

Successive Refinement on Trees: A Special Case of a New MD Coding Region

Raman Venkataramani,[†] Gerhard Kramer* and Vivek K Goyal*

[†]University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

*Bell Labs, Lucent Technologies, 600 Mountain Ave, Murray Hill, NJ 07974, USA

rvenkata@ifp.uiuc.edu, gkr@bell-labs.com, v.goyal@ieee.org

Abstract

New achievability results for the L -stage successive refinement problem with $L > 2$ are presented. These are derived from a recent achievability result for the more general problem of multiple description coding with $L > 2$ channels. It is shown that successive refinability on chains implies successive refinability on trees and that memoryless Gaussian sources are successively refinable on chains and trees.

1 Introduction

For the past half century, information theory has had a canonical Figure 1—literally the first figure of Shannon’s paper [1], reproduced below. In Shannon’s communication problem, there is a single source of information, a single user, and a single communication medium. Though clearly fundamental, this is not the only important communication problem.

Some apparent extensions do not appreciably change the problem. For example, if there are several sources but still a single channel and single user, the problem has

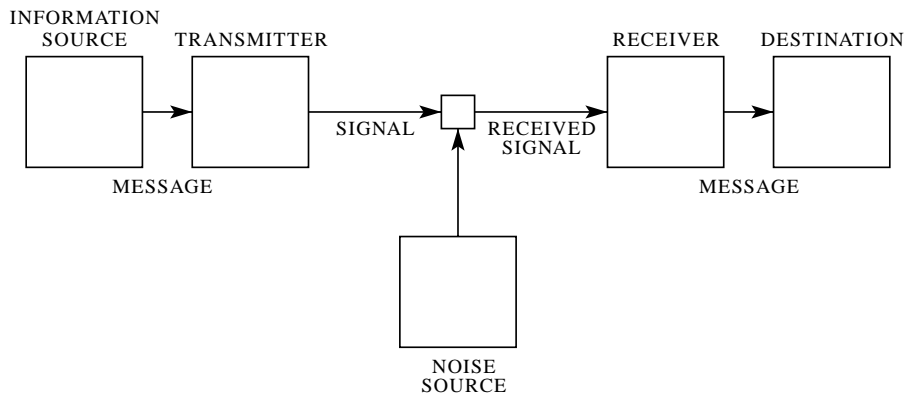


Figure 1: Shannon’s communication system abstraction (reproduced from [1]).

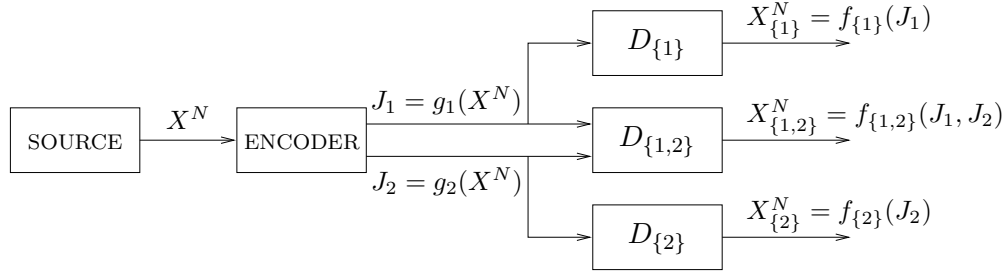


Figure 2: Multiple description source coding with two channels and three receivers. Notation is defined in Section 2. The general case has L channels and $2^L - 1$ receivers.

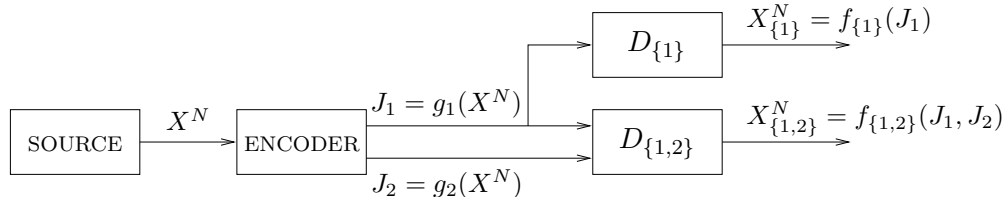


Figure 3: Successive refinement with two channels and two receivers. With the tree-defining relationship described in the text, receiver $D_{\{1,2\}}$ is a child of receiver $D_{\{1\}}$.

not really changed; the source has been replaced by a vector of sources. Similarly, having several channels to communicate a single source to a single user simply makes the channel inputs and outputs into vectors.

