arXiv:0901.1705v1 [cs.IT] 13 Jan 2009

Rate-Distortion with Side-Information at Many Receivers

Roy Timo, Terence Chan and Alexander Grant

Abstract

We present an inner bound for the admissible rate region of the t-stage successive refinement problem with side information, and we present an upper bound for the rate-distortion function for lossy source coding with multiple receivers and side information. A single-letter characterisation of this rate-distortion function is a long standing open problem in multi-terminal information theory, and it is widely believed that the tightest upper bound is provided by Theorem 2 of Heegard and Berger's paper "Rate-Distortion when Side Information may be Absent," *IEEE Trans. Inform. Theory*, 1985. We give a counterexample to Heegard and Berger's result, and we develop our new upper bound as a corollary to our inner bound for the successive refinement problem with side information.

Index Terms

Source coding with receiver side information, rate-distortion function, successive refinement, complementary delivery, side information scalable source coding.

R. Timo, T. Chan and A. Grant are with the Institute for Telecommunications Research, University of South Australia. Email: {roy.timo, terence.chan, alex.grant}@unisa.edu.au. This work was funded by the Australian Research Council Grant DP0880223.

I. INTRODUCTION

One of the most important results in multi-terminal information theory is Wyner and Ziv's solution [1] to the problem of lossy source coding with side information at the receiver; figure 1 shows the problem setup. The main objective is to give a single-letter characterisation [2, Page 259] of the rate-distortion function $R(d)$, which is defined as the smallest rate at which it is possible to encode a discrete memoryless source $X^n = X_1, \ldots, X_n$ such that the receiver with side information $Y^n = Y_1, \ldots, Y_n$ can obtain a reconstruction $\hat{X}^n$ of X^n with an average per-letter distortion less than d. To this end, Wyner and Ziv [1, Theorem 1] showed that

$$R(d) = \min \big\{ I(X;U) - I(U;Y) \big\} \ , \tag{1}$$

where the minimization is over all choices of a discrete finite alphabet auxiliary random variable U such that: (1) $Y \ominus X \ominus U$ forms a Markov chain, and (2) there exists a deterministic function $\hat{X}(U,Y)$ with an expected distortion less than d. In this paper we study two extensions of this problem with multiple receivers.

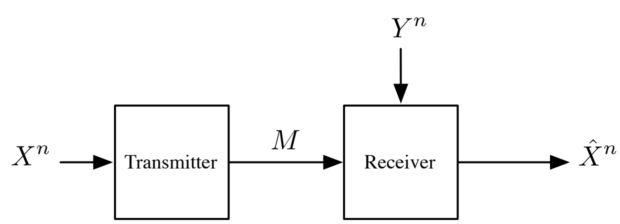

Fig. 1. Lossy source coding with side information at the receiver.

If the side information Y^n in Wyner and Ziv's problem becomes unreliable in the sense that it may, or may not, be available to the receiver, then the coding scheme [1, Section IV] used to prove (1) fails: a more complex coding scheme is required to exploit Y^n. This observation independently inspired Kaspi [3] in 1980 (published by Wyner on behalf of Kaspi in 1994) as well as Heegard and Berger [4] in 1985 to consider problem shown in Figure 2 – the so called Kaspi/Heegard-Berger problem. As before, the objective is to find the smallest rate $R(d_1, d_2)$ such that receiver 1 resp. 2 can find reconstructions with average per-letter distortions d_1 resp. d_2. Heegard and Berger[1] showed that [4, Theorem 1]

$$R(d_1, d_2) = \min \Big\{ I\left(X;W\right) + I\left(X;U \mid Y,W\right) \Big\} \ ,$$

[1]Kaspi's result [3, Theorem 2] gives an alternative characterisation of $R(d_1, d_2)$.

where the minimization is over all choices of two discrete finite alphabet auxiliary random variables U and W such that: (1) $Y \multimap\!\!- X \multimap\!\!- (U,W)$ forms a Markov chain, and (2) there exists functions $\hat{X}_1(Y,U,W)$ and $\hat{X}_2(W)$ with expected distortions bound by d_1 and d_2 respectively.

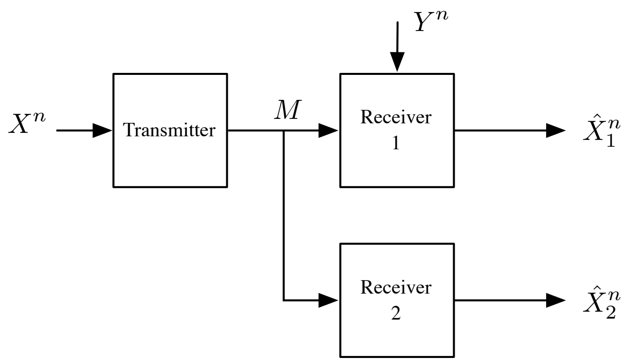

Fig. 2. Lossy source coding when side information may be absent at the receiver.

The Kaspi/Heegard-Berger problem was further generalised by Heegard and Berger in [4, Section VII] to the problem shown in Figure 3. There are t receivers (each with their own side information) and the objective is to characterise the corresponding rate-distortion function $R(d_1, d_2, \ldots, d_t)$. Today, a single-letter characterisation of $R(d_1, d_2, \ldots, d_t)$ is still lacking; since its formulation in 1985, Heegard and Berger's problem has resisted final solution and is now regarded as a classic in multi-terminal information theory. Notwithstanding this difficulty, the problem has stimulated a number of important results over the past two decades [3], [5]–[9], and it has been solved for the special case of degraded side information $X \multimap\!\!- Y_{\{t\}} \multimap\!\!- Y_{\{t-1\}} \multimap\!\!- \cdots \multimap\!\!- Y_{\{1\}}$ [4, Theorem 3].

For arbitrarily correlated side information, Heegard and Berger presented the function $R_{HB}(d_1, d_2, \ldots, d_t)$ in [4, Theorem 2] as an upper bound for $R(d_1, d_2, \ldots, d_t)$. (The expression for $R_{HB}(d_1, d_2, \ldots, d_t)$ follows in (14); however, this expression requires the notation and definitions from Section II.) This function is widely believed to be the tightest upper bound.

The present paper was motivated by our discovery of a counterexample to [4, Theorem 2]. That is, a situation where the claimed upper bound $R_{HB}(d_1, d_2, \ldots, d_t)$ is strictly less than the rate distortion function $R(d_1, d_2, \ldots, d_t)$. The invalidity of $R_{HB}(d_1, d_2, \ldots, d_t)$ as an upper bound for $R(d_1, d_2, \ldots, d_t)$ is by no means obvious. Despite being used with modest frequency in the literature, it appears to have gone unnoticed for more than two decades. The claim is based on a complex random coding argument that uses $2^t - 1$ individual descriptions (via $2^t - 1$ auxiliary random variables) to convey information about X^n to the receivers. We will see, however, that the expression for $R_{HB}(d_1, d_2, \ldots, d_t)$ does not provide appropriate conditional independence between certain auxiliary random variables; thus, there is insufficient rate for each of the $2^t - 1$ descriptions to be reliably decoded at the receivers.

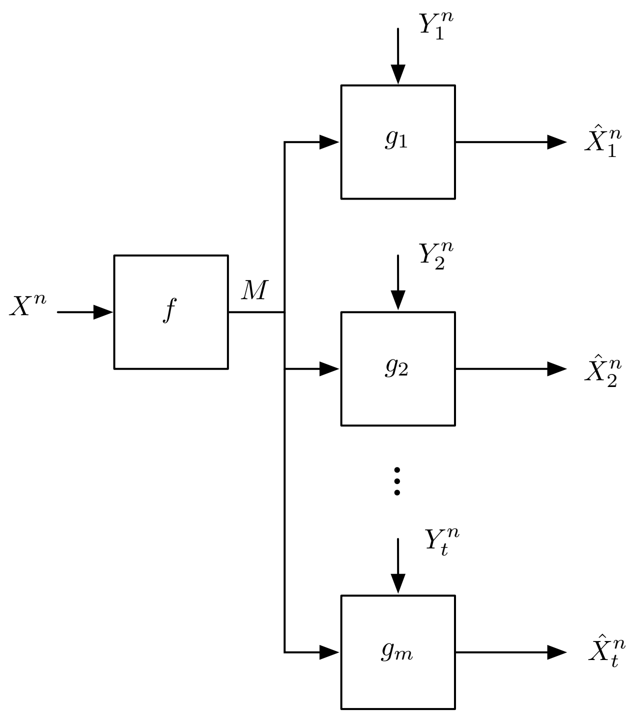

Fig. 3. Lossy source coding with t receivers – each with arbitrary side information.

This observation led us to consider the generalisation of Heegard and Berger's problem shown in Figure 4. The transmitter encodes the source into t messages $M_1, M_2, \ldots, M_t$. Receiver j receives messages M_1 through M_j and forms a reconstruction $\hat{X}_j$ with average per-letter distortion less than d_j. It is readily seen that this generalisation of Heegard and Berger's problem is a multi-stage version of the successive refinement problem with side information [6], [8], [9], and for this reason we refer to it as a successive refinement problem[2].

Steinberg and Merhav [6] introduced and solved the two-receiver successive refinement problem with degraded side information $X \ominus Y_2 \ominus Y_1$, and Tian and Diggavi [9] extended this solution to t-receivers

[2]In this paper we shall be exclusively interested in the characterisation of an inner bound for the region of admissible rate tuples. We will not require, or even define, any notion of successive refinability of the source. For such details, the interested reader is directed to [6], [8], [9].

with degraded side information $X \ominus Y_t \ominus \cdots \ominus Y_2 \ominus Y_1$. More recently, Tian and Diggavi [8] gave inner and outer bounds for the admissible rate region for two receivers assuming $X \ominus Y_1 \ominus Y_2$ forms a Markov chain – a reverse of the degradedness $X \ominus Y_2 \ominus Y_1$ used in [6], [9]. Our main result is a coding theorem for the t-stage successive refinement problem with arbitrarily correlated side information shown in Figure 4. An immediate corollary of this theorem is an upper bound for the rate-distortion function $R(d_1, d_2, \ldots, d_t)$ for Heegard and Berger's problem shown in Figure 3.

