inequalities prove that the last two constraints are also satisfied by these rates and hence $P_{123} \in \mathcal{R}^*$.

$$\begin{aligned} R'_2 + R''_{13} &= r_2 + r_3 \\ &= \frac{1}{2}\log\frac{1}{D_2 D_3} + \delta(D_1, D_2, D_{12}) \\ &\quad + (\delta(D_2, D_3, D_{23}) - \delta(D_1, D_2, D_{12}))^+ \\ &\geq \frac{1}{2}\log\frac{1}{D_2 D_3} + \delta(D_2, D_3, D_{23}) \\ R'_2 + R'_1 + R''_{13} &= \frac{1}{2}\log\frac{1}{D_1 D_2} + \delta(D_1, D_2, D_{12}) \\ &\quad + (\delta(D_2, D_3, D_{23}) - \delta(D_1, D_2, D_{12}))^+ \\ &\geq \frac{1}{2}\log\frac{1}{D_1 D_2} + \delta(D_1, D_2, D_{12}) \end{aligned} \tag{74}$$

Next consider P_{321}:

$$\begin{aligned} r_1 &= \frac{1}{2}\log\frac{1}{D_1} \\ &\quad + (\delta(D_1, D_2, D_{12}) - \delta(D_2, D_3, D_{23}))^+ \\ r_2 &= \frac{1}{2}\log\frac{1}{D_2} + \delta(D_2, D_3, D_{23}) \\ r_3 &= \frac{1}{2}\log\frac{1}{D_3} \end{aligned} \tag{75}$$

Using same arguments as before, it can be shown that $P_{321} \in \mathcal{R}^*$ by using the following auxiliary rates: $R''_{13} = r_3$, $R'_2 = r_2$ and $R'_1 = \frac{1}{2}\log\frac{D_3}{D_1} + (\delta(D_1, D_2, D_{12}) - \delta(D_2, D_3, D_{23}))^+$. Therefore, it follows that $\mathcal{R} = \mathcal{R}^*$ and hence CMS achieves the complete rate region, proving (i).

We next prove (ii)(a). It follows from (i) that the following rate point is achievable $\forall D_{23} \geq D^*_{23}$:

$$\{R_1, R_2, R_3\} = \left\{\frac{1}{2}\log\frac{1}{D_1}, \frac{1}{2}\log\frac{1}{D_2} + \delta(D_1, D_2, D_{12}), \frac{1}{2}\log\frac{1}{D_3}\right\} \tag{76}$$

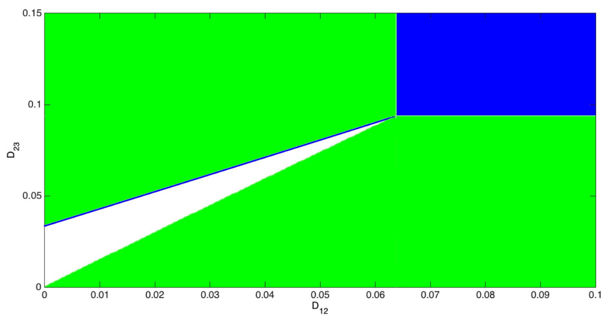

Fig. 8. Example: This figure denotes the regime of distortions wherein the CMS scheme achieves the complete rate region and the minimum sum rate. Here $D_1 = 0.1$, $D_2 = 0.15$ and $D_3 = 0.2$. The blue points correspond to the region of distortions wherein the CMS scheme achieves the complete rate-region and the green points represent the region where the CMS scheme achieves the minimum sum rate.

Also observe that $\forall D_{23} \geq D_{23}^*$, $\delta(D_1, D_2, D_{12}) \geq \delta(D_2, D_3, D_{23})$ and hence a lower bound to the sum rate is $\frac{1}{2}\log\frac{1}{D_1 D_2 D_3} + \delta(D_1, D_2, D_{12})$. Therefore the above point achieves the minimum sum rate $\forall D_{23} \geq D_{23}^*$.

The proof of (ii)(b) follows similarly by noting that if $D_{12} \in \{D_{12} : \delta(D_2, D_3, D_{23}) \geq \delta(D_1, D_2, D_{12})\}$, the minimum sum rate is given by $\frac{1}{2}\log\frac{1}{D_1 D_2 D_3} + \delta(D_2, D_3, D_{23})$ which is achieved by the point:

$$\{R_1, R_2, R_3\} = \left\{\frac{1}{2}\log\frac{1}{D_1}, \frac{1}{2}\log\frac{1}{D_2} + \delta(D_2, D_3, D_{12}), \frac{1}{2}\log\frac{1}{D_3}\right\} \tag{77}$$

This proves the theorem. □

It is interesting to observe that the optimal encoding scheme introduces common codewords (creates an interaction) between descriptions 1 and 3, even though these two descriptions are never received simultaneously at the decoder. While common codewords typically imply redundancy in the system, in this case, introducing them allows for better coordination between the descriptions leading to a smaller rate for the common branch. We will use similar principles in the following section for the $L-$descriptions framework and show that the CMS scheme achieves the complete RD region for several distortion regimes.

Example 1: We consider an asymmetric setting where $D_1 = 0.1$, $D_2 = 0.15$ and $D_3 = 0.2$. Fig. 8 shows the regime of distortions where CMS achieves the complete rate-region and minimum sum rate. The blue region corresponds to the set of distortion pairs (D_{12}, D_{23}) wherein the CMS rate-region is complete. The green region denotes the minimum sum rate points. It is clearly evident from the figure that CMS achieves the minimum sum rate for a fairly large regime of distortions.

C. Points on the Boundary - L-Descriptions

We next extend the coding scheme described in the previous section to the $L-$descriptions framework. Observe that, in all the setups considered so far, we have used only a single common layer random variable. We will see, in the proof of Theorem 6, that multiple common layer random variables are necessary to achieve the complete rate region of several cross-sections of the rate-distortion region. We first describe the particular cross-section we are interested in and then state the result as part of Theorem 6. We impose only the following distortion constraints : $D_1, D_2, \ldots, D_L, D_{12}, D_{13}, \ldots, D_{1L}$ and limit the remaining distortions to 1. This essentially corresponds to an $L-$descriptions framework wherein each of the descriptions are either received individually or the first description is received along with another description from the set $\{2, 3, \ldots, L\}$. Observe that the 3-descriptions framework considered in the previous section is a special case of this $L-$channel setup.[7] In the following theorem, we assume without loss of generality that $D_2 \leq \ldots \leq D_L$. The results follow accordingly for any other permutation of the ordering.

Theorem 6: Consider a cross-section of the $L-$descriptions quadratic Gaussian rate-distortion region, wherein we only impose distortions $D_1, D_2, \ldots, D_L, D_{12}, D_{13}, \ldots, D_{1L}$. Without loss of generality, assume that $D_2 \leq \ldots \leq D_L$ holds. Then the CMS scheme achieves the following complete rate-region:

$$R_i \geq \frac{1}{2}\log\frac{1}{D_i}, \quad \forall i = \{1, 2, \ldots, L\} \tag{78}$$

$$R_1 + R_i \geq \frac{1}{2}\log\frac{1}{D_1 D_i} + \delta(D_1, D_L, D_{1L}) \tag{79}$$

if the distortions $D_{12}, D_{13}, \ldots, D_{1L}$ satisfy the following conditions:

$$D_{1i} = \frac{\sigma_1^2\sigma_i^2(1-\rho_{1i}^2)}{\sigma_1^2\sigma_i^2(1-\rho_{1i}^2) + \sigma_1^2 + \sigma_i^2 - 2\sigma_1\sigma_i\rho_{1i}} \tag{80}$$

where $\sigma_j^2 = \frac{D_j}{1-D_j}$, $\forall j \in \{1, 2, \ldots, L\}$ and $\rho_{12}, \rho_{13}, \ldots, \rho_{1L}$ are given by:

$$\rho_{1i} = \rho_{1L}^*\frac{\sigma_2}{\sigma_i}, \quad \forall i \in \{1, 2, \ldots, L\} \tag{81}$$

Remark 12: It is possible to show that the CMS scheme achieves the minimum sum rate in several distortion regimes similar to Theorem 5. However, we omit explicitly stating the distortion regimes here as they can be derived directly from the above theorem.

