On Common Information and the Encoding of Sources that are Not Successively Refinable

Kumar Viswanatha, Emrah Akyol, Tejaswi Nanjundaswamy and Kenneth Rose
ECE Department, University of California - Santa Barbara
{kumar,eakyol,tejaswi,rose}@ece.ucsb.edu

***Abstract*—This paper focuses on a new framework for scalable coding of information based on principles derived from common information of two dependent random variables. In the conventional successive refinement setting, the encoder generates two layers of information called the base layer and the enhancement layer. The first decoder, which receives only the base layer, produces a coarse reconstruction of the source, whereas the second decoder, which receives both the layers, uses the enhancement layer to refine the information further leading to a finer reconstruction. It is popularly known that asymptotic rate-distortion optimality at both the decoders is possible if and only if the source-distortion pair is successively refinable. However when the source is not successively refinable under the given distortion metric, it is impossible to achieve rate-distortion optimality at both the layers simultaneously. For this reason, most practical system designers resort to storing two individual representations of the source leading to significant overhead in transmission/storage costs. Inspired by the breadth of applications, in this paper, we propose a new framework for scalable coding wherein a subset of the bits sent to the first decoder is not sent to the second decoder. That is, the encoder generates one common bit stream which is routed to both the decoders, but unlike the conventional successive refinement setting, *both* the decoders receive an additional individual bitstream. By relating the proposed framework with the problem of common information of two dependent random variables, we derive a single letter characterization for the minimum sum rate achievable for the proposed setting when the two decoders are constrained to receive information at their respective rate-distortion functions. We show using a simple example that the proposed framework provides a strictly better asymptotic sum rate as opposed to the conventional scalable coding setup when the source-distortion pair is not successively refinable.**

***Index Terms*—Successive refinement, Common information, Rate-Distortion**

I. Introduction

From the view point of rate-distortion (RD) theory, scalable coding has been addressed in the context of successive refinement of information [1], [2]. The problem was originally formulated by Equitz and Cover in [1] as a special case of the more general problem of multiple descriptions [3] and has since been extensively studied by many information theorists [1], [4], [5], [6]. In the conventional successive refinement framework, the encoder generates two layers of information called the base layer (at rate R_{12}) and the enhancement layer (at rate R_2, see Fig. 1a). The base layer provides a coarse reconstruction of the source (at rate R_{12}, achieving distortion D_1), while the enhancement layer is used to 'refine' the reconstruction beyond the base layer (at a net enhancement rate of $R_2 + R_{12}$, achieving distortion $D_2 < D_1$). It is interesting to observe the inherent conflict between maintaining optimality at the two layers. On the one hand the system could be designed to maintain RD optimality at the base layer, i.e., $R_{12} = R(D_1)$ (where $R(\cdot)$ denotes the RD function) and then seek the minimum R_2 that achieves D_2. On the other hand the system could be designed to maintain optimality at the enhancement layer, i.e., $R_2+R_{12} = R(D_2)$ and minimize R_{12} to achieve D_1. Rimoldi [2] derived the complete achievable RD region for this setup and Equitz and Cover [1] established the conditions for a source-distortion pair to achieve asymptotic RD optimality at both the layers simultaneously. Such source-distortion pairs are called *successively refinable* in the literature. Specifically, they showed that, any source is successively refinable for a given distortion measure at distortions D_1 and D_2 ($D_2 < D_1$) if and only if there exists a conditional probability distribution $P(\hat{X}_1, \hat{X}_2|X))$ satisfying $E\left\{d(X, \hat{X}_1)\right\} \leq D_1,\ E\left\{d(X, \hat{X}_2)\right\} \leq D_2$, such that:

$$I(X, \hat{X}_1) = R(D_1), \qquad I(X, \hat{X}_2) = R(D_2) \tag{1}$$

$$X \leftrightarrow \hat{X}_2 \leftrightarrow \hat{X}_1 \tag{2}$$

where $X \leftrightarrow Y \leftrightarrow Z$ denotes that X, Y, Z form a Markov chain in that order. They also showed that Gaussian and Laplacian sources are asymptotically successively refinable under mean squared and absolute error distortion metrics respectively.