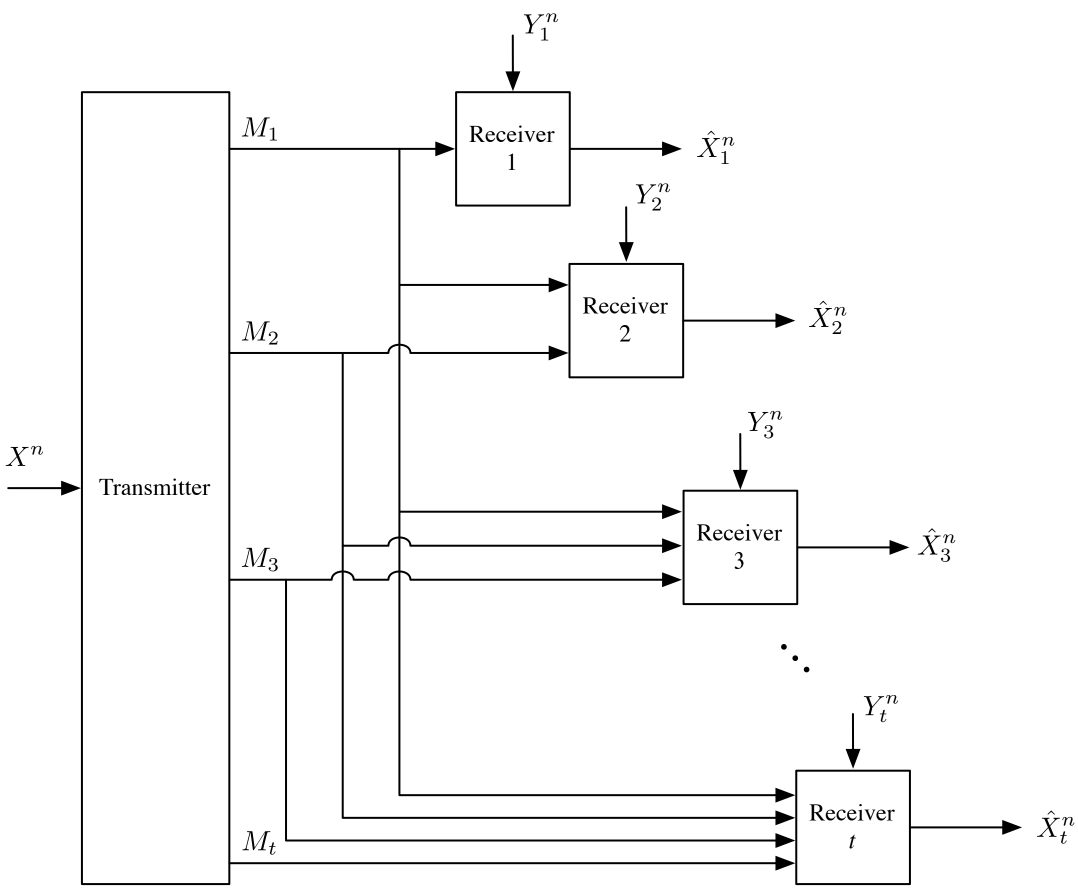

Fig. 4. Successive Refinement with t stages and side information.

An outline of the remainder of this paper is as follows. In Section II we formally define the t-receiver successive refinement problem shown in Figure 4; we present an inner bound for the admissible rate region in Theorem 1; and we show that this inner bound includes the coding theorems of Steinberg and Merhav [6] as well as Tian and Diggavi [8], [9] as special cases. Our proof of Theorem 1 is given in Section III. In Section IV we formally define Heegard and Berger's problem shown in Figure 3; we show that there exists a situation where $R_{HB}(d_1, d_2, \ldots, d_t) < R(d_1, d_2, \ldots, d_t)$; we present a new upper bound for $R(d_1, d_2, \ldots, d_t)$ in Corollary 1; and we show that this bound includes the coding theorems of Wyner and Ziv [1], Heegard and Berger [4], and Kimura and Uyematsu [10] as a special case. In Section V we describe a new lossless source coding problem and present an achievable rate. Finally, the paper is concluded in Section VI.

Notation: Sets will be identified using calligraphic typeface, e.g. $\mathscr{X}$; random variables will be identified by upper case characters[3] e.g. $X \in \mathscr{X}$; and particular realizations of random variables will be identified by lowercase characters e.g. x. Superscripts will be used to denote sequences, e.g. $X^j = X_1, X_2, \ldots, X_j$, similarly $X_i^j = X_i, X_{i+1}, \ldots, X_j$. For any natural number $t \in \mathbb{N}$ we let $[t] = \{1, 2, \ldots, t\}$, and for $s < t$ we let $[s, t] = \{s, s+1, \ldots, t\}$. Set-valued subscripts will serve as indices[4], e.g. $U_{\mathscr{S}}$ with $\mathscr{S} \subset [t]$ denotes the random variable assigned to the subset $\mathscr{S}$. For brevity, singletons or other small sets will be written without brace notation, e.g. $U_{\{1\}}$ and $U_{\{1,2\}}$ will be written as U_1 and U_{12} respectively. Sequences of so-labelled variables will be denoted by $U_{\mathscr{S},i}^j = U_{\mathscr{S},i}, U_{\mathscr{S},i+1}, \ldots, U_{\mathscr{S},j}$. Finally, tuples will be denoted by boldface, e.g. $\mathbf{d} = (d_1, d_2, \ldots, d_t)$.

II. Successive Refinement with Receiver Side Information

We begin with a formal definition of the problem that is shown in Figure 4. Let $\mathscr{X}$ and $\mathscr{Y}_j$ (for all receivers $j \in [t]$) be discrete finite alphabets. We assume that

$$\left(X^n, Y_1^n, Y_2^n, \ldots, Y_t^n\right) \triangleq \left\{\left(X_i, Y_{1,i}, Y_{2,i}, \ldots, Y_{t,i}\right)\right\}_{i=1}^n$$

are n independent and identically distributed (i.i.d.) tuples of random variables emitted by a discrete memoryless source $(\mathscr{X} \times \mathscr{Y}_1 \times \mathscr{Y}_2 \times \cdots \times \mathscr{Y}_t,\ Q)$, where Q is an arbitrary probability mass function on the cartesian product space $\mathscr{X} \times \mathscr{Y}_1 \times \mathscr{Y}_2 \times \cdots \times \mathscr{Y}_t$:

$$Q\left(x, y_1, \ldots, y_t\right) \triangleq \Pr\left[X_1 = x_1, Y_1 = y_1, \ldots, Y_t = y_t\right].$$

The transmitter encodes X^n with an encoder

$$f^{(n)} : \mathscr{X}^n \to \mathscr{M}_1 \times \mathscr{M}_2 \times \cdots \times \mathscr{M}_t\ ,$$

where $\mathscr{M}_j$ is a discrete finite set with $|\mathscr{M}_j|$ elements. The resulting t indices $(M_1,\ M_2,\ \ldots,\ M_t) = f^{(n)}\left(X^n\right)$ are sent to the receivers over channels 1 through t respectively.

[3]With the exception of H and I, which will be respectively reserved for the entropy and mutual information functions as defined in [11]. Similarly, R will be reserved for rate-distortion functions and admissible rates.

[4]Rather than to denote the set $\{U_i, i \in \mathscr{S}\}$ of random variables, which is common usage in the literature.

At the j^{th}-receiver, let $\hat{\mathscr{X}_j}$ be a reconstruction alphabet and $\delta_j : \mathscr{X} \times \hat{\mathscr{X}_j} \to \mathbb{R}_+ \triangleq [0, \infty)$ be a per-letter distortion measure. (The reconstruction alphabet and distortion measure used at each receiver need not be identical.) The j^{th}-receiver is required to generate a reconstruction $\hat{X}_j^n = g_j^{(n)}(M_1, M_2, \ldots, M_j, Y_j^n)$ of X^n using a decoder

$$g_j^{(n)} : \mathscr{M}_1 \times \mathscr{M}_2 \times \cdots \times \mathscr{M}_j \times \mathscr{Y}_j^n \to \hat{\mathscr{X}_j}^n \ ,$$

and the quality of this reconstruction is measured by the average distortion

$$\Delta_j \triangleq \frac{1}{n} \mathbb{E} \sum_{i=1}^{n} \delta_j(X_i, \hat{X}_{j,i}) \ ,$$

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

Definition 1 ($\mathbf{d}$-Admissible Rate Tuple): Suppose $\mathbf{d} = (d_1, d_2, \ldots, d_t) \in \mathbb{R}_+^t$ is an arbitrary distortion tuple. A rate tuple $\mathbf{R} = (R_1, R_2, \ldots, R_t) \in \mathbb{R}_+^t$ is said to be $\mathbf{d}$-admissible if, for arbitrary $\epsilon > 0$, there exists a sufficiently large n, an encoder $f^{(n)}$ and t decoders $g_1^{(n)}$, $g_2^{(n)}$, ..., $g_t^{(n)}$, where $\Delta_j \leq d_j + \epsilon$ and

$$\frac{1}{n} \log_2 |\mathscr{M}_j| \leq R_j + \epsilon$$

for every $j \in [t]$. We let $\mathscr{R}(\mathbf{d})$ denote the closure of the set of all $\mathbf{d}$-admissible rate tuples.