Proof: Following the arguments as in the proof of Theorem 5, the converse for the rate region in (78) follows directly using Ozarow's converse results in [5]. The encoding scheme which achieves this sum rate resembles that in Theorem 5. First, a quantized version of the source, optimized to achieve D_L, is sent in description L. This information is also sent as part of all descriptions $\{2\ldots, L-l\}$. Then, refinement information is sent as part of description $L-1$ which helps in achieving a lower distortion D_{L-1}. This first layer of refinement information is also sent as part of all descriptions $\{2, \ldots, L-2\}$. Such an encoding mechanism is repeated successively. The refinement information generated at layer l is sent in description l, as well as all descriptions $\{2\ldots, L-l-1\}$. Note that successive refinability of a Gaussian source under MSE [33] ensures that there is no rate loss in encoding the source using such a mechanism. Finally, the first

[7]Note that 'description 1' here plays the role of 'descriptions 2' considered in the 3-descriptions setting in the previous section.

description carries enough information to achieve D_1 and D_{1i}, when the respective descriptions are received at the decoder.

Specifically, we choose the following joint density for the auxiliary random variables:

$$\begin{aligned} U_1 &= X + W_1 \\ U_2 &= X + W_2 \\ V_{23} &= X + W_2 + W_3 \\ V_{234} &= X + W_2 + W_3 + W_4 \\ &\vdots \\ V_{23\ldots L} &= X + W_2 + W_3 + \ldots + W_L \\ U_3 &= V_{23}, \quad U_4 = V_{234}, \ldots, U_L = V_{23\ldots L} \end{aligned} \tag{82}$$

We set all the other auxiliary random variables to constants and set the functions $\psi(\cdot)$ to be the respective optimal MSE estimators where $\{W_1, W_2, \ldots, W_L\}$ are zero mean jointly Gaussian random variables with the covariance matrix given by:

$$K_{W_1 \ldots W_L} = \begin{bmatrix} \tilde{\sigma}_1^2 & \rho_{12}\tilde{\sigma}_1\tilde{\sigma}_2 & 0 & \cdots & 0 \\ \rho_{12}\tilde{\sigma}_1\tilde{\sigma}_2 & \tilde{\sigma}_2^2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & & 0 \\ & \vdots & & & \vdots \\ 0 & 0 & 0 & \cdots & \tilde{\sigma}_L^2 \end{bmatrix} \tag{83}$$

where $\tilde{\sigma}_1^2 = \sigma_1^2 = \frac{D_1}{1-D_1}$, $\tilde{\sigma}_2^2 = \sigma_2^2 = \frac{D_2}{1-D_2}$ and $\tilde{\sigma}_j^2 = \sigma_j^2 - \sigma_{j-1}^2 = \frac{D_j}{1-D_j} - \frac{D_{j-1}}{1-D_{j-1}}$, $\forall j \in \{3, \ldots, L\}$. ρ_{12} is equal to ρ_{12}^* (ρ_{12}^* is defined in eq. (56)). This induces a correlation of ρ_{1i} (defined in (81)) between U_1 and U_i. Observe that for any fixed $\rho_{12}, \rho_{13}, \ldots, \rho_{1L}$, any distortion tuple satisfying the following conditions is achievable:

$$\begin{aligned} D_{1i} &\geq \text{Var}(X|U_1, U_i)\, i \in \{2, 3, \ldots, L-1\} \\ &= \frac{\sigma_1^2\sigma_i^2(1-\rho_{1i}^2)}{\sigma_1^2\sigma_i^2(1-\rho_{1i}^2) + \sigma_1^2 + \sigma_i^2 - 2\sigma_1\sigma_i\rho_{1i}} \end{aligned} \tag{84}$$

Next, we have from Theorem 1 that the following rates are achievable: $\forall l \in \{3, \ldots, L\}$:

$$\begin{aligned} R_1' &\geq \frac{1}{2}\log\frac{1}{D_1} \\ R_2' + \sum_{i=3}^{L} R''_{2\ldots l} &\geq \frac{1}{2}\log\frac{1}{D_2} \\ \sum_{i=l}^{L} R''_{2\ldots i} &\geq \frac{1}{2}\log\frac{1}{D_l} \\ R_1' + R_2' + \sum_{i=l}^{L} R''_{2\ldots i} &\geq \frac{1}{2}\log\frac{1}{D_1 D_2} + \delta(D_1, D_2, D_{12}) \\ R_1' + \sum_{i=l}^{L} R''_{2\ldots i} &\geq \frac{1}{2}\log\frac{1}{D_1 D_l} + \delta(D_1, D_L, D_{1L}) \\ R_1 &= R_1' \\ R_2 &= R_2' + \sum_{i=l}^{L} R''_{2\ldots i} \end{aligned}$$

$$R_l = \sum_{i=l}^{L} R''_{2\ldots i} \tag{85}$$

Following similar arguments as in Theorem 5, it follows directly that the above rate-region is equivalent to (78), completing the proof. □

VI. Conclusion

A novel encoding scheme for the general $L-$channel multiple descriptions problem was proposed which results in a new achievable rate-distortion region that subsumes the achievable region due to Venkataramani, Kramer and Goyal. The proposed encoding scheme adds controlled redundancy by including a common codeword in every subset of the descriptions. The common codewords assist in better coordination between the descriptions leading to a strictly larger region for a fairly general class of sources and distortion measures. In particular, we showed that the proposed scheme achieves a strictly larger rate-distortion region for a Gaussian source under MSE and for a binary symmetric source under Hamming distortion for all $L \geq 3$. We further showed that CMS achieves the complete rate-distortion region for several asymmetric cross-sections of the $L-$channel quadratic Gaussian MD problem. The possible impact of the new ideas presented here, on practical multiple descriptions encoder design and on certain special cases of the general setting, such as the symmetric multiple descriptions problem, will be studied as part of the future work.

Appendix A
Error Bounds in Theorem 1

Proof: We follow the notation and the notion of strong typicality defined in [34]. We refer to [34] (section 3) for formal definitions and basic Lemmas associated with typicality.

Let $\mathcal{E}$ denote the event of an encoding error. We have:

$$\begin{aligned} P(\mathcal{E}) = &P(\mathcal{E}|x^n \in \mathcal{T}_\epsilon^n)P(x^n \in \mathcal{T}_\epsilon^n) \\ &+ P(\mathcal{E}|x^n \notin \mathcal{T}_\epsilon^n)P(x^n \notin \mathcal{T}_\epsilon^n) \end{aligned} \tag{86}$$

From standard typicality arguments [28], we have $P(x^n \notin \mathcal{T}_\epsilon^n) < \epsilon$ as $n \to \infty$. Hence, it is sufficient to find conditions on the rates to bound $P(\mathcal{E}|x^n \in \mathcal{T}_\epsilon^n)$.