However, when a source-distortion pair is not successively refinable, it is impossible to maintain asymptotic optimality at both the layers simultaneously in the scalable coding framework. Inevitably, one of the two layers is forced to carry excess information than necessary to achieve the distortion at that layer. Unfortunately, in several practical applications, source-distortion pairs are not successively refinable, see eg. [7], [6], [8], and due to this inherent and significant shortcoming of the scalable coding framework, most practical system designers or content provides resort to maintaining two independent copies of the source, one at rate $R(D_1)$ and the other at $R(D_2)$, without recourse to scalable coding. We see that, even in this over-simplified scenario with just two layers of information, the implication of such an approach is the need to (almost) double the data center resources. Next, consider broadcast of a real-time multimedia event over a network

The work was supported by the NSF under grants CCF - 1016861 and CCF-1118075

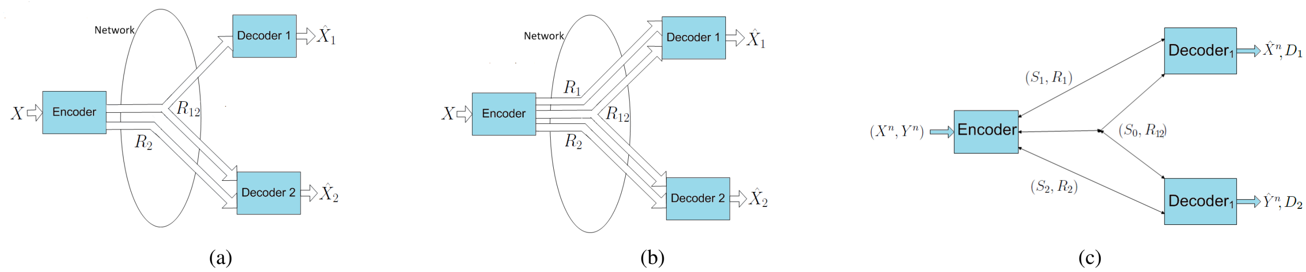

Figure 1: (a) The conventional scalable coding framework : The encoder generates two bit streams. First one at rate R_{12} is sent to both the decoders, while the second one at rate R_2 is only sent to the second decoder. (b) Proposed scalable coding framework: The encoder generates three bit streams. Two bit streams at rates R_{12} and R_2 are sent similar to the conventional scalable coding framework. However a third bit-stream at rate R_1 is sent individually to the first decoder (c) Lossy Gray-Wyner network

to multiple users with differing quality requirements. In this scenario, transmitting independent representations of source optimized at each quality imposes additional strain on the network where many intermediate links must carry multiple versions of the signal. We note that the inherent rate loss in the current successive refinement framework can be upper bounded by $\frac{1}{2}$ bit/sample for any source under mean squared error distortion metric as shown in [9]. This $\frac{1}{2}$ bit in itself represents a considerable loss at low bit rates, while several practical applications use more complex distortion measures for which this rate loss is significantly higher. For eg., in audio compression, the most widely used distortion metric is the weighted mean squared error and the inherent loss due to scalable coding for audio compression has been demonstrated in [7].

In this paper, we focus on a new encoding paradigm for optimal joint layered coding at multiple quality levels. The primary drawback of the conventional scalable coder is that it enforces a rigid hierarchical structure on the bit-stream of different layers. This structure is inspired by the classical communication 'broadcast' scenario wherein the transmitter simply broadcasts all the bits and receivers manage to decode a subset of the (more significant) bits, depending on their channel conditions. The natural assumption underlying this scalable coding framework is that a user with a better channel must always decode all the base layer bits which were necessary for the base layer reconstruction. However, it is worthwhile to question this assumption in today's network scenario wherein an intermediate node forwards packets to multiple receivers. This routing can be effectively leveraged given modern communication protocols that allow entities to exchange information about relevant network conditions. Thus, as a fundamental shift in strategy, we propose to enable routing to a high-end receiver *only a subset of the bits* that are sent to the base receiver as shown in Fig. 1b. We will show that this relaxation of the hierarchical structure provides enough freedom for a well optimized design to achieve best attainable quality at both the receivers, yet significantly reduces the total transmission and storage rates.