For each $j \in [t]$, let $d_{j,min} = \mathbb{E}[\min_{\hat{x} \in \hat{\mathscr{X}_j}} \delta_j(X, \hat{x})]$. If $d_j < d_{j,min}$ for any receiver $j \in [t]$, then there is no rate tuple $\mathbf{R} \in \mathbb{R}_+^t$ for which the distortion tuple $\mathbf{d}$ is admissible; the set $\mathscr{R}(\mathbf{d})$ is empty. In the following, we are interested in the characterisation of $\mathscr{R}(\mathbf{d})$ for distortion tuples where $d_j \geq d_{j,min}$ for all $j \in [t]$. The following proposition shows that this region is always convex; its proof follows from the standard code time sharing argument [12, Appedix].

Proposition 1: If $\mathbf{d} \in \mathbb{R}_+^t$ with $d_j \geq d_{j,min}$ for all $j \in [j]$, then the $\mathbf{d}$-admissible rate region $\mathscr{R}(\mathbf{d})$ is a closed convex subset of $\mathbb{R}_+^t$.

The inner bound that we will develop in Theorem 1 requires $2^t - 1$ auxiliary random variables – one for each non-empty subset of receivers. For this purpose we let, for each non-empty subset $\mathscr{S} \subseteq [t]$, $\mathscr{A}_\mathscr{S}$ be a discrete finite alphabet and $U_\mathscr{S}$ be an auxiliary random variable defined on $\mathscr{A}_\mathscr{S}$. Additionally, we let $\mathscr{U} \triangleq \{U_\mathscr{T};\ \emptyset \neq \mathscr{T} \subseteq [t]\}$ be the set of all such auxiliary variables, and we define the following two subsets of $\mathscr{U}$:

$$\mathscr{U}^*(\mathscr{S}) \triangleq \big\{ U_\mathscr{T} \ \ : \ \ \mathscr{S} \subsetneq \mathscr{T} \subseteq [t] \big\} \ , \text{ and}$$

$$\mathscr{U}^\dagger(\mathscr{S}) \triangleq \left\{ U_\mathscr{T} \ \ : \ \ \begin{matrix} \mathscr{T} \subseteq [t], & \mathscr{T} \cap \mathscr{S} \neq \emptyset, \\ \mathscr{T} \neq \mathscr{S}, & |\mathscr{T}| = |\mathscr{S}| \end{matrix} \right\} .$$

We define the auxiliary random variables in $\mathscr{U}$ via a family of probability mass functions $\mathscr{P}(\mathbf{d}, Q)$ on $\prod_{\mathscr{S}} \mathscr{A}_{\mathscr{S}} \times \mathscr{X} \times \mathscr{Y}_1 \times \cdots \times \mathscr{Y}_t$. Specifically, a probability mass function p is a member of $\mathscr{P}(\mathbf{d}, Q)$ if it satisfies the following four properties:

(P1) The $\mathscr{X} \times \mathscr{Y}_1 \times \mathscr{Y}_2 \times \cdots \times \mathscr{Y}_t$ marginal of p is equal to Q.

(P2) p factors to form the Markov chain $\mathscr{U} \multimap X \multimap (Y_1, Y_2, \ldots, Y_t)$.

(P3) For every subset $\mathscr{S} \subseteq [t]$, p factors to form the Markov chain $U_{\mathscr{S}} \multimap (\mathscr{U}^*(\mathscr{S}), X) \multimap \mathscr{U}^\dagger(\mathscr{S})$.

(P4) For every receiver $j \in [t]$, there exists a deterministic function $\hat{X}_j\left(Y_j, U_j, \mathscr{U}^*(j)\right)$ with

$$\mathbb{E}\, \delta_j \left(X, \hat{X}_j\left(Y_j, U_j, \mathscr{U}^*(j)\right)\right) \leq d_j.$$

For each mass function $p \in \mathscr{P}(\mathbf{d}, Q)$, define

$$\mathscr{R}(\mathbf{d}, p) \triangleq \left\{ \mathbf{R} \in \mathbb{R}_+^t \ : \ \sum_{i=1}^{j} R_i \geq \sum_{\substack{\mathscr{S} \subseteq [t] \\ \mathscr{S} \cap [j] \neq \emptyset}} \max_{i \in \mathscr{S} \cap [j]} I\left(X; U_{\mathscr{S}} \mid Y_i, \mathscr{U}^*(\mathscr{S})\right) \ , \ \forall j \in [t]) \right\} \tag{2}$$

and let

$$\mathscr{R}^*(\mathbf{d}) = \mathrm{co} \left(\bigcup_{p \in \mathscr{P}(\mathbf{d}, Q)} \mathscr{R}(\mathbf{d}, p) \right),$$

where $\mathrm{co}(\cdot)$ denotes the closure of the convex hull. The following theorem is the main result of the paper.

Theorem 1: If $\mathbf{d} \in \mathbb{R}_+^t$ and $d_j \geq d_{j,\min}$ for every receiver $j \in [t]$, then every rate tuple within $\mathscr{R}^*(\mathbf{d})$ is $\mathbf{d}$-admissible:

$$\mathscr{R}^*(\mathbf{d}) \subseteq \mathscr{R}(\mathbf{d}) \ .$$

The proof of this coding theorem is provided in the next section. The following two examples show that this inner bound yields the entire $\mathbf{d}$-admissible rate region when the side information is degraded, and it reduces to the largest known inner bound for the side information scalable source coding problem.

Example 1 (Degraded Side Information): The side information is said to be degraded if $X \multimap Y_t \multimap Y_{t-1} \multimap \cdots \multimap Y_1$ forms a Markov chain. The first result for degraded side information was provided by Steinberg and Merhav [6, Theorem 1] for two receivers $t = 2$. This result was subsequently extended by Tian and Diggavi [9, Theorem 1] to any finite number of receivers $t > 2$. To see how Theorem 1 gives the coding part of [9, Theorem 1] consider the following. Suppose $\mathbf{d} \in \mathbb{R}_+^t$ with $d_j \geq d_{j,min}$ for all $j \in [t]$, and let $\mathscr{P}_{deg}(\mathbf{d}, Q)$ denote those mass functions in $\mathscr{P}(\mathbf{d}, Q)$ where $U_{\mathscr{S}}$ is degenerate (constant) whenever $\mathscr{S} \neq [j, t]$ for some $j \in [t]$. For each $p \in \mathscr{P}_{deg}(\mathbf{d}, Q)$, the region specified by (2) simplifies to

$$\mathscr{R}(\mathbf{d}, p) \triangleq \left\{ \mathbf{R} \in \mathbb{R}_+^t \ : \ \sum_{i=1}^{j} R_i \geq \sum_{i=1}^{j} I\left(X; U_{[i,t]} \mid Y_i, U_{[1,t]}, U_{[2,t]}, \ldots, U_{[i-1,t]}\right), \ \forall j \in [t] \right\},$$

which is the desired result.

Example 2 (Side Information Scalable Source Coding): When $t = 2$ and $X \multimap Y_2 \multimap Y_1$ forms a Markov chain, Heegard and Berger [4, Theorem 3] showed that an optimal compression strategy should satisfy distortion constraints of receiver 2 after the distortion constraints of receiver 1 have been satisfied. However, if the side information is not degraded, then this ordering may not be optimal. This observation lead Tian and Diggavi [8] to propose the side information scalable source coding problem, where it is assumed that $X \multimap Y_1 \multimap Y_2$ forms a Markov chain. Under this Markov constraint, the region defined by (2) reduces to

$$\mathscr{R}(\mathbf{d}, p) = \left\{ (R_1, R_2) \in \mathbb{R}_+^2 \ : \ \begin{array}{ll} R_1 & \geq I\left(X; U_1, U_{12} \mid Y_1\right) \\ R_1 + R_2 & \geq I\left(X; U_2, U_{12} \mid Y_2\right) \\ & \quad + I\left(X; U_1 \mid Y_1, U_{12}\right) \end{array} \right\} ,$$

which yields the inner bound reported in [8, Theorem 1]. A single-letter solution for this problem remains open.

III. Proof of Theorem 1

We now show that every rate tuple $\mathbf{R} \in \mathscr{R}(\mathbf{d}, p)$ is $\mathbf{d}$-admissible for any $p \in \mathscr{P}(\mathbf{d}, Q)$. (The $\mathbf{d}$-admissibility of rate tuples in $\mathscr{R}^*(\mathbf{d})$ follows by the standard code time sharing argument.) The main ingredient of the proof is a multi-layered random coding argument, which uses Kramer's notion of ϵ-letter typical sequences [13]. For convenience, we have reviewed the relevant ϵ-letter typical results in Appendix II. Finally, to help elucidate the main ideas of the random coding argument, we present the special case of $t = 3$ receivers as a series of examples in parallel to the main proof.

A. Code Construction

Suppose $\mathbf{d} \in \mathbb{R}_+^t$ (with $d_j \geq d_{j,min}$ for every receiver $j \in [t]$) and $p \in \mathscr{P}(\mathbf{d}, Q)$ are given. For each non-empty subset $\mathscr{S} \subseteq [t]$, construct an $|\mathscr{S}|$-layer nested codebook in the following manner: for each vector valued index

$$\mathbf{k}_{\mathscr{S}} \triangleq \left(k_{\mathscr{S},1}, k_{\mathscr{S},2}, \ldots, k_{\mathscr{S},|\mathscr{S}|}, k'_{\mathscr{S}}\right) ,$$

with $k_{\mathscr{S},i} \in [2^{nR_{\mathscr{S},i}}]$, $i = 1, 2, \ldots, |\mathscr{S}|$ and $k'_{\mathscr{S}} \in [2^{nR'_{\mathscr{S}}}]$, generate a length n codeword $a^n_{\mathscr{S}}(\mathbf{k}_{\mathscr{S}}) \in \mathscr{A}^n_{\mathscr{S}}$ by selecting n symbols from $\mathscr{A}_{\mathscr{S}}$ in an i.i.d. manner using the $U_{\mathscr{S}}$ marginal of p. The quantities $R_{\mathscr{S},i}$ and $R'_{\mathscr{S}}$ will be defined shortly.