Towards finding conditions on the rate to bound $P(\mathcal{E}|x^n \in \mathcal{T}_\epsilon^n)$, we define event $\mathcal{A}(\{j\}_\mathcal{Q})$, for any index tuple $\{j\}_\mathcal{Q}$, as,

$$\mathcal{A}(\{j\}_\mathcal{Q}) = \left\{\left(x^n, v^n(\{j\}_\mathcal{Q}), u^n(\{j\}_\mathcal{Q})\right) \in \mathcal{T}_\epsilon^n\right\} \tag{87}$$

$\mathcal{Q} \subseteq \mathcal{J}(\mathcal{L})$, where for any set $\mathcal{Q} \in \mathcal{Q}^*$ (defined in (15)), $v^n(\{j\}_\mathcal{Q})$ denotes the codeword tuple $\{v^n_\mathcal{K}(\{j\}_{\mathcal{I}_{W+}(\mathcal{K})}, j_\mathcal{K})\forall\mathcal{K} \in \mathcal{Q}\}$ and $u^n(\{j\}_\mathcal{Q})$ denotes the tuple $\{u^n_\mathcal{K}(\{j\}_{\mathcal{I}_{1+}(\mathcal{K})}, \{j\}_\mathcal{K})\forall\mathcal{K} \subseteq [\mathcal{Q}]_1, \mathcal{K} \neq \emptyset\}$. Define random variables

$$\chi(\{j\}_{\mathcal{J}(\mathcal{L})}) = \begin{cases} 1 & \text{if } \mathcal{A}(\{j\}_{\mathcal{J}(\mathcal{L})})\text{occurs} \\ 0 & \text{else} \end{cases} \tag{88}$$

We have $P(\mathcal{E}|x^n \in \mathcal{T}_\epsilon^n) = P(\Psi = 0)$ where $\Psi = \sum_{\mathcal{J}(\mathcal{L})} \chi(\{j\}_{\mathcal{J}(\mathcal{L})})$. From Chebyshev's inequality, it follows that:

$$P(\Psi = 0) \leq P\left[|\Psi - E(\Psi)| \geq E(\Psi)/2\right] \leq \frac{4\text{Var}(\Psi)}{(E(\Psi))^2} \tag{89}$$

From standard typicality arguments, we can bound $E(\Psi)$ as follows:

$$E(\Psi) \geq 2^{n\sum_{\mathcal{K}\in\mathcal{I}_{1+}} R''_{\mathcal{K}} + n\sum_{l\in\mathcal{L}} R'_l - n(\alpha(\mathcal{J}(\mathcal{L}))+\epsilon)} \tag{90}$$

where for any set $\mathcal{Q}$ satisfying (15), we define:

$$\begin{aligned}\alpha(\mathcal{Q}) = &\sum_{\mathcal{K}\in\mathcal{Q}-[\mathcal{Q}]_1} H\left(V_{\mathcal{K}}|\{V\}_{\mathcal{I}_{|\mathcal{K}|+}(\mathcal{K})}\right)\\ &+ \sum_{\mathcal{K}\in 2^{[\mathcal{Q}]_1}-\emptyset} H\left(U_{\mathcal{K}}|\{V\}_{\mathcal{I}_{1+}(\mathcal{K})}, \{U\}_{2^{\mathcal{K}}-\emptyset-\mathcal{K}}\right)\\ &- H\left(\{V\}_{\mathcal{Q}-[\mathcal{Q}]_1}, \{U\}_{2^{[\mathcal{Q}]_1}-\emptyset}|X\right)\end{aligned} \tag{91}$$

We follow the convention, $\alpha(\emptyset)=0$. Next consider $\text{Var}(\Psi) = E(\Psi^2) - (E(\Psi))^2$ where,

$$\begin{aligned}E(\Psi^2) &= \sum_{\{j\}_{\mathcal{J}(\mathcal{L})}} \sum_{\{k\}_{\mathcal{J}(\mathcal{L})}} E\left[\chi(\{j\}_{\mathcal{J}(\mathcal{L})})\chi(\{k\}_{\mathcal{J}(\mathcal{L})})\right]\\ &= \sum_{\{j\}_{\mathcal{J}(\mathcal{L})}} \sum_{\{k\}_{\mathcal{J}(\mathcal{L})}} P\left[\mathcal{A}(\{j\}_{\mathcal{J}(\mathcal{L})}) \cap \mathcal{A}(\{k\}_{\mathcal{J}(\mathcal{L})})\right]\end{aligned} \tag{92}$$

The probability in (92) depends on whether $\{j\}_{\mathcal{J}(\mathcal{L})}$ and $\{k\}_{\mathcal{J}(\mathcal{L})}$ are equal for a subset of indices. Let $\mathcal{Q} \subseteq \mathcal{J}(\mathcal{L})$ such that $\{j\}_{\mathcal{Q}} = \{k\}_{\mathcal{Q}}$. Observe that, due to the hierarchical structure in the conditional codebook generation mechanism, for $\left\{v^n(\{j\}_{\mathcal{Q}}), u^n(\{j\}_{\mathcal{Q}})\right\} = \left\{v^n(\{k\}_{\mathcal{Q}}), u^n(\{k\}_{\mathcal{Q}})\right\}$ to hold, $\mathcal{Q}$ must satisfy (15), i.e., it must be such that,

$$\text{if } \mathcal{K} \in \mathcal{Q} \Rightarrow \mathcal{I}_{|\mathcal{K}|+}(\mathcal{K}) \subset \mathcal{Q} \tag{93}$$

It follows from codebook generation that given the codeword tuple $\{v^n(\{j\}_{\mathcal{Q}}), u^n(\{j\}_{\mathcal{Q}})\}$, tuples $\{v^n(\{j\}_{\mathcal{J}(\mathcal{L})-\mathcal{Q}}), u^n(\{j\}_{\mathcal{J}(\mathcal{L})-\mathcal{Q}})\}$ and $\{v^n(\{k\}_{\mathcal{J}(\mathcal{L})-\mathcal{Q}}), u^n(\{k\}_{\mathcal{J}(\mathcal{L})-\mathcal{Q}})\}$ are independent and identically distributed. Hence we can rewrite the probability in (92) as:

$$\begin{aligned}P&\left[\mathcal{A}(\{j\}_{\mathcal{J}(\mathcal{L})}) \cap \mathcal{A}(\{k\}_{\mathcal{J}(\mathcal{L})})\right]\\ &= \left(P\left[\mathcal{A}(\{j\}_{\mathcal{J}(\mathcal{L})-\mathcal{Q}})\middle|\mathcal{A}(\{j\}_{\mathcal{Q}}\right]\right)^2\\ &\quad\times P\left[\mathcal{A}(\{j\}_{\mathcal{Q}})\right] \qquad (94)\\ &= \left(P\left[\mathcal{A}(\{j\}_{\mathcal{J}(\mathcal{L})})\middle|\mathcal{A}(\{j\}_{\mathcal{Q}}\right]\right)^2\\ &\quad\times P\left[\mathcal{A}(\{j\}_{\mathcal{Q}})\right] \qquad (95)\\ &= \frac{\left(P\left[\mathcal{A}(\{j\}_{\mathcal{J}(\mathcal{L})})\right]\right)^2}{P\left[\mathcal{A}(\{j\}_{\mathcal{Q}})\right]} \qquad (96)\end{aligned}$$