Our objective in this paper is to establish an information theoretic characterization for the minimum transmit rate, denoted by $R^*_{sum}(D_1, D_2)$, for the setup shown in Fig. 1b when the two receivers are constrained to achieve their respective RD functions. We formally state the new framework in section II and provide an information theoretic characterization for $R^*_{sum}(D_1, D_2)$ in section III. We then show using an example that the proposed common layer framework provides a strict reduction in total transmit rate compared to conventional scalable coding when the source-distortion pair is not successively refinable. We will then see in section IV that this problem is closely related to the concept of 'common information' (CI) of two dependent random variables [10], [11], [12], [13] which provides further insights on the workings of the proposed framework.

II. Proposed Framework

It is obvious that, when a source is not successively refinable, the information required to achieve D_1 is not a proper subset of the information required to achieve D_2, or alternatively, part of the information sent to achieve D_1, is not useful in decoding at distortion D_2. However, it is also obvious that there is considerable overlap in information required to achieve reconstructions at the two distortion levels, and hence transmitting/storing two individual representations of the source is clearly wasteful. This is precisely the theoretical intuition which we exploit in the proposed framework, where a part of the information that is sent to a receiver that achieves D_1 will not be sent to a receiver reconstructing at lower distortion D_2. Hence, the encoder generates 3 different packets. One at rate R_1, which is only sent to the receiver reconstructing at D_1 and a second packet at rate R_2 is sent only to the decoder reconstructing at D_2. A third packet at rate R_{12} is sent to both the decoders. The two decoders receive at rates of $R_1 + R_{12}$ and $R_2 + R_{12}$, respectively. This setup is shown schematically in Fig. 1b. Observe that conventional scalable coding is obtained as a degenerate special case when R_1 is set to 0, and hence the proposed framework performs at least as well as conventional scalable coding. However, when the source-distortion pair is not successively refinable and scalable coding suffers inherent loss, this framework provides an extra degree of freedom achieving RD optimality at both the layers, yet exploiting the dependencies across the layers to minimize the total transmission rate.

The complete set of RD tuples, consisting of all achievable $(R_1, R_2, R_{12}, D_1, D_2)$ for the proposed framework, denoted hereafter by $\mathcal{RD}$, can be derived easily from the results of Gray and Wyner [12]. We will return to this characterization in the following section. Our primary objective in this paper is to establish a single letter information theoretic characterization for the minimum sum rate (total transmit rate at the encoder) when both the decoders are constrained to achieve their respective RD functions. Towards deriving this, we define the quantity $R^*(D_1, D_2)$ by,

$$R^*(D_1, D_2) = \sup R_{12} : (R_1, R_2, R_{12}, D_1, D_2) \in \mathcal{RD}, \\ R_1 + R_{12} = R(D_1),\ \ R_2 + R_{12} = R(D_2) \quad (3)$$

Observe that the minimum sum rate when the decoders are constrained to receive at their respective RD functions is equal to $R^*_{sum}(D_1, D_2) = R(D_1) + R(D_2) - R^*(D_1, D_2)$. Note that in the conventional scalable coding framework, if the source is not successively refinable, the only way to achieve RD optimality at both the layers is to store two individual representations, in which case, $R_1 = R(D_1)$, $R_2 = R(D_2)$ and $R_{12} = 0$. Hence if $R^*(D_1, D_2)$ is strictly greater than 0, the proposed framework can achieve a sum rate strictly lower than conventional scalable coding, yet achieving RD optimality at both the decoders. Ofcourse, for a successively refinable source-distortion pair, $R^*(D_1, D_2) = R(D_1)$.