Example 3 (3-Receivers Code Construction): We construct seven nested codebooks; one codebook for each non-empty subset of $\{1, 2, 3\}$. Figure 5 shows the 3-layer nested codebook associated with the subset

$\{1,2,3\}$. In the first layer, there are $2^{nR_{123,1}}$ bins (labelled with the index $k_{123,1}$) each of which contain $2^{n(R'_{123}+R_{123,2}+R_{123,3})}$ codewords. The set of codewords inside a particular layer one bin define the second layer of the codebook. Specifically, each layer one index $k_{123,1} \in [2^{nR_{123,1}}]$ identifies $2^{nR_{123,2}}$ layer two bins. These bins are labelled with the index $k_{123,2}$, and each bin contains $2^{n(R'_{123}+R_{123,3})}$ codewords. Similarly, each pair $k_{123,1} \in [2^{nR_{123,1}}]$ and $k_{123,2} \in [2^{nR_{123,2}}]$ identifies $2^{nR_{123,3}}$ layer three bins. There are $2^{n(R'_{123})}$ codewords in each one of the layer three bins.

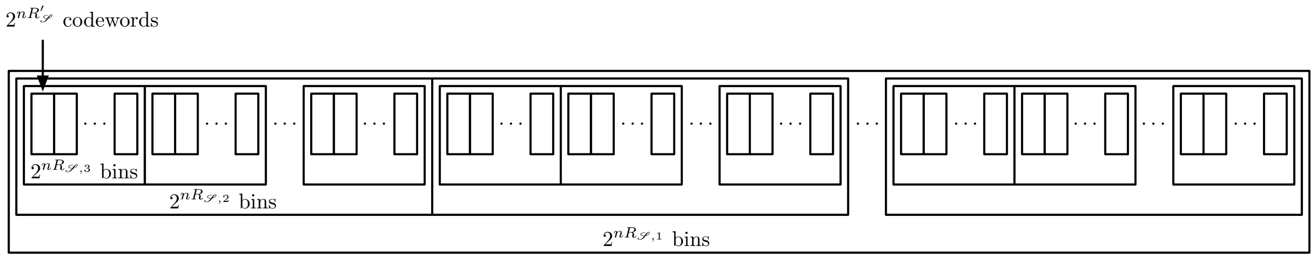

Fig. 5. Three-layer codebook.

B. Encoding

To describe the encoding and decoding procedure, it will be convenient to introduce some additional notation. Arrange the subsets of $[t]$ into a list with descending cardinality. (For subsets with the same cardinality, use lexicographical ordering). With a slight abuse of notation, label the resulting list with the sequence $\mathscr{S}_1, \mathscr{S}_2, \ldots, \mathscr{S}_{2^t-1}$. For example, for $t=3$ receivers we have: $\mathscr{S}_1 = \{1,2,3\}$, $\mathscr{S}_2 = \{1,2\}$, $\mathscr{S}_3 = \{1,3\}$, $\mathscr{S}_4 = \{2,3\}$, $\mathscr{S}_5 = \{1\}$, $\mathscr{S}_6 = \{2\}$ and $\mathscr{S}_7 = \{3\}$. Now define

$$\mathscr{U}^{\ddagger}(\mathscr{S}_j) \triangleq \{U_{\mathscr{S}_i} \;: \mathscr{S}_j \cap \mathscr{S}_i \neq \emptyset,\ |\mathscr{S}_j| = |\mathscr{S}_i|,\ i<j\}$$

to be those auxiliary random variables labelled by lower indexed sets, which share at least one element with $\mathscr{S}_j$ and have the same size as $\mathscr{S}_j$. Finally, let $\mathscr{S}(i)$ denote the i-th element of $\mathscr{S}$ under natural ordering. For example, if $\mathscr{S} = \{1,3\}$ then $\mathscr{S}(1) = 1$ and $\mathscr{S}(2) = 3$.

Encoding proceeds sequentially in 2^t-1 stages using ϵ-letter typical set encoding rules. For this purpose, choose $0 < \epsilon_0 < \epsilon_1 < \cdots < \epsilon_{2^t}$ to be arbitrarily small real numbers.

The transmitter is given a vector $x^n \in \mathscr{X}^n$. At encoding stage j (for $j = 1,2,\ldots,2^t-1$), it selects the codebook with label $\mathscr{S}_j$ and looks for an index vector $\mathbf{k}_{\mathscr{S}_j}$ where the corresponding codeword $a^n_{\mathscr{S}_j}(\mathbf{k}_{\mathscr{S}_j})$

$\mathscr{S}_j$	Subset	$A.R.V.$	$U^{\ddagger}(\mathscr{S}_j)$	Index-to-Channel Map
$\mathscr{S}_1$	$\{1,2,3\}$	$U_{\mathscr{S}_1}$	$\emptyset$	$k_{\mathscr{S}_1,1} \rightarrow$ Channel 1 $k_{\mathscr{S}_1,2} \rightarrow$ Channel 2 $k_{\mathscr{S}_1,3} \rightarrow$ Channel 3
$\mathscr{S}_2$	$\{1,2\}$	$U_{\mathscr{S}_2}$	$\emptyset$	$k_{\mathscr{S}_2,1} \rightarrow$ Channel 1 $k_{\mathscr{S}_2,2} \rightarrow$ Channel 2
$\mathscr{S}_3$	$\{1,3\}$	$U_{\mathscr{S}_3}$	$\{U_{\mathscr{S}_2}\}$	$k_{\mathscr{S}_3,1} \rightarrow$ Channel 1 $k_{\mathscr{S}_3,2} \rightarrow$ Channel 3
$\mathscr{S}_4$	$\{2,3\}$	$U_{\mathscr{S}_4}$	$\{U_{\mathscr{S}_2}, U_{\mathscr{S}_3}\}$	$k_{\mathscr{S}_4,1} \rightarrow$ Channel 2 $k_{\mathscr{S}_4,2} \rightarrow$ Channel 3
$\mathscr{S}_5$	$\{1\}$	$U_{\mathscr{S}_5}$	$\emptyset$	$k_{\mathscr{S}_5,1} \rightarrow$ Channel 1
$\mathscr{S}_6$	$\{2\}$	$U_{\mathscr{S}_6}$	$\emptyset$	$k_{\mathscr{S}_6,1} \rightarrow$ Channel 2
$\mathscr{S}_7$	$\{3\}$	$U_{\mathscr{S}_7}$	$\emptyset$	$k_{\mathscr{S}_7,1} \rightarrow$ Channel 3

Fig. 6. The figure illustrates the label and channel-to-index assignments for three receivers.

is ϵ_j-letter typical with x^n,

$$\left\{a^n_{\mathscr{S}_i}(\mathbf{k}_{\mathscr{S}_i}) : \mathscr{S}_i \supsetneq \mathscr{S}_j,\ i < j\right\},\ \text{and}$$

$$\left\{a^n_{\mathscr{S}_i}(\mathbf{k}_{\mathscr{S}_i}) : \mathscr{S}_i \cap \mathscr{S}_j \neq \emptyset, |\mathscr{S}_i| = |\mathscr{S}_j|,\ i < j\right\}\ . \tag{3}$$

If successful[5], the transmitter sends the bin index $k_{\mathscr{S}_j,i}$ over channel $\mathscr{S}_j(i)$ for each $i = 1, 2, \ldots, |\mathscr{S}_j|$. If unsuccessful, the transmitter sends $k_{\mathscr{S}_j,i} = 1$ over each of these channels.

Example 4 (3-Receivers Encoding): Figure 6 illustrates the labels used to identify the seven non-empty subsets of $\{1,2,3\}$; the assignment of seven auxiliary random variables; the members of each of the sets $U^{\ddagger}(\mathscr{S}_j)$; and the channels on which the bin indices are sent. In the first encoding stage, the transmitter looks for a vector $\mathbf{k}_{\mathscr{S}_1} = (k_{\mathscr{S}_1,1}, k_{\mathscr{S}_1,2}, k_{\mathscr{S}_1,3}, k'_{\mathscr{S}_1})$ such that the corresponding codeword $a^n_{\mathscr{S}_1}(\mathbf{k}_{\mathscr{S}_1})$ is typical with x^n. The indices $k_{\mathscr{S}_1,1}$, $k_{\mathscr{S}_1,2}$ and $k_{\mathscr{S}_1,3}$ are sent over channels 1, 2 and 3 respectively. In the fourth encoding stage, the encoder looks for a vector $\mathbf{k}_{\mathscr{S}_4} = (k_{\mathscr{S}_4,1}, k_{\mathscr{S}_4,2}, k'_{\mathscr{S}_4})$ such that the corresponding codeword $a^n_{\mathscr{S}_4}(\mathbf{k}_{\mathscr{S}_4})$ is typical with $a^n_{\mathscr{S}_1}(\mathbf{k}_{\mathscr{S}_1})$, $a^n_{\mathscr{S}_2}(\mathbf{k}_{\mathscr{S}_2})$, $a^n_{\mathscr{S}_3}(\mathbf{k}_{\mathscr{S}_3})$ and x^n. Similarly, in the sixth encoding stage, the transmitter looks for a vector $\mathbf{k}_{\mathscr{S}_6} = (k_{\mathscr{S}_6,1}, k'_{\mathscr{S}_6})$ such that the corresponding codeword $a^n_{\mathscr{S}_6}(\mathbf{k}_{\mathscr{S}_6})$ is typical with $a^n_{\mathscr{S}_1}(\mathbf{k}_{\mathscr{S}_1})$, $a^n_{\mathscr{S}_2}(\mathbf{k}_{\mathscr{S}_2})$, $a^n_{\mathscr{S}_4}(\mathbf{k}_{\mathscr{S}_4})$ and x^n.