Note that if $\mathcal{Q} = \emptyset$ (i.e., $\{j\}_{\mathcal{J}(\mathcal{L})}$ are $\{k\}_{\mathcal{J}(\mathcal{L})}$ are not equal for any subset of the indices), then $P\left[\mathcal{A}(\{j\}_{\mathcal{J}(\mathcal{L})}) \cap \mathcal{A}(\{k\}_{\mathcal{J}(\mathcal{L})})\right] = \left(P\left[\mathcal{A}(\{j\}_{\mathcal{J}(\mathcal{L})})\right]\right)^2$. Hence, we bound $\text{Var}(\Psi)$ by (see [1] for a similar argument):

$$\begin{aligned}\text{Var}(\Psi) &\leq \sum\Bigg\{N(\mathcal{Q})P\left[\mathcal{A}(\{j\}_{\mathcal{Q}})\right]\\ &\qquad\times\left(P\left[\mathcal{A}(\{j\}_{\mathcal{J}(\mathcal{L})})\middle|\mathcal{A}(\{j\}_{\mathcal{Q}})\right]\right)^2\Bigg\}\\ &= \sum\left\{N(\mathcal{Q})\frac{\left(P\left[\mathcal{A}(\{j\}_{\mathcal{J}(\mathcal{L})})\right]\right)^2}{P\left[\mathcal{A}(\{j\}_{\mathcal{Q}})\right]}\right\}\end{aligned} \tag{97}$$

where the summation is over all $\mathcal{Q} \subseteq 2^{\mathcal{L}} - \emptyset$ such that (15) holds (i.e., over all $\mathcal{Q} \in \mathcal{Q}^*$) and $\mathcal{A}(\{j\}_{\mathcal{Q}})$ denotes the event that $\left\{\mathcal{A}(\{j\}_{\mathcal{Q}}) = \left(x^n, v^n(\{j\}_{\mathcal{Q}}), u^n(\{j\}_{\mathcal{Q}})\right) \in \mathcal{T}^n_{\epsilon}\right\}$. $N(\mathcal{Q})$ denotes the total number of ways of choosing $\{j\}_{\mathcal{J}(\mathcal{L})}$ and $\{k\}_{\mathcal{J}(\mathcal{L})}$ such that they overlap in the subset $\mathcal{Q}$. An upper bound on $N(\mathcal{Q})$ is given in eq. (98), as shown at the bottom of this page. Also observe that the term corresponding to $\mathcal{Q}=\emptyset$ gets canceled with the '$-(E(\Psi))^2$' term in $\text{Var}(\Psi)$ leaving a summation over all non-empty $\mathcal{Q} \in \mathcal{Q}^*$.

On substituting the upper bound for $N(\mathcal{Q})$, we have:

$$\begin{aligned}\text{Var}(\Psi) \leq \sum\Bigg\{&2^{-2n\left(\alpha(\mathcal{J}(\mathcal{L})) - \sum_{\mathcal{K}\in\mathcal{I}_{1+}} R''_{\mathcal{K}} - \sum_{l\in\mathcal{L}} R'_l\right)}\\ &2^{n\left(\alpha(\mathcal{Q}) - \sum_{\mathcal{K}\in\mathcal{Q}-[\mathcal{Q}]_1} R''_{\mathcal{K}} - \sum_{l\in[\mathcal{Q}]_1} R'_l\right)+3n\epsilon}\Bigg\}\end{aligned} \tag{99}$$

where the summation is over all $\mathcal{Q} \in 2^{\mathcal{L}} - \emptyset$ such that (15) holds, i.e., over all $\mathcal{Q} \in \mathcal{Q}^*$.

Inserting (99) and (90) in (89), we get,

$$P(\mathcal{E}) \leq 4\sum 2^{n\left(\alpha(\mathcal{Q}) - \sum_{\mathcal{K}\in\mathcal{Q}-[\mathcal{Q}]_1} R''_{\mathcal{K}} - \sum_{l\in[\mathcal{Q}]_1} R'_l\right)+5n\epsilon} \tag{100}$$

where the summation is over all $\mathcal{Q} \in 2^{\mathcal{L}} - \emptyset$ such that (15) holds, i.e., over all $\mathcal{Q} \in \mathcal{Q}^*$. $P(\mathcal{E})$ can be made arbitrarily small if $\sum_{\mathcal{K}\in\mathcal{Q}-[\mathcal{Q}]_1} R''_{\mathcal{K}} + \sum_{l\in[\mathcal{Q}]_1} R'_l > \alpha(\mathcal{Q}) \forall \mathcal{Q} \in \mathcal{Q}^*$. □

Appendix B
Proof of Theorem 4.iii

In this appendix, we will show that a Gaussian random variable, under MSE distortion belongs to both $\mathcal{Z}_{EC}$ and $\mathcal{Z}_{ZB}$ and hence CMS achieves points strictly outside the VKG region. Throughout this section, we use the notation $\mathcal{C}_G$, to denote the set of all jointly Gaussian distributions.

First, let us recall the definitions of $\mathcal{Z}_{ZB}$ and $\mathcal{Z}_{EC}$. $\mathcal{Z}_{ZB}$ is defined in Definition 2 as the set of all random variables X for which there exists a strict sub-optimality in the ZB region (with respect to the given distortion measures) when the closure of the rates in (4) is defined only over joint distributions

$$\begin{aligned}N(\mathcal{Q}) &= 2^{n\left(\sum_{\mathcal{K}\in\mathcal{Q}-[\mathcal{Q}]_1} R''_{\mathcal{K}} + \sum_{l\in[\mathcal{Q}]_1} R'_l\right)} \prod_{\mathcal{K}\in\mathcal{J}(\mathcal{L})-\mathcal{Q}-[\mathcal{J}(\mathcal{L})-\mathcal{Q}]_1} 2^{nR'_{i,\mathcal{K}}}(2^{nR'_{i,\mathcal{K}}}-1) \prod_{l\in[\mathcal{J}(\mathcal{L})-\mathcal{Q}]_1} 2^{nR'_{i,l}}(2^{nR'_{i,l}}-1)\\ &\leq 2^{n\{\sum_{\mathcal{K}\in\mathcal{Q}-[\mathcal{Q}]_1} R''_{i,\mathcal{K}} + \sum_{l\in[\mathcal{Q}]_1} R'_{i,l} + 2\sum_{\mathcal{K}\in\mathcal{J}(\mathcal{L})-\mathcal{Q}-[\mathcal{J}(\mathcal{L})-\mathcal{Q}]_1} R''_{i,\mathcal{K}} + 2\sum_{l\in[\mathcal{J}(\mathcal{L})-\mathcal{Q}]_1} R'_{i,l}\}}\\ &= 2^{n\{2\sum_{\mathcal{K}\in I_{1+}} R''_{i,\mathcal{K}} + 2\sum_{l\in\mathcal{L}} R'_{i,l} - \sum_{\mathcal{K}\in\mathcal{Q}-[\mathcal{Q}]_1} R''_{i,\mathcal{K}} - \sum_{l\in[\mathcal{Q}]_1} R'_{i,l}\}}\end{aligned} \tag{98}$$

that satisfy the following conditions:

$$\begin{aligned} U_1 &\leftrightarrow (X, V_{12}) \leftrightarrow U_2 \\ E\left[d_{\mathcal{K}}(X, \psi_{\mathcal{K}}(U_{\mathcal{K}}))\right] &\leq D_{\mathcal{K}}, \mathcal{K} \in \{1, 2, 12\} \\ U_{12} &= f(U_1, U_2, V_{12}) \end{aligned} \tag{101}$$

where f is some deterministic function. Note that the statement in Definition 2 is more general than the above description. However, for a Gaussian source, under MSE, it is possible to show strict sub-optimality in ZB region even for $\epsilon = 0$, which leads to the above constraints for the joint distributions. Similarly, X is said to belong to $\mathcal{Z}_{EC}$ if there exists a strict sub-optimality in the EC region when the closure of the rates in (3) is defined only over joint distributions that satisfy the following conditions:

$$\begin{aligned} U_1 &\leftrightarrow X \leftrightarrow U_2 \\ E\left[d_{\mathcal{K}}(X, \psi_{\mathcal{K}}(U_{\mathcal{K}}))\right] &\leq D_{\mathcal{K}}, \mathcal{K} \in \{1, 2, 12\} \\ U_{12} &= f(U_1, U_2) \end{aligned} \tag{102}$$

where f is some deterministic function. We next show that a Gaussian source under MSE belongs to both $\mathcal{Z}_{EC}$ and $\mathcal{Z}_{ZB}$. We first prove the result for $\mathcal{Z}_{EC}$ and then extend similar arguments for $\mathcal{Z}_{ZB}$.