III. Main Results

A. Gray-Wyner Region

Gray and Wyner considered the network shown in Fig. 1c where the encoder observes n iid versions of two correlated random variables (X, Y) and generates three bit streams which are routed to the respective decoders as shown. The two decoders reconstruct X^n and Y^n within distortions D_1 and D_2 respectively based on some well defined single letter distortion measures $d_X(\cdot,\cdot)$ and $d_Y(\cdot,\cdot)$. An achievable rate-distortion tuple is defined in the standard information theoretic sense. The convex closure of all achievable rate-distortion tuples for the setup shown in Fig. 1c is denoted here by $\mathcal{RD}_{GW}$.

Gray and Wyner [12] gave the complete characterization for $\mathcal{RD}_{GW}$. Let $(U, \hat{X}, \hat{Y})$ be any random variables jointly distributed with (X, Y) and taking values over alphabets $\mathcal{U}, \hat{\mathcal{X}}$ and $\hat{\mathcal{Y}}$ respectively for some arbitrary $\mathcal{U}$ and where $\hat{\mathcal{X}}$ and $\hat{\mathcal{Y}}$ are the respective reconstruction alphabets. Let the joint density, $P(X, Y, U, \hat{X}, \hat{Y})$, be such that $E(d_X(X, \hat{X})) \leq D_1$ and $E(d_Y(Y, \hat{Y})) \leq D_2$ hold. Then all RD tuples $(R_{12}, R_1, R_2, D_1, D_2)$ satisfying the following conditions are achievable:

$$R_{12} \geq I(X, Y; U) \\ R_1 \geq I(X; \hat{X}|U),\ \ R_2 \geq I(Y; \hat{Y}|U) \quad (4)$$

The closure of the achievable RD tuples over all such joint densities is the complete RD region for the Gray-Wyner network. It is important to observe that the proposed framework for successive refinement is in fact a special case of the Gray-Wyner network with $X = Y$. Hence, it immediately follows that the region $\mathcal{RD}$ defined in section II can be characterized using Gray-Wyner's result by setting $X = Y$ and $d_X(\cdot,\cdot) = d_Y(\cdot,\cdot) = d(\cdot,\cdot)$. We next derive an information theoretic characterization for $R^*(D_1, D_2)$.

B. Characterization of $R^*(D_1, D_2)$

Recall the definition of $R^*(D_1, D_2)$ defined in section II. The following theorem provides an information theoretic characterization.

Theorem 1. *An information theoretic characterization for $R^*(D_1, D_2)$ is given by:*

$$R^*(D_1, D_2) = \sup I(X; U) \quad (5)$$

where the supremum is over all conditional densities $P(U, \hat{X}_1, \hat{X}_2|X)$ such that, $P(\hat{X}_1|X)$ and $P(\hat{X}_2|X)$ achieve RD-optimality at D_1 and D_2 respectively and for which the following two Markov chains hold:

$$X \leftrightarrow \hat{X}_1 \leftrightarrow U, \qquad X \leftrightarrow \hat{X}_2 \leftrightarrow U \quad (6)$$

Remark 1. Note that a source is successively refinable if and only if $R^* = R(D_1)$. From Theorem 1, it is easy to verify that $R^* = R(D_1)$ if an only if Equitz and Cover's conditions given by (2) are satisfied.

Proof: We begin by assuming that there exist unique channels $P(\hat{X}_1|X)$ and $P(\hat{X}_2|X)$ that achieve RD optimality at D_1 and D_2 respectively. The proof follows in very similar lines to the case when there are multiple RD optimal channels.

We will first characterize the set of all points in $\mathcal{RD}$ which lie on both the planes $R_{12} + R_1 = R(D_1)$ and $R_{12} + R_2 = R(D_2)$. Let us denote by $\mathcal{R}(D_1, D_2)$ the set of all rate tuples that are achievable for given distortions (D_1, D_2), i.e., $\mathcal{R}(D_1, D_2) = \{(R_{12}, R_1, R_2) : (R_{12}, R_1, R_2, D_1, D_2) \in \mathcal{RD}\}$. It follows from the Gray-Wyner theorem that for every point (R_{12}, R_1, R_2) in $\mathcal{RD}(D_1, D_2)$, we can find random variables $(U, \hat{X}_1, \hat{X}_2)$ satisfying $E(d(X, \hat{X}_1)) \leq D_1$ and $E(d(X, \hat{X}_2)) \leq D_2$ such that:

$$R_{12} \geq I(X; U) \\ R_1 \geq I(X; \hat{X}_1|U),\ \ R_2 \geq I(X; \hat{X}_2|U) \quad (7)$$