[5]If there are two-or-more such codewords, we assume that the transmitter selects one codeword arbitrarily and sends the corresponding indices.

C. Decoding

Like the encoding procedure, receiver l (for each $l \in [t]$) forms its reconstruction $\hat{X}_l^n$ using $2^t - 1$ sequential decoding stages. Recall, receiver l recovers every bin index transmitted on channels 1 through l; it does not have access to any index transmitted on channels $l+1$ through t. In stage j (for all stages $j = 1, 2, \ldots, 2^t - 1$) it considers subset $\mathscr{S}_j$. If $l \notin \mathscr{S}_j$, then it does nothing and moves to decoding stage $j+1$. If $l \in \mathscr{S}_j$, then it takes the bin indices

$$\left\{k_{\mathscr{S}_j,i};\ i = 1, 2, \ldots, |[l] \cap \mathscr{S}_j|\right\}$$

and looks for an index vector $\tilde{\mathbf{k}}_{\mathscr{S}_j}$, with $\tilde{k}_{\mathscr{S}_j,i} = k_{\mathscr{S}_j,i}$ for all $i = 1, 2, \ldots, |[l] \cap \mathscr{S}_j|$, such that the corresponding codeword $a^n(\tilde{\mathbf{k}}_{\mathscr{S}_j})$ is ϵ_{j+1}-letter typical with y_l^n and those codewords which belong to supersets of $\mathscr{S}_j$:

$$\left(\left\{a^n_{\mathscr{S}_i}(\hat{\mathbf{k}}_{\mathscr{S}_i}) : \mathscr{S}_i \supsetneq \mathscr{S}_j, i < j\right\}, a^n_{\mathscr{S}_j}(\tilde{\mathbf{k}}_{\mathscr{S}_j}), y_l^n\right) \in T^{(n)}_{\epsilon_{j+1}}(p). \tag{4}$$

There are exactly

$$\exp_2\left[n\left(R'_{\mathscr{S}_j} + \sum_{i=|[l]\cap\mathscr{S}_j|+1}^{|\mathscr{S}_j|} R_{\mathscr{S}_j,i}\right)\right]$$

codewords in the bin specified by the indices $\{k_{\mathscr{S}_j,i} : i = 1, 2, \ldots, |[l] \cap \mathscr{S}_j|\}$. If one or more of these codewords satisfy this typicality condition, then receiver l selects one arbitrarily and sets $\hat{\mathbf{k}}_{\mathscr{S}_j} = \tilde{\mathbf{k}}_{\mathscr{S}_j}$. If there is no such codeword, it sets each of the unknown indices equal to 1.

Example 5 (3-Receivers Decoding): Consider the second receiver ($l = 2$). In stage one, take $k_{\mathscr{S}_1,1}$ (from channel 1) and $k_{\mathscr{S}_1,2}$ (from channel 2) and look for a vector $\tilde{\mathbf{k}}_{\mathscr{S}_1} = (k_{\mathscr{S}_1,1}, k_{\mathscr{S}_1,2}, \tilde{k}_{\mathscr{S}_1,3}, \tilde{k}'_{\mathscr{S}_1})$ such that the corresponding codeword $a^n_{\mathscr{S}_1}(\tilde{\mathbf{k}}_{\mathscr{S}_1})$ is typical with y_2^n. Similarly, in stage four take $k_{\mathscr{S}_4,1}$ (from channel 2) and look for $\tilde{\mathbf{k}}_{\mathscr{S}_4} = (k_{\mathscr{S}_4,1}, \tilde{k}_{\mathscr{S}_4,2}, \tilde{k}'_{\mathscr{S}_4})$ such that the corresponding codeword $a^n_{\mathscr{S}_4}(\tilde{\mathbf{k}}_{\mathscr{S}_4})$ is jointly typical with $a^n_{\mathscr{S}_1}(\hat{\mathbf{k}}_{\mathscr{S}_1})$ and y_2^n. Finally, in stage six take $k_{\mathscr{S}_6,1}$ (from channel 2) and look for $\tilde{\mathbf{k}}_{\mathscr{S}_6} = (k_{\mathscr{S}_6,1}, \tilde{k}'_{\mathscr{S}_6})$ such that the corresponding codeword $a^n_{\mathscr{S}_6}(\tilde{\mathbf{k}}_{\mathscr{S}_6})$ is jointly typical with $a^n_{\mathscr{S}_1}(\hat{\mathbf{k}}_{\mathscr{S}_1})$, $a^n_{\mathscr{S}_2}(\hat{\mathbf{k}}_{\mathscr{S}_2})$, $a^n_{\mathscr{S}_4}(\hat{\mathbf{k}}_{\mathscr{S}_4})$ and y_2^n.

D. Error Analysis: Encoding

The coding scheme is based on ϵ-letter typical set encoding and decoding techniques. As such, the distortion criteria at each receiver will not be satisfied when $(x^n, y_1^n, y_2^n, \ldots, y_t^n) \notin T^{(n)}_{\epsilon_0}(p)$ – an event we denoted by E_1. From Lemma 2, the probability of this event may be bound by

$$\Pr\left[E_1\right] \le \delta_1\left(n, \epsilon_0, \mu(p)\right)\ ,$$

where $\delta_1\left(n, \epsilon_0, \mu(p)\right) \rightarrow 0$ as $n \to \infty$.

Now let $E_{2,\mathscr{S}_j}$ denote the event that the transmitter fails to find an ϵ_j-letter typical codeword during stage j of encoding procedure given that it found an ϵ_i-letter typical codeword for every stage $i \in [j-1]$. From Lemma 3 and the inequality $(1-x)^t \leq e^{-tx}$ we have

$$\Pr\left[E_{2,\mathscr{S}_j}\right] = \left[1 - \Pr\Big[\left(\left\{a^n_{\mathscr{S}_i}(\mathbf{k}_{\mathscr{S}_i})\right\}, U^n_{\mathscr{S}_j}(\mathbf{k}_{\mathscr{S}_j}), x^n\right) \in T^{(n)}_{\epsilon_{j+1}}(p)\Big]\right]^{2^{n\left(R'_{\mathscr{S}_j} + \sum_{i=1}^{|S_j|} R_{\mathscr{S}_j,i}\right)}}$$
$$\leq \exp\left(- \left(1-\delta_2\right) 2^{n\left(R'_{\mathscr{S}_j} + \sum_{i=1}^{|S_j|} R_{\mathscr{S}_j,i}\right)} \cdot 2^{-n\left(I\left(\mathscr{U}^*(\mathscr{S}_j), \mathscr{U}^\ddagger(\mathscr{S}_j), X; U_{\mathscr{S}_j}\right) + 2\epsilon_j H\left(U_{\mathscr{S}_j}\right)\right)}\right) \quad (5)$$

where, for compact representation, we have written the ϵ_j-letter typicality condition in (3) as $(\{a^n_{\mathscr{S}_i}(\mathbf{k}_{\mathscr{S}_i})\}$, $U^n_{\mathscr{S}_j}(\mathbf{k}_{\mathscr{S}_j}),\ x^n) \in T^{(n)}_{\epsilon_j}(p)$ and the function $\delta_2\left(n, \epsilon_{j-1}, \epsilon_j, \mu(p)\right)$ as δ_2.

From property **(P4)** we have that $U_{\mathscr{S}_j} \multimap (\mathscr{U}^*(\mathscr{S}_j), X) \multimap \mathscr{U}^\dagger(\mathscr{S}_j)$ forms a Markov chain. Since $U^\ddagger(\mathscr{S}_j) \subseteq \mathscr{U}^\dagger(\mathscr{S}_j)$, $U_{\mathscr{S}_j} \multimap (\mathscr{U}^*(\mathscr{S}_j), X) \multimap \mathscr{U}^\ddagger(\mathscr{S}_j)$ also forms a Markov chain; therefore,

$$I\left(\mathscr{U}^*(\mathscr{S}_j), \mathscr{U}^\ddagger(\mathscr{S}_j), X; U_{\mathscr{S}_j}\right) = I\left(\mathscr{U}^*(\mathscr{S}_j), X; U_{\mathscr{S}_j}\right)\ .$$

Consequently (5) simplifies to

$$\Pr\left[E_{2,\mathscr{S}_j}\right] \leq \exp\left(- \left(1-\delta_2\right) 2^{n\left(R'_{\mathscr{S}_j} + \sum_{i=1}^{|S_j|} R_{\mathscr{S}_j,i}\right)} \cdot 2^{-n\left(I\left(\mathscr{U}^*(\mathscr{S}_j), X; U_{\mathscr{S}_j}\right) + 2\epsilon_j H\left(U_{\mathscr{S}_j}\right)\right)}\right)\ .$$

Let E_2 denote the event where a typical codeword cannot be found at any one of the encoding stages. By the union bound we get the following upper bound for $\Pr[E_2]$:

$$\Pr\left[E_2\right] \leq \sum_{j=1}^{2^t-1} \exp\left(- \left(1-\delta_2\right) 2^{n\left(R'_{\mathscr{S}_j} + \sum_{i=1}^{|S_j|} R_{\mathscr{S}_j,i}\right)} \cdot 2^{-n\left(I\left(\mathscr{U}^*(\mathscr{S}_j), X; U_{\mathscr{S}_j}\right) + 2\epsilon_j H\left(U_{\mathscr{S}_j}\right)\right)}\right)\ .$$

Finally, note that if

$$R'_{\mathscr{S}_j} + \sum_{i=1}^{|\mathscr{S}_j|} R_{\mathscr{S}_j,i} > I\left(\mathscr{U}^*(\mathscr{S}_j), X; U_{\mathscr{S}_j}\right) + 2\epsilon_j H\left(U_{\mathscr{S}_j}\right) \quad (6)$$

for every encoding stage $j \in [2^t-1]$, then $\Pr[E_2] \to 0$ as $n \to \infty$.