1) Proof for $\mathcal{Z}_{EC}$: We first give an intuitive argument to justify the claim and then follow it up with the formal proof. It follows from Ozarow's results (see also [2]) that there exists a regime of distortions (D_1, D_2, D_{12}) for which the following rate region is achievable (and complete):

$$\begin{aligned} R_1 &\geq \frac{1}{2}\log\frac{1}{D_1} \\ R_2 &\geq \frac{1}{2}\log\frac{1}{D_2} \\ R_1 + R_2 &\geq \frac{1}{2}\log\frac{1}{D_{12}} \end{aligned} \tag{103}$$

i.e., there is no excess rate incurred due to encoding the source using two descriptions. Equivalently, there exists a regime of distortions for which the excess sum rate term in the EC region (i.e., $I(U_1; U_2)$) must be set to zero to achieve the complete rate-region. We will show that, if we restrict the optimization to conditionally independent joint densities, then it is impossible to simultaneously satisfy all the distortions and achieve $I(U_1; U_2) = 0$.

We next provide a formal proof. Our objective is to show that for the 2-descriptions quadratic Gaussian MD problem, there is a strict sub-optimality if we perform the EC encoding scheme and restrict the optimization to joint densities satisfying the following conditions:

$$\begin{aligned} P(U_1, U_2|X) &= P(U_1|X) \\ &\quad \times P(U_2|X) \\ E\left[(X - \psi_i(U_i))^2\right] &\leq D_i, \quad i \in \{1, 2\} \\ E\left[(X - \psi_{12}(U_{12}))^2\right] &\leq D_{12} \\ U_{12} &= f(U_1, U_2) \end{aligned} \tag{104}$$

for some functions $\psi_1, \psi_2, \psi_{12}$ and f. We denote by $\mathcal{RD}_{EC}^{IQ}$, the closure of all the achievable RD tuples using the EC encoding scheme, over all joint densities satisfying (104). We need to show that $\mathcal{RD}_{EC}^{IQ}$ is strictly smaller than $\mathcal{RD}_G^2$. We consider one point in $\mathcal{RD}_G^2$ and show that it is not contained in $\mathcal{RD}_{EC}^{IQ}$.

Let the distortions be such that $D_{12} \leq D_1 + D_2 - 1$ holds. This regime of distortions has been termed as the 'high distortion regime' in the literature and the complete rate region under these constraints is given by (103) (see for example [5], [29], [32]). Let us consider the following rate point : $P_0 \triangleq (R_1, R_2) = (\frac{1}{2}\log\frac{1}{D_1}, \frac{1}{2}\log\frac{D_1}{D_{12}})$. We will show that this point in not contained in $\mathcal{RD}_{EC}^{IQ}$ when $D_{12} \leq D_1 + D_2 - 1$. We first rewrite the EC region for any joint density over the random variables (X, U_1, U_2, U_{12}) and functions ψ_1, ψ_2 and ψ_{12}:

$$\begin{aligned} R_1 &\geq I(X; U_1) \\ R_2 &\geq I(X; U_2) \\ R_1 + R_2 &\geq I(X; U_1, U_2, U_{12}) + I(U_1; U_2) \\ D_{\mathcal{K}} &\geq E\left[(X - \psi_{\mathcal{K}}(U_{\mathcal{K}}))^2\right], \quad \mathcal{K} \in 2^{\{1,2\}} - \emptyset \end{aligned} \tag{105}$$

Observe that, to achieve P_0, we have to impose the joint density for (X, U_1) and the function ψ_1 to be RD optimal at D_1. This requires (X, U_1) to be jointly Gaussian and ψ_1 to be the optimal estimator of X given U_1. Note that, if $P(U_1|X)$ is not Gaussian (with ψ_1 being the MMSE estimator), it is possible to construct a jointly Gaussian distribution which achieves the same distortion with smaller rate. For example, one could construct a distribution $(X, \tilde{U}_1) \in \mathcal{C}_G$, with variance of $P(\tilde{U}_1|X)$ equal to the variance of $P(\psi_1(U_1)|X)$. As a Gaussian distribution achieves maximum entropy among all distributions with the same variance, $I(X; \tilde{U}_1) < I(X; U_1)$, leading to a smaller rate for R_1. Therefore, we can assume that $(X, U_1) \in \mathcal{C}_G$. We denote this joint density by $P_G(X, U_1)$. It follows that the infimum R_2 achievable using an independent quantization scheme when $R_1 = \frac{1}{2}\log\frac{1}{D_1}$ is given by:

$$\begin{aligned} R_{EC}^{IQ} &= \inf R_2 : \Big\{R_1 = \frac{1}{2}\log\frac{1}{D_1}, \\ &\qquad (R_1, R_2) \in \mathcal{RD}_{EC}^{IQ}\Big\} \\ &= \inf\Big\{I(X; U_2, U_{12}|U_1) + I(U_1; U_2)\Big\} \end{aligned} \tag{106}$$

where the infimum is over all joint densities $P(X, U_1, U_2, U_{12}) = P_G(X, U_1)P(U_2, U_{12}|X, U_1)$ and functions $\psi_1, \psi_2, \psi_{12}$ and f satisfying (104), where $(X, U_1) \in \mathcal{C}_G$ is RD optimal at D_1. We will next show that $R_{EC}^{IQ} > \frac{1}{2}\log\frac{D_1}{D_{12}}$.

In the following Lemma, we begin by proving that R_{EC}^{IQ} is greater than or equal R_{EC}^G, where R_{EC}^G is defined as:

$$R_{EC}^G = \inf\Big\{I(X; \tilde{U}_2, \tilde{U}_{12}|U_1) + I(U_1; \tilde{U}_2)\Big\} \tag{107}$$

where the infimum is over all jointly Gaussian densities $Q(X, U_1, \tilde{U}_2, \tilde{U}_{12}) = P_G(X, U_1)Q(\tilde{U}_2, \tilde{U}_{12}|X, U_1)$ satisfying the following conditions:

$$Q(U_1, \tilde{U}_2|X) = Q(U_1|X) \times Q(\tilde{U}_2|X)$$
$$E\left[(X - \tilde{U}_2)^2\right] \le D_2$$
$$E\left[(X - \tilde{U}_{12})^2\right] \le D_{12} \tag{108}$$