We further require these points to satisfy $R_{21} + R_1 = R(D_1)$ and $R_{12} + R_2 = R(D_2)$. Hence, we have the following series of inequalities:

$$\begin{aligned} R(D_1) &= R_{12} + R_1 \geq I(X; U) + I(X; \hat{X}_1|U) \\ &= I(X; U, \hat{X}_1) \geq I(X; \hat{X}_1) \geq R(D_1) \end{aligned} \quad (8)$$

and similar chain of inequalities for $R(D_2)$. As the LHS and the RHS of the above series of inequalities are the same, all the inequalities must be equalities leading to:

$$I(X; \hat{X}_1) = R(D_1) \qquad I(X; \hat{X}_2) = R(D_2) \\ I(X; U, \hat{X}_1) = I(X; \hat{X}_1) \qquad I(X; U, \hat{X}_2) = I(X; \hat{X}_2) \quad (9)$$

i.e, every point which lies on both the planes $R_{12} + R_1 = R(D_1)$ and $R_{12} + R_2 = R(D_2)$ and is part of $\mathcal{R}(D_1, D_2)$ must be achieved by a joint density which satisfies (9). Clearly, every joint density that satisfies (9) leads to an achievable point in $\mathcal{R}(D_1, D_2)$ which satisfies both the constraints $R_{12} + R_1 = R(D_1)$ and $R_{12} + R_2 = R(D_2)$. Hence, defining a convex closure of all achievable points over all joint densities satisfying (6) leads to every point in the intersection of $R_{12} + R_1 = R(D_1)$, $R_{12} + R_2 = R(D_2)$ and $\mathcal{R}(D_1, D_2)$. Therefore the supremum of R_{12} (recall the definition of $R^*(D_1, D_2)$) is achieved by taking a supremum of $I(X;U)$ over all densities satisfying (6). ■

C. Example

In this section, we show that $R^*(D_1, D_2) > 0$ for the counter-example considered by Equitz and Cover in [1], which was used to prove that there exist source-distortion pairs that are not successively refinable. The example is described as follows. Consider a discrete source X over alphabet $\mathcal{X} = \{0, 1, 2\}$ with PMF $P_X = [\frac{1-P}{2}, P, \frac{1-P}{2}]$. Let the reconstruction alphabet be $\hat{\mathcal{X}} = \{0, 1, 2\}$ and the distortion measure be defined by the absolute difference, i.e. $|X - \hat{X}|$. Let $z = e^{R^{'}(D)}$. Then for $P < 3 - 2\sqrt{2}$, it can be shown that the RD curve has three regions, denoted by $\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3$. In the first and the third regions, i.e., $\forall D \in \mathcal{D}_1, \mathcal{D}_3$, the RD optimal channel, $P(\hat{X}|X)$, has reconstruction alphabet of cardinality three. However, when $D \in \mathcal{D}_2$ the cardinality of the optimal reconstruction alphabet is just two. Equitz and Cover showed that when we pick $D_2 \in \mathcal{D}_2$ and $D_1 \in \mathcal{D}_3$ $(D_2 < D_1)$, it is impossible to find a joint density $P(\hat{X}_1, \hat{X}_2|X)$ with $X \leftrightarrow \hat{X}_2 \leftrightarrow \hat{X}_1$ for which the PMFs $P(\hat{X}_2|X)$ and $P(\hat{X}_1|X)$ achieve RD optimality at D_1 and D_2 respectively. The expressions for $P(X|\hat{X})$ in each of the three regions is given in [1].