E. Error Analysis: Decoding

Consider the l-th receiver (for all $l \in [t]$) and a set $\mathscr{S}_j$ with $j \in \mathscr{S}_j$. Let $D_{l,\mathscr{S}_j}$ be the event that it cannot find a unique codeword during decoding stage j, which satisfies the typicality condition (4); given that at every stage $i < j$ it found a unique codeword satisfying this typicality condition.

By the Markov lemma (Lemma 4), the probability that the codeword $a^n_{\mathscr{S}_j}(\mathbf{k}_{\mathscr{S}_j})$ selected by the transmitter is not jointly typical with y^n_l is small for large n:

$$\Pr\left[Y^n_l \notin T^{(n)}_{\epsilon_{j+1}}\Big(p \mid \{a^n_{\mathscr{S}_i}(\mathbf{k}_{\mathscr{S}_i})\}, a^n_{\mathscr{S}_j}(\mathbf{k}_{\mathscr{S}_j}), x^n\Big)\right] \leq \delta_2\left(n, \epsilon_j, \epsilon_{j+1}, \mu(p)\right),$$

where for brevity we have used

$$\left\{a^n_{\mathscr{S}_i}(\mathbf{k}_{\mathscr{S}_i})\right\} = \left\{a^n_{\mathscr{S}_i}(\mathbf{k}_{\mathscr{S}_i}) : \mathscr{S}_i \supsetneq \mathscr{S}_j, i < j\right\}. \tag{7}$$

An upper bound for the probability that there exists one or more codewords $a^n_{\mathscr{S}_j}(\tilde{\mathbf{k}}_{\mathscr{S}_j}) \neq a^n_{\mathscr{S}_j}(\mathbf{k}_{\mathscr{S}_j})$, which satisfy (4), is

$$\Pr\left[\bigcup_{\mathscr{K}_j}\left\{\{a^n_{\mathscr{S}_i}(\mathbf{k}_{\mathscr{S}_i})\}, y^n_l, a^n_{\mathscr{S}_j}(\tilde{\mathbf{k}}_{\mathscr{S}_j})\right\} \in T^{(n)}_{\epsilon_{j+1}}(p)\right]$$

$$< \exp_2\left[n\Big(R'_{\mathscr{S}_j} + \sum_{i=|[l]\cap S_j|+1}^{|\mathscr{S}_j|} R_{\mathscr{S}_j,i} - I\big(U_{\mathscr{S}_j}; \mathscr{U}^*(\mathscr{S}_j), Y_l\big) + 2\epsilon_{j+1}H\big(U_{\mathscr{S}_j}\big)\Big)\right], \tag{8}$$

where we take the union over all codewords

$$\mathscr{K}_j = \left\{\tilde{\mathbf{k}}_{S_j} \neq \mathbf{k}_{S_j},\ \{\tilde{k}_{S_j,i} = k_{S_j,i}\}_{i=1}^{|\{1,2,\ldots,l\}\cap S_j|}\right\},$$

and we have again used (7) for brevity. Applying the union bound we get

$$\Pr\left[D_{l,\mathscr{S}_j}\right] < \delta_2 + \exp_2\left[n\left(R'_{\mathscr{S}_j} + \sum_{i=|[l]\cap\mathscr{S}_j|}^{\mathscr{S}_j} R_{\mathscr{S}_j,i}\right) - n\left(I\Big(U_{\mathscr{S}_j}; \mathscr{U}^*(\mathscr{S}_j), Y_l\Big) - 2\epsilon_{j+1}H(U_{\mathscr{S}_j})\right)\right].$$

Thus, if

$$R'_{\mathscr{S}_j} + \sum_{i=|[l]\cap S_j|+1}^{|\mathscr{S}_j|} R_{\mathscr{S}_j,i} < I\left(U_{\mathscr{S}_j}; \mathscr{U}^*(\mathscr{S}_j), Y_l\right) - 2\epsilon_{j+1}H\left(U_{\mathscr{S}_j}\right) \tag{9}$$

then $\Pr[D_{l,\mathscr{S}_j}] \to 0$ as $n \to \infty$.

F. Rate Constraints

Consider receiver l and any subset $\mathscr{S}$ where $l \in \mathscr{S}$. On combining the rate constraints (6) and (9) we get

$$\sum_{i=1}^{|\mathscr{S}\cap[l]|} R_{\mathscr{S},i} > I\big(\mathscr{U}^*(\mathscr{S}), X; U_{\mathscr{S}}\big) - I\big(U_{\mathscr{S}}; \mathscr{U}^*(\mathscr{S}), Y_l\big). \tag{10}$$

(Since ϵ_j and ϵ_{j+1} may be selected arbitrarily small, we can ignore the $2(\epsilon_j + \epsilon_{j+1})H(\mathscr{S}_j)$ term.) From property **(P2)** we have that $\mathscr{U} \multimap X \multimap Y_l$ forms a Markov chain. This implies $U_{\mathscr{S}} \multimap (\mathscr{U}^*(\mathscr{S}), X) \multimap Y_l$

forms a Markov chain and $I\big(\mathscr{U}^*(\mathscr{S}), X; U_{\mathscr{S}}\big) = I\big(\mathscr{U}^*(\mathscr{S}), X, Y_l; U_{\mathscr{S}}\big)$. Consequently, simplifies the rate constraint (10) simplifies to

$$\sum_{i=1}^{|[l]\cap\mathscr{S}|} R_{\mathscr{S},i} > I\left(X; U_{\mathscr{S}} \mid \mathscr{U}^*(\mathscr{S}), Y_l\right) \ . \tag{11}$$

Repeating this procedure for any receiver $\tilde{l} \in [l] \cap \mathscr{S}_j$, we obtain

$$\sum_{i=1}^{|[\tilde{l}]\cap\mathscr{S}_j|} R_{\mathscr{S},i} > I\left(X; U_{\mathscr{S}} \mid \mathscr{U}^*(\mathscr{S}), Y_{\tilde{l}}\right) \ .$$

Since $R_{\mathscr{S},i} \geq 0$ for all i, it must be true that

$$\sum_{i=1}^{|[l]\cap\mathscr{S}|} R_{\mathscr{S},i} > \max_{\tilde{l}\in[l]\cap\mathscr{S}} I\left(X; U_{\mathscr{S}} \mid \mathscr{U}^*(\mathscr{S}), Y_{\tilde{l}}\right) \ ; \tag{12}$$

that is, the rate constraint for receiver l must be at least as large as the rate constraint for receiver $\tilde{l}$.

The rate constraint (12) is valid for any set $\mathscr{S}$ where $l \in \mathscr{S}$. For those subsets $\mathscr{S}$ with $l \in \mathscr{S}$, define $l^* \triangleq \max_{i\in[l]\cap\mathscr{S}} i$. Since $l^* \in \mathscr{S}$ and $[l^*] \cap \mathscr{S} = [l] \cap \mathscr{S}$, it follows that (12) is also valid for any set $\mathscr{S}$ where $[l] \cap \mathscr{S} \neq \emptyset$.

Finally, consider the sum rate $\sum_{i=1}^{l} R_i$ for the first l channels. By construction, we have that

$$\sum_{i=1}^{l} R_i = \sum_{\substack{\mathscr{S}\subseteq[t] \\ \mathscr{S}\cap[l]\neq\emptyset}} \sum_{i=1}^{|[l]\cap\mathscr{S}|} R_{\mathscr{S},i} \ . \tag{13}$$

Substituting the rate constraint (12) into (13) yields the desired result.

IV. Rate-Distortion with Receiver Side Information

We now turn attention to Heegard and Berger's lossy source coding problem shown in Figure 3. This problem may be recovered from the setup of Section II by choosing $|\mathscr{M}_2| = |\mathscr{M}_3| = \cdots = |\mathscr{M}_t| = 1$. We are interested in the characterisation of the rate-distortion function

$$R(\mathbf{d}) = \inf\left\{R_1 \in \mathbb{R}_+ : (R_1, 0, \ldots, 0) \in \mathscr{R}(\mathbf{d})\right\} \ .$$

A single-letter characterisation of $R(\mathbf{d})$ is an open problem, and in this section we provide an upper bound.

Given a distortion tuple $\mathbf{d}$ and family $\mathscr{Q}$ of probability mass functions on $\prod_{\mathscr{S}} \mathscr{A}_{\mathscr{S}} \times \mathscr{X} \times \mathscr{Y}_1 \times \cdots \times \mathscr{Y}_t$, define

$$R^*(\mathbf{d}, \mathscr{Q}) \triangleq \min_{p\in\mathscr{Q}} \left\{ \sum_{\mathscr{S}\subseteq[t]} \max_{j\in\mathscr{S}} I\left(X; U_{\mathscr{S}} \mid Y_j, \mathscr{U}^*(\mathscr{S})\right) \right\} .$$

The following upper bound for $R(\mathbf{d})$ follows directly from Theorem 1.