Lemma 1: Let R_{EC}^{IQ} be defined as in (106) and R_{EC}^{G} be defined as in (107). Then, we have:

$$R_{EC}^{IQ} \ge R_{EC}^{G} \tag{109}$$

Proof: Consider any joint density $P_G(X, U_1)$ $P(U_2, U_{12}|X, U_1)$ and functions $\psi_1, \psi_2, \psi_{12}, f$, satisfying (104), where (X, U_1) are distributed according to a jointly Gaussian density which is RD optimal at D_1. We now construct a joint density $P_G(X, U_1)Q(\tilde{U}_2, \tilde{U}_{12}|X, \tilde{U}_1)$ that achieves a smaller value for the quantity in (106) and is in $\mathcal{C}_G$. Consider $P_G(X, U_1)Q(\tilde{U}_2, \tilde{U}_{12}|X, U_1) \in \mathcal{C}_G$ such that $K_{\tilde{U}_2, \tilde{U}_{12}|X, U_1} = K_{\psi_2(U_2), \psi_{12}(U_{12})|X, U_1}$, i.e., the random variables $(\tilde{U}_2, \tilde{U}_{12})$ have the same covariance matrix as $(\psi_2(U_2), \psi_{12}(U_{12}))$, given (X, U_1). Clearly, all the conditions in (108) are satisfied by the joint density Q. We need to show that the quantity in (106) is smaller for the joint density Q compared to P. Recall that, for a fixed covariance matrix, a Gaussian distribution over the relevant random variables maximizes the conditional entropy [35]. Hence, we have:

$$\begin{aligned} I(X; U_2, U_{12}|U_1) &= H(X|U_1) - H(X|U_1, U_2, U_{12}) \\ &\quad + I(U_1; U_2) + H(U_1) - H(U_1|U_2) \\ &\ge H(X|U_1) - H(X|U_1, \tilde{U}_2, \tilde{U}_{12}) \\ &\quad + H(U_1) - H(U_1|\tilde{U}_2) \\ &= I(X; \tilde{U}_2, \tilde{U}_{12}|U_1) + I(U_1; \tilde{U}_2) \end{aligned} \tag{110}$$

leading to $R_{EC}^{IQ} \ge R_{EC}^{G}$. □

Equipped with this result, we continue with our proof to show that $R_{EC}^{G} > \frac{1}{2}\log\frac{D_1}{D_{12}}$. Consider the following series of inequalities:

$$\begin{aligned} R_{EC}^{IQ} \ge R_{EC}^{G} &= \inf\left\{I(X; \tilde{U}_2, \tilde{U}_{12}|U_1) + I(U_1; \tilde{U}_2)\right\} \\ &= \inf\left\{I(X; \tilde{U}_2, \tilde{U}_{12}, U_1) + I(U_1; \tilde{U}_2) - \frac{1}{2}\log\frac{1}{D_1}\right\} \\ &\ge^{(a)} \frac{1}{2}\log\frac{D_1}{D_{12}} + \inf I(U_1; \tilde{U}_2) \\ &>^{(b)} \frac{1}{2}\log\frac{D_1}{D_{12}} \end{aligned} \tag{111}$$

where the infimum is over all $Q(\tilde{U}_2, \tilde{U}_{12}|X, \tilde{U}_1) \in \mathcal{C}_G$ satisfying (108). Note that (a) follows from the fact that $E\left[(X - \tilde{U}_{12})^2\right] \le D_{12}$ (and therefore $I(X; \tilde{U}_2, \tilde{U}_{12}, U_1) \ge I(X; \tilde{U}_{12}) \ge \frac{1}{2}\log\frac{1}{D_{12}}$). (b) follows from the fact that $(X, U_1, \tilde{U}_2)$ are jointly Gaussian satisfying the Markov condition $U_1 \leftrightarrow X \leftrightarrow \tilde{U}_2$, where $I(X; U_1) = \frac{1}{2}\log\frac{1}{D_1}$ and $I(X; \tilde{U}_2) \ge \frac{1}{2}\log\frac{1}{D_2}$. Therefore, $\inf I(U_1; \tilde{U}_2) > 0$, proving that there is a strict sub-optimality if we restrict the EC coding scheme to an independent quantization mechanism for a Gaussian random variable under MSE. Hence a Gaussian random variable, under MSE, belongs to $\mathcal{Z}_{EC}$.

2) Proof for $\mathcal{Z}_{ZB}$: Our next objective is to prove that a Gaussian random variable under MSE also belongs to $\mathcal{Z}_{ZB}$. Equivalently, we need to show that there is a strict sub-optimality in the ZB region when we restrict the optimization to joint densities satisfying the following constraints:

$$\begin{aligned} P(U_1, U_2|X, V_{12}) &= P(U_1|X, V_{12}) \times P(U_2|X, V_{12}) \\ E\left[(X - \psi_i(U_i))^2\right] &\le D_i, i \in \{1, 2\} \\ E\left[(X - \psi_{12}(U_{12}))^2\right] &\le D_{12} \\ U_{12} &= f(U_1, U_2, V_{12}) \end{aligned} \tag{112}$$

The proof follows in very similar lines to that for $\mathcal{Z}_{EC}$, the main difference being that all the quantities are now defined conditioned on every value of V_{12}. Recall that the ZB region achievable using any joint density $P(X, V_{12}, U_1, U_2, U_{12})$ is given by:

$$\begin{aligned} R_1 &\ge I(X; V_{12}, U_1) \\ R_2 &\ge I(X; V_{12}, U_2) \\ R_1 + R_2 &\ge I(X; V_{12}) + I(U_1; U_2|V_{12}) \\ &\quad + I(X; V_{12}, U_1, U_2, U_{12}) \\ D_{\mathcal{K}} &\ge E\left[(X - \psi_{\mathcal{K}}(U_{\mathcal{K}}))^2\right], \quad \mathcal{K} \subseteq \{1, 2\}, \ \mathcal{K} \ne \emptyset \end{aligned} \tag{113}$$

Before we continue with the proof, we take a small digression and prove certain properties of the ZB region. We summarize these properties in the following lemma.

Lemma 2: The closure of the RD-tuples in $\mathcal{RD}_{ZB}$ can be restricted to joint distributions that satisfy the following Markov chain conditions, without any loss in the achievable region:

$$\begin{aligned} &X \leftrightarrow U_1 \leftrightarrow V_{12} \\ &X \leftrightarrow U_1 \leftrightarrow V_{12} \\ &X \leftrightarrow U_{12} \leftrightarrow U_1, U_2, V_{12} \end{aligned} \tag{114}$$

Proof: Let $P(X, V_{12}, U_1, U_2, U_{12})$ be any joint density, which may or may not satisfy the above properties, and let ψ_1, ψ_2 and ψ_{12} be functions that satisfy the respective distortion constraints. We will construct another joint density $Q(X, \tilde{V}_{12}, \tilde{U}_1, \tilde{U}_2, \tilde{U}_{12})$ that satisfies (114) and achieves the same RD region. Towards constructing such a joint density, we set:

$$\begin{aligned} \tilde{V}_{12} &= V_{12} \\ \tilde{U}_1 &= (U_1, V_{12}) \\ \tilde{U}_2 &= (U_2, V_{12}) \\ \tilde{U}_{12} &= (U_{12}, U_1, U_2, V_{12}) \end{aligned}$$

First, observe that the above choice of the joint density satisfies (114). Also observe that, $I(X; V_{12}, U_1) = I(X; \tilde{V}_{12}, \tilde{U}_1)$, $I(X; V_{12}, U_2) = I(X; \tilde{V}_{12}, \tilde{U}_2)$, $I(X; V_{12}, U_1, U_2, U_{12}) = I(X; \tilde{V}_{12}, \tilde{U}_1, \tilde{U}_2, \tilde{U}_{12})$ and $I(U_1; U_2|V_{12}) = I(\tilde{U}_1; \tilde{U}_2|\tilde{V}_{12})$.

Hence, the rate region achievable by $Q(\cdot)$ is the same as that of $P(\cdot)$. Moreover, all the distortion constraints can be satisfied using the same functions ψ_1, ψ_2 and ψ_{12}. Hence it follows that the entire ZB region can be achieved by considering a closure only over random variable that satisfy (114). □

Corollary 1: The closure of the RD-tuples in $\mathcal{RD}^{IQ}_{ZB}$ can be restricted to joint distributions that satisfy (114), without any loss in $\mathcal{RD}^{IQ}_{ZB}$.