We will next show that we can always find a joint density $P(U, \hat{X}_1, \hat{X}_2|X)$ with $I(X;U) > 0$ such that $X \leftrightarrow \hat{X}_1 \leftrightarrow U$ and $X \leftrightarrow \hat{X}_2 \leftrightarrow U$ hold and for which the the PMFs $P(\hat{X}_2|X)$ and $P(\hat{X}_1|X)$ are the RD optimal channels and distortions D_1 and D_2 respectively where $D_2 \in \mathcal{D}_2$ and $D_1 \in \mathcal{D}_3$ $(D_2 < D_1)$. Towards finding such a joint distribution, note that it is sufficient for us to find conditional PMFs $P(U|\hat{X}_1)$ and $P(U|\hat{X}_2)$ which satisfy the following conditions for some distribution $P(U|X)$:

$$\begin{aligned} P(U|X) &= \sum_{\hat{\mathcal{X}}_1} P(U|\hat{X}_1)P(\hat{X}_1|X) \\ &= \sum_{\hat{\mathcal{X}}_2} P(U|\hat{X}_2)P(\hat{X}_2|X) \end{aligned} \tag{10}$$

where $P(\hat{X}_1|X)$ and $P(\hat{X}_2|X)$ achieve optimality at D_1 and D_2 respectively. Once we find such conditional PMFs, we can always generate the density $P(U, \hat{X}_1, \hat{X}_2|X)$ as follows:

$$P(U, \hat{X}_1, \hat{X}_2|X) = P(U|X)P(\hat{X}_1|X, U)P(\hat{X}_2|X, U) \tag{11}$$

where $P(\hat{X}_1|X, U)$ and $P(\hat{X}_2|X, U)$ are the conditional densities induced by the Markov chain structure in (10). Then $P(U, \hat{X}_1, \hat{X}_2|X)$ satisfies all the required conditions in Theorem 1 and hence $R^*(D_1, D_2) \geq I(X;U) > 0$. Let us pick U to be a binary symmetric random variable taking values over the alphabet $\{0, 2\}$ with equal probability. Recall that $\hat{X}_2$ takes values over just two reconstruction alphabets, while $\hat{X}_1$ takes values over an alphabet of cardinality three. Let us choose $P(U|\hat{X}_1)$ and $P(U|\hat{X}_2)$ as follows. The channel $P(U|\hat{X}_2)$ is a binary symmetric channel with cross over probability Q and the transition probabilities $P(U|\hat{X}_1)$ are given by:

$$\begin{aligned} P(U = 0|\hat{X}_1 = 0) &= P(U = 2|\hat{X}_1 = 2) = 1 \\ P(U = 0|\hat{X}_1 = 1) &= P(U = 2|\hat{X}_1 = 1) = \tfrac{1}{2} \end{aligned} \tag{12}$$

Q is set to ensure that (10) is satisfied. Let us denote by $z_i = e^{R'(D_i)}$ $i = \{1, 2\}$. Then it is easy to verify that (12) induces the following conditional distribution for $P(X|U)$:

$$P(X|U) = \begin{bmatrix} \frac{1-z_1-P}{1-z_1} & P & \frac{Pz_1}{1-z_1} \\ \frac{Pz_1}{1-z_1} & P & \frac{1-z_1-P}{1-z_1} \end{bmatrix} \tag{13}$$

Then by using (10), we have:

$$Q = \frac{z_1 z_2^2 + P z_2^2 + P z_1 - z_2^2}{(1 - z_1)(1 - P)(1 - z_2^2)} \tag{14}$$

It is easy to check that for all $D_2 \in \mathcal{D}_2$ and $D_1 \in \mathcal{D}_3$, we have $0 < Q < \frac{1}{2}$ and $I(X;U) > 0$. In fact, for this example, it can also be shown that the above construction achieves the maximum $I(X;U)$ in Theorem 1. However, we omit the details here due to space constraints.