Corollary 1: If $\mathbf{d} \in \mathbb{R}_+^t$ and $d_j \geq d_{j,min}$ for every receiver $j \in [t]$, then

$$R^*(\mathbf{d}, \mathscr{P}(\mathbf{d}, Q)) \geq R(\mathbf{d}).$$

At this point it is useful to recall Heegard and Berger's function $R_{HB}(\mathbf{d})$ from [4, Theorem 2]. Adopting the above notation, this function may be written as

$$R_{HB}(\mathbf{d}) = R^*(\mathbf{d}, \mathscr{P}_{HB}(\mathbf{d}, Q)) \tag{14}$$

where $\mathscr{P}_{HB}(\mathbf{d}, Q)$ is set of probability mass functions satisfying **(P1)**, **(P2)** and **(P4)** – but not **(P3)**. Hence $R^*(\mathbf{d}, \mathscr{P}(\mathbf{d}, Q))$ and $R_{HB}(\mathbf{d})$ differ only in the set on which the minimization takes place. The following example shows that the Markov condition provided by **(P3)** is not superfluous, and it provides a counterexample to the claim of [4, Theorem 2].

Example 6 ($R_{HB}(\mathbf{d})$ can be smaller than $R(\mathbf{d})$): Let $t = 3$ and suppose $Y_1 = Y_2 = Y_3 =$ constant. Let $\hat{\mathscr{X}}_j = \mathscr{X} = \{0, 1, 2\}$ for all j with Hamming distortion,

$$\delta_H(x, \hat{x}) = \begin{cases} 0, & \text{if } \hat{x} = x \\ 1, & \text{otherwise,} \end{cases} \tag{15}$$

Additionally, consider the situation where it is desired that X^n is recovered at each receiver with $d_1 = d_2 = d_3 = 0$. Finally, suppose that B and C are independent random variables, uniform on $\{0, 1, 2\}$, and set: $X = B$; $U_1 = U_2 = U_3 = U_{123} =$ constant; $U_{12} = C$; $U_{13} = B \oplus C$; and $U_{23} = B \oplus 2C$ in modulo-3 arithmetic.

The above selection of random variables X, $U_{\{1,2\}}$, $U_{\{1,3\}}$ and $U_{\{2,3\}}$ is by no means arbitrary. Both functions R_{HB} and R^* require existence of functions $\hat{X}_1 : \mathscr{A}_{12} \times \mathscr{A}_{13} \to \mathscr{X}$, $\hat{X}_2 : \mathscr{A}_{12} \times \mathscr{A}_{23} \to \mathscr{X}$ and $\hat{X}_3 : \mathscr{A}_{13} \times \mathscr{A}_{23}, \to \mathscr{X}$ where $x = \hat{X}_1(a_{12}, a_{13})$, $x = \hat{X}_2(a_{12}, a_{23})$ and $x = \hat{X}_3(a_{13}, a_{23})$ whenever $p(x, a_{12}, a_{13}) > 0$, $p(x, a_{12}, a_{23}) > 0$ and $p(x, a_{13}, a_{23}) > 0$ respectively. It is readily checked that this selection of random variables implies the existence of such functions. Finally, note that the Markov chains $U_{12} \multimap X \multimap (U_{13}, U_{12})$, $U_{13} \multimap X \multimap (U_{12}, U_{23})$ and $U_{23} \multimap X \multimap (U_{12}, U_{23})$ do not hold; therefore, this is not a valid selection of auxiliary random variables for the upper bound in Corollary 1.

It is clear that $R_{HB}(0,0,0) = I(X; U_{12}) + I(X; U_{13}) + I(X; U_{23})$. However, on closer inspection, it can be seen that each of these mutual information terms are equal to zero. In Appendix I we prove that $R(0,0,0) \geq H(X) > 0$; therefore, $R_{HB}(0,0,0) < R(0,0,0)$.

In the following two examples, we show that Corollary (1) gives the rate-distortion function for the degraded side information and complementary delivery problems.

Example 7 (Degraded Side Information): In [4, Theorem 3], Heegard and Berger characterised $R(\mathbf{d})$ under the assumption of degraded side information $X \multimap Y_t \multimap Y_{t-1} \multimap \cdots \multimap Y_1$. To see how Corollary (1) gives the coding theorem of [4, Theorem 3] consider the following. Recall the set of probability mass functions $\mathscr{P}_{deg}(\mathbf{d}, Q)$ from Example 1. On substituting $\mathscr{P}_{deg}(\mathbf{d}, Q)$ into Corollary (1) we get

$$R^*(\mathbf{d}, \mathscr{P}_{deg}(\mathbf{d}, Q)) \geq \min_{p \in \mathscr{P}_{deg}(\mathbf{d},Q)} \sum_{j=1}^{t} I\left(X; U_{[j,t]} \mid Y_j, U_{[1,t]}, U_{[2,t]}, \ldots, U_{[j-1,t]}\right) \ ,$$

which is the desired result.

Example 8 (Complementary Delivery): The complementary delivery problem was originally proposed and solved as a source coding problem (with vanishing block error probability) by Wyner, Wolf and Willems [14]. Recently, Kimura and Uyematsu [10] solved this problem in the rate-distortion setting.

Suppose that $X = (X_1, X_2, \ldots, X_t)$ is a product source on $\mathscr{X} \triangleq \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_t$, where the $\mathscr{X}_j$ are discrete finite alphabets. Additionally, suppose that the side information at receiver j is given by $Y_j = X_j^c \triangleq (X_1, X_2, \ldots, X_{j-1}, X_{j+1}, \ldots, X_t)$. The resulting rate-distortion function is given by [10]

$$R(\mathbf{d}) = \min \max_{j \in [t]} I\left(X_j; C \mid X_j^c\right),$$

where the minimization is over all choices of an auxiliary random variable C. The forward implication of this result is a special case of Corollary 1 when $U_{\mathscr{S}}$ is set to be a constant whenever $\mathscr{S} \subsetneq [t]$.

V. Lossless Source Coding with Individual Messages

If $\hat{\mathscr{X}}_j = \mathscr{X}$ for all $j = 1, 2, \ldots, t$, and the distortion measure δ_j is Hamming, then[6]

$$R(0, 0, \ldots, 0) = \max_{j \in [t]} H\left(X \mid Y_j\right).$$

The forward part of this result follows directly from Corollary 1 when U_S is set to X if $S = [t]$ and constant otherwise, and the converse is given in Appendix I. For this reason, it is generally accepted that the lossless version of the t receiver problem of Figure 3 is well understood.

In the following, we present a second lossless source coding problem for which the set of achievable rates is not known. In fact, this problem appears to be just as difficult as the rate-distortion problem.

In the same manner as the complementary delivery problem, suppose $X = (X_1,\ X_2,\ \ldots,\ X_t)$ is a product source on $\mathscr{X} = \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_t$. Now let $\hat{\mathscr{X}}_j = \mathscr{X}_j$ and assume that receiver j is interested only in lossless reconstruction of X_j^n. Of interest is the smallest rate R_{IM} such that this is possible. A direct application of Corollary 1 with the Hamming distortion measure yields an upper bound for R_{IM}. Unfortunately, however, it is not known if this bound is tight. The following corollary shows that this upper bound matches the rate distortion function when the side information is degraded.

[6]The vanishing block error probability version of this problem was solved by Sgarro in [15].

Corollary 2: If the side information is degraded $X \multimap Y_t \multimap Y_{t-1} \multimap \cdots \multimap Y_1$, then

$$R_{IM} = \sum_{j=1}^{t} H\left(X_j \mid X_1, X_2, \ldots, X_{j-1}, Y_j\right).$$

The coding theorem follows from evaluation of $R(0,0,\ldots,0)$ using Corollary 1, with $\delta_j(x,\hat{x}_j) = \delta_H(x_j,\hat{x}_j)$, and by setting $U_{\mathscr{S}} = X_j$ if $\mathscr{S} = [j,t]$ and constant otherwise. The converse follows by making some minor changes to the converse of [4, Section VII]. For expedience, these details are omitted.

VI. Conclusion

The main result of the paper (Theorem 1) gives an inner bound for the region of admissible rate tuples for the t-stage successive refinement problem with side information. This result unifies the existing inner bounds of Steinberg, Merhav, Tian and Diggavi [6], [8], [9]. An immediate application of Theorem 1 yields an upper bound for the rate-distortion function for the problem of lossy source coding problem with side information at many receivers (Corollary 1). This bound reduces to the rate-distortion function for Heegard and Berger's degraded side information problem [4, Theorem 3], as well as Kimura's complementary delivery problem [10]. Of particular interest is a counterexample to Heegard and Berger's general upper bound [4, Theorem 2] for arbitrary side information. Although the successive refinement and rate-distortion bounds presented in this paper subsume existing results in the literature, it is not clear if either bound is tight.

Appendix I
A Lower Bound for $R(\mathbf{d})$ under Hamming Distortion

Lemma 1: Consider the rate-distortion function $R(\mathbf{d})$, which is defined in Section IV. If $\hat{\mathscr{X}}_j = \mathscr{X}$ and δ_j is Hamming distortion measure (for all receivers $j \in [t]$), then

$$R(0,0,\ldots,0) \geq \max_{i\in[t]} H(X|Y_i)\ .$$

Proof: Consider the j-th receiver (for some $j \in [t]$). Let $P_{e,i} = \Pr[X_{j,i} \neq \hat{X}_{j,i}]$ denote the probability that this receiver incorrectly reconstructs symbol i (for $i \in [n]$), and let

$$P_e = \frac{1}{n}\sum_{i=1}^{n} P_{e,i} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\,\delta_H(X_{j,i},\hat{X}_{j,i}) \leq \epsilon$$

denote the average probability of symbol error over n symbols. By definition, we have