Proof: The proof is very similar to that of Lemma 2. We omit restating it here. □

Equipped with these results, we continue our proof. Let us again consider the corner point $P_0 \triangleq (R_1, R_2) = (\frac{1}{2}\log\frac{1}{D_1}, \frac{1}{2}\log\frac{D_1}{D_{12}})$ for some (D_1, D_2, D_{12}) satisfying $D_{12} \leq D_1 + D_2 - 1$, and show that it is not achievable by ZB when we restrict the joint densities to satisfy (112). Note that, as a result of Lemma (2) and Corollary (1), it is sufficient to consider joint densities that also satisfy (114). Observe that, as $I(X; V_{12}, U_1, U_2, U_{12}) \geq \frac{1}{2}\log\frac{1}{D_{12}}$, P_0 can be achieved only by joint densities that satisfy $I(X; V_{12}) = 0$. Hence, to prove that a Gaussian random variable under MSE belongs to $\mathcal{Z}_{ZB}$, it is sufficient to show that P_0 is not achievable when we restrict the optimization to joint densities satisfying (112), (114) and $I(X; V_{12}) = 0$.

Let $P(V_{12}, U_1, U_2, U_{12}, X)$ be any such joint density and let $\mathcal{V}_{12}$ be the corresponding alphabet for V_{12}. Then the associated achievable region can be rewritten as:

$$\begin{aligned}
R_1 &\geq I(X; U_1|V_{12}) \\
&= \sum_{v_{12}\in\mathcal{V}_{12}} P(v_{12}) I(X; U_1|V_{12} = v_{12}) \\
R_2 &\geq I(X; U_2|V_{12}) \\
&= \sum_{v_{12}\in\mathcal{V}_{12}} P(v_{12}) I(X; U_2|V_{12} = v_{12}) \\
R_1 + R_2 &\geq I(U_1; U_2|V_{12}) \\
&\quad + I(X; U_1, U_2, U_{12}|V_{12}) \\
&= \sum_{v_{12}\in\mathcal{V}_{12}} P(v_{12}) \Big[I(U_1; U_2|V_{12} = v_{12}) \\
&\qquad\qquad + I(X; U_1, U_2, U_{12}|V_{12} = v_{12}) \Big] \\
D_{\mathcal{K}} &\geq E\left[(X - \psi_{\mathcal{K}}(U_{\mathcal{K}}))^2\right], \quad \mathcal{K} \subset \{1, 2\},\ \mathcal{K} \neq \emptyset \\
&= E\left[E\left[(X - \psi_{\mathcal{K}}(U_{\mathcal{K}}))^2 \Big| V_{12}\right]\right]
\end{aligned} \tag{115}$$

We will next show that the optimization can be further restricted to joint densities $P(X, V_{12}) Q(\tilde{U}_1, \tilde{U}_2, \tilde{U}_{12}|X, V_{12})$ such that $(X, \tilde{U}_1, \tilde{U}_2, \tilde{U}_{12}) \in \mathcal{C}_G$ given $V_{12} = v_{12}$, $\forall v_{12} \in \mathcal{V}_{12}$. First, as X is independent of V_{12}, $P(X|V_{12} = v_{12})$ is Gaussian $\forall v_{12} \in \mathcal{V}_{12}$. Recall that P_0 is obtained by first minimizing R_1, followed by minimizing R_2 given R_1 subject to all the distortion constraints. From (115), we have:

$$R_1 = \inf \sum_{v_{12}\in\mathcal{V}_{12}} P(v_{12}) I(X; U_1|V_{12} = v_{12}) \tag{116}$$

where the infimum is over all joint densities $P(X, V_{12}, U_1)$ satisfying the distortion constraint D_1.

Towards proving that $(X, U_1) \in \mathcal{C}_G$ given $V_{12} = v_{12}$, $\forall v_{12} \in \mathcal{V}_{12}$, let $P(X, V_{12}, U_1)$ be any joint density over the given alphabets and let function $\psi_1(\cdot)$ achieve the distortion constraint at D_1. Consider the following joint density $P(X, V_{12}) Q(\tilde{U}_1|X, V_{12})$, such that $(X, \tilde{U}_1)$ are jointly Gaussian given $V_{12} = v_{12}$ and $K_{X,\tilde{U}_1|V_{12}=v_{12}} = K_{X,\psi_1(U_1)|V_{12}=v_{12}}$, $\forall v_{12} \in \mathcal{V}_{12}$. Observe that this joint density satisfies the distortion constraint for D_1. Moreover, it achieves a smaller rate, as a Gaussian density over the relevant random variables maximizes the conditional entropy for a fixed covariance matrix, i.e., $I(X; U_1|V_{12} = v_{12}) \geq I(X; \tilde{U}_1|V_{12} = v_{12})$, $\forall v_{12} \in \mathcal{V}_{12}$. Therefore, to achieve minimum R_1, we can consider only densities wherein $(X, \tilde{U}_1)$ are jointly Gaussian given V_{12}.

We next focus on minimizing R_2 given $R_1 = \frac{1}{2}\log\frac{1}{D_1}$. From (115), we have:

$$\begin{aligned}
R_2 = \inf\Big\{ &\sum_{v_{12}\in\mathcal{V}_{12}} P(v_{12}) \Big[I(\tilde{U}_1; U_2|V_{12} = v_{12}) \\
&+ I(X; \tilde{U}_1, U_2, U_{12}|V_{12} = v_{12}) \Big] - R_1 \Big\}
\end{aligned} \tag{117}$$

where the infimum is over all joint densities $P(X, V_{12}, \tilde{U}_1) P(U_2, U_{12}|X, V_{12}, \tilde{U}_1)$ satisfying (112), (114) and $I(X; V_{12}) = 0$ and where $(X, \tilde{U}_1)$ are jointly Gaussian given $V_{12} = v_{12}$, $\forall v_{12} \in \mathcal{V}_{12}$ (required to minimize R_1). It is easy to show using similar arguments that, to achieve the infimum in (117), it is sufficient to consider joint densities where $(X, \tilde{U}_1, \tilde{U}_2, \tilde{U}_{12}) \in \mathcal{C}_G$ given $V_{12} = v_{12}$ and $Q(\tilde{U}_1, \tilde{U}_2|X, V_{12}) = Q(\tilde{U}_1|X, V_{12}) Q(\tilde{U}_2|X, V_{12})$, $\forall v_{12} \in \mathcal{V}_{12}$.

To summarize what we have so far, we have shown that to achieve minimum R_1 followed by minimum R_2 in $\mathcal{RD}^{IQ}_{ZB}$, it is sufficient to consider only joint densities $P(X, V_{12}, U_1, U_2, U_{12})$ that satisfy the following properties:

1) $I(U_1; U_2|X, V_{12}) = 0$, $U_{12} = f(U_1, U_2, V_{12})$ and $P(X, V_{12}, U_1, U_2, U_{12})$ satisfies all the distortion constraints
2) $I(X; V_{12}) = 0$
3) (X, U_1, U_2, U_{12}) are jointly Gaussian given $V_{12} = v_{12}$ $\forall v_{12} \in \mathcal{V}_{12}$
4) The Markov chain conditions in (114) are satisfied
5) U_1 achieves RD-optimality at D_1