IV. Relations to Common Information

A. Gács-Körner's Common Information

Gács and Körner [11] defined CI of X and Y as the maximum rate of the codeword that can be generated individually at two encoders observing X^n and Y^n separately. Gács-Körner's original definition of CI was predated and naturally unrelated to the Gray-Wyner network. However, an alternate and insightful characterization of $C_{GK}(X, Y)$ was given by Ahlswede and Körner [10] in terms of $\mathcal{RD}_{GW}$ as follows:

$$C_{GK}(X, Y) = \sup R_{12} : \bar{\mathbf{R}} \in \mathcal{R}_{GW}(0, 0) \tag{15}$$

subject to,

$$R_{12} + R_1 = H(X), \qquad R_{12} + R_2 = H(Y) \tag{16}$$

where $\bar{\mathbf{R}} = (R_{12}, R_1, R_2)$ and $\mathcal{R}_{GW}(D_1, D_2)$ denotes the cross-section of $\mathcal{RD}_{GW}$ at distortions (D_1, D_2). Although the original definition of Gács-Körner's CI does not have a direct lossy interpretation, the alternate definition given by Ahlswede and Körner in terms of the lossless Gray-Wyner region can be extended to the lossy setting. The lossy generalization of Gács-Körner's CI at (D_1, D_2) is defined as:

$$C_{GK}(X, Y; D_1, D_2) = \sup R_{12} : \bar{\mathbf{R}} \in \mathcal{R}_{GW}(D_1, D_2) \tag{17}$$

subject to,

$$R_{12} + R_1 = R_X(D_1), \qquad R_{12} + R_2 = R_Y(D_2) \tag{18}$$

where $R_X(\cdot)$ and $R_Y(\cdot)$ denote the respective rate-distortion functions. A single letter characterization for $C_{GK}(X,Y;D_1,D_2)$ was recently derived in [14]. It is very interesting to observe that $R^*(D_1,D_2)$ is obtained as an extreme special case when we set $X=Y$ in $C_{GK}(X,Y;D_1,D_2)$. Essentially, the maximum shared rate in the proposed framework is a measure of the 'amount of common information' between the bitstreams required to decode the source at distortions D_1 and D_2 respectively. What makes it particularly surprising is the fact that, if X and Y are correlated, it can be shown following the footsteps of Gács and Körner in lossless setting that $C_{GK}(X,Y;D_1,D_2)$ is typically very small $\forall D_1,D_2\geq 0$ and depends only on the zeros of the joint distribution of (X,Y). However, when X is set equal to Y, $R^*(D_1,D_2)$ is usually strictly greater than zero leading to important practical implications in scalable coding.

B. Wyner's Common Information

An alternate definition of CI was given by Wyner, which is stated as the minimum rate on the shared branch of the lossless Gray-Wyner network when the sum rate is constrained to be the joint entropy, i.e.,

$$\begin{aligned} C_W(X,Y) &= \inf R_{12} : \bar{\mathbf{R}} \in \mathcal{R}_{GW}(0,0) \\ & \quad R_{12}+R_1+R_2 = H(X,Y) \end{aligned} \tag{19}$$

Wyner showed that $C_W(X,Y)=\inf I(X,Y;U)$ where the infimum is over all joint densities $P(U|X,Y)$ which satisfy $X\leftrightarrow U\leftrightarrow Y$. It is of both theoretical and practical interest to understand the potential implications of Wyner's definition in the successive refinement framework that we consider in this paper. To clarify the underlying relations, we define a lossy extension of Wyner's CI analogous to the Gács-Körner's extension as follows:

$$\begin{aligned} C_W(X,Y;D_1,D_2) &= \inf R_{12} : \bar{\mathbf{R}} \in \mathcal{R}_{GW}(D_1,D_2) \\ & \quad R_{12}+R_1+R_2 = R_{X,Y}(D_1,D_2) \end{aligned} \tag{20}$$