$$\begin{aligned}
\frac{1}{n}\log|\mathscr{M}| &\geq \frac{1}{n}H(M_1) \geq \frac{1}{n}H\left(M_1 \mid Y_{j,1}^n\right) \\
&\geq \frac{1}{n}I\left(X_1^n; M_1 \mid Y_{j,1}^n\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} I\left(X_i; M_1 \mid X_1^{i-1}, Y_{j,1}^n\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} I\left(X_i; M_1, X_1^{i-1}, Y_{j,1}^{i-1}, Y_{j,i+1}^n \mid Y_{j,i}\right) && (16) \\
&\geq \frac{1}{n}\sum_{i=1}^{n} I\left(X_i; M_1, Y_{j,1}^{i-1}, Y_{j,i+1}^n \mid Y_{j,i}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \left[H\left(X_i \mid Y_{j,i}\right) - H\left(X_i \mid M_1, Y_{j,1}^n\right)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \left[H\left(X_i \mid Y_{j,i}\right) - H\left(X_i \mid M_1, Y_{j,1}^n, \hat{X}_{j,i}\right)\right] && (17) \\
&\geq \frac{1}{n}\sum_{i=1}^{n} \left[H\left(X_i \mid Y_{j,i}\right) - H\left(X_i \mid \hat{X}_{j,i}\right)\right] \\
&\geq \frac{1}{n}\sum_{i=1}^{n} \left[H\left(X_i \mid Y_{j,i}\right) - h\left(P_{e,i}\right) - P_{e,i}\log|\mathscr{X}|\right] && (18) \\
&= H\left(X \mid Y_j\right) - \frac{1}{n}\sum_{i=1}^{n} h\left(P_{e,i}\right) - P_e \log|\mathscr{X}| \\
&\geq H\left(X \mid Y_j\right) - h\left(P_e\right) - P_e\log|\mathscr{X}| && (19) \\
&\geq H\left(X \mid Y_j\right) - h(\epsilon) - \epsilon\log|\mathscr{X}| && (20)
\end{aligned}$$

where (16) follows because $X_i \multimap Y_{j,i} \multimap (X_1^{i-1}, Y_{j,1}^{i-1}, Y_{j,i+1}^n)$ forms a Markov chain, (17) is due to $\hat{X}_j^n = g^{(n)}(M_1, Y_{j,1}^n)$ (18) follows from Fano's inequality [11, Page 39] with $h(\cdot)$ as the binary entropy function [11, Page 14], (19) follows from Jensen's inequality, (20) follows by assuming ϵ is small (i.e. $0 < \epsilon < 1/2$). Finally, $h(\epsilon) + \epsilon\log|\mathscr{X}| \to 0$ as $\epsilon \to 0$. ■

APPENDIX II

ϵ-LETTER TYPICALITY

For $\epsilon \geq 0$, a sequence $x^n \in \mathscr{X}^n$ is said to be ϵ-letter typical with respect to a discrete memoryless source $(\mathscr{X}, p_X)$ if

$$\left|\frac{1}{n}N(a|x^n) - p_X(a)\right| \leq \epsilon \cdot p_X(a) \quad \forall a \in \mathscr{X} \ ,$$

where $N(a|x^n)$ is the number of times the letter a occurs in the sequence x^n. The collection of all ϵ-letter typical sequences is denoted by $T_\epsilon^{(n)}(p_X)$.

In a similar fashion, a pair of sequences x^n and y^n are said to jointly ϵ-letter typical with respect to a discrete memoryless two source $(\mathscr{X} \times \mathscr{Y}, p_{XY})$ if

$$\left|\frac{1}{n}N(a,b|x^n,y^n) - p_{XY}(a,b)\right| \leq \epsilon \cdot p_{XY}(a,b) \quad \forall (a,b) \in \mathscr{X} \times \mathscr{Y} \ ,$$

where $N(a,b|x^n,y^n)$ is the number of times the pair of letters (a,b) occurs in the pair (x^n, y^n). The collection of all joint ϵ-typical sequence pairs is denoted by $T_\epsilon^{(n)}(p_{XY})$.

Given $(\mathscr{X} \times \mathscr{Y}, p_{XY})$ and $x^n \in \mathscr{X}^n$, the set

$$T_\epsilon^{(n)}\left(p_{XY} \mid x^n\right) = \left\{y^n \ : \ (x^n, y^n) \in T_\epsilon^{(n)}(p_{XY})\right\}$$

is called the set of conditionally ϵ-letter typical sequences.

Let $\mu(\mathscr{X}, p_X) = \min\{p_X(x) : \ x \in \text{support}(p_X)\}$ and define

$$\delta_1\left(n, \epsilon, \mu(p_X)\right) = 2|\mathscr{X}| \cdot e^{-n\epsilon^2 \mu(p_X)}.$$

Note, $\delta\big(n, \epsilon, \mu(p_X)\big) \to 0$ as $n \to \infty$.

Lemma 2 (Theorem 1.1, [13]): Suppose X^n is emitted by a discrete memoryless source $(\mathscr{X}, p_X)$. If $0 < \epsilon \leq \mu(p_X)$, then

$$1 - \delta_1\left(n, \epsilon, \mu(p_X)\right) \leq \Pr\left[X^n \in T_\epsilon^{(n)}(p_X)\right] \leq 1 \ .$$

Now consider a discrete memoryless two-source $(\mathscr{X} \times \mathscr{Y}, p_{XY})$, let

$$\delta_2\big(n, \epsilon_1, \epsilon_2, \mu(p_{XY})\big) = 2|\mathscr{X}||\mathscr{Y}| \cdot e^{-n\frac{(\epsilon_2 - \epsilon_1)^2}{1+\epsilon_1}\mu(p_{XY})},$$

and note that $\delta_2\big(n, \epsilon_1, \epsilon_2, \mu(p_X)\big) \to 0$ as $n \to \infty$.

Lemma 3 (Theorem 1.3, [13]): Suppose Y^n is emitted by $(\mathscr{Y}, p_Y)$ where p_Y is equal to the Y-marginal of p_{XY}. If $0 < \epsilon_1 < \epsilon_2 \leq \mu(p_{XY})$ and $x^n \in T_{\epsilon_1}^{(n)}(p_X)$, then

$$\left(1 - \delta_2\left(n, \epsilon_1, \epsilon_2, \mu(p_{XY})\right)\right) 2^{-n(I(X;Y) + 2\epsilon_2 H(Y))}$$
$$\leq \Pr\left[Y^n \in T_{\epsilon_2}^{(n)}\left(p_{XY} \mid x^n\right)\right] \leq 2^{-n(I(X;Y) - 2\epsilon_2 H(Y))}.$$

Finally, a direct consequence of Lemma 3 for Markov sources is the following result.

Lemma 4 (Markov Lemma [13]): Suppose (X^n, Y^n, Z^n) is emitted by a discrete memoryless three-source $(\mathscr{X} \times \mathscr{Y} \times \mathscr{Z}, p_{XYZ})$ where $X \ominus Y \ominus Z$. If $0 < \epsilon_1 < \epsilon_2 \leq \mu(p_{XYZ})$ and $(x^n, y^n) \in T_{\epsilon_1}^{(n)}(p_{XY})$, then

$$\begin{aligned}&\Pr\left[Z^n \in T_{\epsilon_2}^{(n)}\left(p_{XYZ} \mid x^n, y^n\right) \mid Y^n = y^n\right]\\&= \Pr\left[Z^n \in T_{\epsilon_2}^{(n)}\left(p_{XYZ} \mid x^n, y^n\right) \mid X^n = x^n, Y^n = y^n\right]\\&\geq 1 - \delta_2\left(n, \epsilon_1, \epsilon_2, \mu(p_{XYZ})\right).\end{aligned}$$

REFERENCES

[1] A. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder," *Information Theory, IEEE Transactions on*, vol. 22, no. 1, pp. 1–10, 1976.

[2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.

[3] A. H. Kaspi, "Rate-Distortion Function when Side-Information May Be Present at the Decoder," *Information Theory, IEEE Transactions on*, vol. 40, no. 6, pp. 2031–2034, 1994.

[4] C. Heegard and T. Berger, "Rate Distortion when Side Information May Be Absent," *Information Theory, IEEE Transactions on*, vol. 31, no. 6, pp. 727–734, 1985.

[5] F. Fang-Wei and R. W. Yeung, "On the Rate-Distortion Region for Multiple Descriptions," *Information Theory, IEEE Transactions on*, vol. 48, no. 7, pp. 2012–2021, 2002.

[6] Y. Steinberg and N. Merhav, "On Successive Refinement for the Wyner-Ziv Problem," *Information Theory, IEEE Transactions on*, vol. 50, no. 8, pp. 1636–1654, 2004.

[7] C. Tian and S. N. Diggavi, "A Calculation of the Heegard-Berger Rate-Distortion Function for a Binary Source," *Information Theory Workshop, 2006. ITW'06 Chengdu. IEEE*, pp. 342–346, 2006.

[8] ——, "Side-Information Scalable Source Coding," *Information Theory, IEEE Transactions on*, vol. 54, no. 12, pp. 5591–5608, 2008.

[9] C. Tian and S. Diggavi, "On Multistage Successive Refinement for Wyner-Ziv Source Coding with Degraded Side Informations," *Information Theory, IEEE Transactions on*, vol. 53, no. 8, pp. 2946–2960, 2007.

[10] A. Kimura and T. Uyematsu, "Multiterminal Source Coding with Complementary Delivery," *Arxiv preprint arXiv:0804.1602 (to appear in: IEICE Transactions on Fundamentals)*, 2008.

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

[12] R. Gray and A. Wyner, "Source Coding for a Simple network," *Bell System Technical Journal*, vol. 53, no. 9, pp. 1681–1721, 1974.

[13] G. Kramer, "Topics in Multi-User Information Theory," *Foundations and Trends in Communications and Information Theory*, vol. 4, no. 45, pp. 265–444, 2008.

[14] A. Wyner, J. Wolf, and F. Willems, "Communicating via a Processing Broadcast Satellite," *Information Theory, IEEE Transactions on*, vol. 48, no. 6, pp. 1243–1249, 2002.

[15] A. Sgarro, "Source Coding with Side Information at Several Decoders," *Information Theory, IEEE Transactions on*, vol. 23, no. 2, pp. 179–182, 1977.