We will show that the point P_0 cannot be achieved by any joint density that satisfies the above properties. Observe that, given $V_{12} = v_{12}$, (X, U_1, U_2) are jointly Gaussian random variables satisfying the Markov condition $U_1 \leftrightarrow X \leftrightarrow U_2$. This leads to two possibilities for the joint distribution $P(X, V_{12}, U_1, U_2, U_{12})$:

i) $\exists v_{12} \in \mathcal{V}_{12}$ such that $I(X; U_1|V_{12} = v_{12}) > 0$ and $I(X; U_2|V_{12} = v_{12}) > 0$: As (X, U_1, U_2) are jointly Gaussian, this leads to the conclusion that $I(U_1; U_2|V_{12} = v_{12}) > 0$ and hence $I(U_1; U_2|V_{12}) > 0$. This clearly implies that there is excess rate on R_2 and hence point P_0 is not achievable.

ii) $\forall v_{12} \in \mathcal{V}_{12}$, either $I(X; U_1|V_{12} = v_{12}) = 0$ or $I(X; U_2|V_{12} = v_{12}) = 0$. This implies that the alphabet space of V_{12} can be divided into two disjoint subsets such that, when V_{12} takes values in the first set, X and U_2 are independent, and when V_{12} takes values in the second set X and U_1 are independent. Let the two sets be denoted by $\mathcal{V}_{12}^1$ and $\mathcal{V}_{12}^2$, respectively. We have:

$$I(X; U_2|V_{12} \in \mathcal{V}_{12}^1) = 0$$
$$I(X; U_1|V_{12} \in \mathcal{V}_{12}^2) = 0 \tag{118}$$

However, recall that the joint distribution can be restricted to satisfy, $I(X; V_{12}) = 0$, $I(X; V_{12}|U_1) = 0$ and $I(X; V_{12}|U_2) = 0$. Hence, we have:

$$\begin{aligned} H(X) &= H(X|V_{12} \in \mathcal{V}_{12}^1) \\ &=^{(a)} H(X|U_2, V_{12} \in \mathcal{V}_{12}^1) \\ &= H(X|U_2) \end{aligned} \tag{119}$$

$$\begin{aligned} H(X) &= H(X|V_{12} \in \mathcal{V}_{12}^2) \\ &=^{(b)} H(X|U_1, V_{12} \in \mathcal{V}_{12}^2) \\ &= H(X|U_1) \end{aligned} \tag{120}$$

where (a) and (b) follow from (118). This implies that (X, U_1) and (X, U_2) must be pair-wise independent and hence the distortion constraints on D_1 and D_2 cannot be satisfied.

Hence, it follows that the point P_0 cannot be achieved by an 'independent quantization mechanism' using the ZB coding scheme for a Gaussian source under MSE, proving the theorem.

REFERENCES

[1] 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.

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

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

[4] R. Ahlswede, "The rate-distortion region for multiple descriptions without excess rate," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 721–726, Nov. 1985.

[5] 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.

[6] J. Wang, J. Chen, L. Zhao, P. Cuff, and H. Permuter, "On the role of the refinement layer in multiple description coding and scalable coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1443–1456, Mar. 2011.

[7] 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.

[8] 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. 2004.

[9] 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.

[10] 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.

[11] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 821–834, May 1993.

[12] P. Koulgi, S. L. Regunathan, and K. Rose, "Multiple description quantization by deterministic annealing," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 2067–2075, Aug. 2003.

[13] 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.

[14] S. Mohajer, C. Tian, and S. N. Diggavi, "Asymmetric multilevel diversity coding and asymmetric Gaussian multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4367–4387, Sep. 2010.

[15] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.

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

[17] 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.

[18] C. Tian, S. Mohajer, and S. N. 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.

[19] L. Song, S. Shuo, and J. Chen, "A lower bound on the sum rate of multiple description coding with symmetric distortion constraints," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 7547–7567, Dec. 2014.

[20] K. Viswanatha, E. Akyol, and K. Rose, "On optimum communication cost for joint compression and dispersive information routing," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Aug./Sep. 2010, pp. 1–5.

[21] J. Nayak, S. Ramaswamy, and K. Rose, "Correlated source coding for fusion storage and selective retrieval," in *Proc. IEEE Int. Symp. Inf. Theory*, Sep. 2005, pp. 92–96.

[22] K. Viswanatha, E. Akyol, and K. Rose, "Combinatorial message sharing for a refined multiple descriptions achievable region," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul./Aug. 2011, pp. 1312–1316.

[23] K. Viswanatha, E. Akyol, and K. Rose, "A strictly improved achievable region for multiple descriptions using combinatorial message sharing," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Oct. 2011, pp. 100–104.

[24] E. Akyol, K. Viswanatha, and K. Rose, "Combinatorial message sharing and random binning for multiple description coding," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2012, pp. 1371–1375.

[25] E. Akyol, K. Viswanatha, and K. Rose, "On random binning versus conditional codebook methods in multiple descriptions coding," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Sep. 2012, pp. 312–316.

[26] K. Viswanatha, E. Akyol, and K. Rose, "On the role of common codewords in quadratic Gaussian multiple descriptions coding," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2013, pp. 1367–1371.

[27] F. Shirani and S. S. Pradhan, "An achievable rate-distortion region for the multiple descriptions problem," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun./Jul. 2014, pp. 576–580.

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

[29] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[30] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder—II: General sources," *Inf. Control*, vol. 38, no. 1, pp. 60–80, Jul. 1978.

[31] Z. Zhang and T. Berger, "Multiple description source coding with no excess marginal rate," *IEEE Trans. Inf. Theory*, vol. 41, no. 2, pp. 349–357, Mar. 1995.

[32] R. Zamir, "Gaussian codes and Shannon bounds for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2629–2636, Nov. 1999.

[33] 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.

[34] T. S. Han and K. Kobayashi, "A unified achievable rate region for a general class of multiterminal source coding systems," *IEEE Trans. Inf. Theory*, vol. 26, no. 3, pp. 277–288, May 1980.

[35] J. Thomas, "Feedback can at most double Gaussian multiple access channel capacity (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 33, no. 5, pp. 711–716, Sep. 1987.

Kumar B. Viswanatha (S'08–M'15) received his PhD degree in 2013 in the Electrical and Computer Engineering department from University of California at Santa Barbara (UCSB), USA. He is currently working for Qualcomm research center in San Diego. Prior to joining Qualcomm, he was an intern associate in the equity volatility desk at Goldman Sachs Co., New York, USA. His research interests include multi-user information theory, wireless communications, joint compression and routing for networks and distributed compression for large scale sensor networks.

Emrah Akyol (S'03–M'12) received the Ph.D. degree in 2011 in electrical and computer engineering from the University of California at Santa Barbara. From 2006 to 2007, he held positions at Hewlett-Packard Laboratories and NTT Docomo Laboratories, both in Palo Alto, where he worked on topics in video compression.

Currently, Dr. Akyol is a postdoctoral researcher in the Department of Electrical and Computer Engineering, University of Illinois Urbana-Champaign. Prior to that, he was a post doctoral researcher at Electrical Engineering Department at University of Southern California and at University of California - Santa Barbara. His current research is on the interplay of networked information theory, communications and control.

Kenneth Rose (S'85–M'91–SM'01–F'03) received the Ph.D. degree in 1991 from the California Institute of Technology, Pasadena.

He then joined the Department of Electrical and Computer Engineering, University of California at Santa Barbara, where he is currently a Professor. His main research activities are in the areas of information theory and signal processing, and include rate-distortion theory, source and source-channel coding, audio and video coding and networking, pattern recognition, and nonconvex optimization. He is interested in the relations between information theory, estimation theory, and statistical physics, and their potential impact on fundamental and practical problems in diverse disciplines.

Dr. Rose was co-recipient of the 1990 William R. Bennett Prize Paper Award of the IEEE Communications Society, as well as the 2004 and 2007 IEEE Signal Processing Society Best Paper Awards.