An information theoretic characterization of $C_W(X,Y;D_1,D_2)$ was also recently derived in [14]. We omit restating it here. However, what we are particularly interested in here is to understand the physical interpretation of this quantity when we set $X=Y$, i.e., for the setup shown in Fig. 1b, we are interested in characterizing the minimum shared rate when the sum rate is set to its minimum. Observe that as $X=Y$, the minimum sum rate is, in fact, equal to $R(D_2)$ (as $D_2<D_1$). Hence the quantity $C_W(X,X;D_1,D_2)$ corresponds to the minimum shared rate when the sum rate is equal to $R(D_2)$. However, $R(D_2)$ is the minimum rate at which the second decoder must receive information for it to decode X at distortion D_2. Therefore it follows that R_1 must be equal to zero and the framework degenerates to the conventional scalable coding setting. Hence the quantity $C_W(X,X;D_1,D_2)$ corresponds to the minimum rate for the base layer when the enhancement layer is set to receive information at the RD function in the context of conventional successive refinement. It can be shown that:

$$C_W(X,X;D_1,D_2) = \inf I(X;\hat{X}_1) \tag{21}$$

where the infimum is over all joint densities $P(\hat{X}_1,\hat{X}_2|X)$ such that $P(\hat{X}_2|X)$ achieves RD optimality at D_2 and $P(\hat{X}_1|X,\hat{X}_2)$ is such that the following conditions holds:

$$X\leftrightarrow \hat{X}_2 \leftrightarrow \hat{X}_1, \qquad E(d(X,\hat{X}_1))\leq D_1 \tag{22}$$

V. Conclusion

In this paper, inspired by the inherent drawback with the conventional scalable coding framework, we introduced a new encoding paradigm for successive refinement of information. Unlike the conventional scalable coding framework, in the proposed setting, a subset of the information sent to the base layer decoder is not routed to the enhancement layer decoder. What makes the proposed framework particularly interesting is the fact that, unlike conventional scalable coding, in the context of non-successively refinable sources both the decoders can receive information at their respective RD functions and yet the sum rate of transmission can be significantly lower than the sum of the individual RD functions. We showed the close relation of the proposed framework with the concept of CI of two dependent random variables and derived an information theoretic characterization for the minimum sum rate achievable when the two decoders receive information at their respective RD functions. We also demonstrated using an example, the potential gains of the proposed paradigm when the source-distortion pair is not successively refinable.

References

[1] W.H.R. Equitz and T.M. Cover, "Successive refinement of information," *Information Theory, IEEE Transactions on*, vol. 37, 1991.

[2] B. Rimoldi, "Successive refinement of information: characterization of the achievable rates," *Information Theory, IEEE Transactions on*, vol. 40, 1994.

[3] A.E. Gamal and T. Cover, "Achievable rates for multiple descriptions," *Information Theory, IEEE Transactions on*, vol. 28, 1982.

[4] E. Tuncel and K. Rose, "Additive successive refinement," *Information Theory, IEEE Transactions on*, vol. 49, 2003.

[5] R. Venkataramani, G. Kramer, and V.K. Goyal, "Multiple description coding with many channels," *Information Theory, IEEE Transactions on*, vol. 49, 2003.

[6] J. Chow and T. Berger, "Failure of successive refinement for symmetric gaussian mixtures," *Information Theory, IEEE Transactions on*, vol. 43, 1997.

[7] A. Aggarwal, S.L. Regunathan, and K. Rose, "Efficient bit-rate scalability for weighted squared error optimization in audio coding," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 14, 2006.

[8] F. Wu, S. Li, and Y.-Q. Zhang, "A framework for efficient progressive fine granularity scalable video coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 11, 2001.

[9] L. Lastras and T.; Berger, "All sources are nearly successively refinable," *Information Theory, IEEE Transactions on*, vol. 47, 2001.

[10] R. Ahlswede and J. Körner, "On common information and related characteristics of correlated information sources," in *Preprint. Presented at the 7th Prague Conference on Information Theory*, 1974.

[11] P. Gács and J. Körner, "Common information is far less than mutual information," *Problems of Control and Information Theory*, 1973.

[12] R. M. Gray and A. D. Wyner, "Source coding over simple networks," *The Bell Systems Technical Journal*, vol. 53, 1974.

[13] A. Wyner, "The common information of two dependent random variables," *Information Theory, IEEE Transactions on*, vol. 21, 1975.

[14] K. Viswanatha, E. Akyol, and K. Rose, "Lossy common information of two dependent random variables," in *Information Theory, IEEE International Symposium on (ISIT)*, Jul 2012.