New communication problems arise when there are users that receive different information.¹ An early example is the study of broadcast channels [3], where the transmitter produces a single output and multiple receivers estimate the source from different noisy received signals. Broadcast channels are generally studied in the context of channel coding. However, generalizations to Shannon's model give rise to new rate-distortion problems as well.

This paper addresses a set of rate-distortion problems in which a single source is communicated to several users over several channels. In multiple description (MD) problems, there is a user for each nonempty subset of the channels. The simplest case, with $L = 2$ channels, is shown in Figure 2. Successive refinement (SR) problems are restricted versions of MD problems in which the users can be arranged in a tree where a child node receives the channels received by its parent and more. With two channels, the only possible tree has one parent and one child. This is shown in Figure 3, with receiver $D_{\{1,2\}}$ a child of receiver $D_{\{1\}}$.

For both MD and SR, the Shannon theory problem is to determine the achievable combinations of rates over the channels and distortions at the receivers. This is called the rate-distortion region. In this paper, we are primarily concerned with the rate-distortion region for SR. Since SR is equivalent to MD with certain receivers

¹This is not meant to be exhaustive. In particular, multiple receivers are not necessary for a material departure from Shannon's model; distributed, Slepian-Wolf coding [2] comes to mind.

removed, an achievable region for MD directly yields an achievable region for SR by disregarding the irrelevant distortions. This was recognized by Equitz and Cover [4], whose main tool in proving an SR result for $L = 2$ was the MD region for $L = 2$ of El Gamal and Cover [5]. We attack the SR problem for $L > 2$ with a new achievable MD region for $L > 2$. The MD problem and the new achievable region are described in Sections 2 and 3. The new SR results are given in Section 4.

For further context, it should be noted that Rimoldi [6] provided a rate region for SR with more than two stages. He remarked that a proof using an MD region was not possible because El Gamal and Cover's MD region had no clear extension to more than two descriptions. Using a new MD region for $L > 2$, we obtain an alternative proof for [6, Thm. 3] and are able to study SR on trees as well as chains.

2 Multiple Description Coding

Consider a source that emits a sequence $X^N = (X^{(1)}, X^{(2)}, \dots, X^{(N)})$ of N independent and identically distributed (i.i.d.) random variables. X^N is encoded and is transmitted to a receiver over L channels at rates R_1, \dots, R_L nats per source symbol. The L transmitted quantities are called the *descriptions* of the source X^N and are denoted by

$$J_l = g_l(X^N), \quad l = 1, \dots, L, \quad (1)$$

where the $g_l(\cdot)$'s are some functions. Suppose that the channels over which the descriptions are transmitted are unreliable in that any channel can break down. If all the channels are perfect, the receiver gets all L descriptions, but whenever a channel breaks down, the receiver does not get the corresponding description. In other words, the description transmitted on each channel is either transmitted error-free or lost completely. The receiver can thus encounter a total of 2^L different cases, since each of the L channels can be in one of two states. The case where all the channels are broken down is trivial. For the non-trivial cases we can represent the receiver as a collection of $2^L - 1$ decoders, where each decoder receives a non-empty subset of $\{J_1, \dots, J_L\}$. We can now pretend that the channels are perfect, since all $2^L - 1$ decoding scenarios are being considered.

Let \mathcal{L} denote the index set $\{1, \dots, L\}$ and let $2^{\mathcal{L}}$ be its power set, *i.e.*, the collection of all the subsets of \mathcal{L} . For every non-empty subset $\mathcal{K} \in 2^{\mathcal{L}}$, let $D_{\mathcal{K}}$ denote the decoder whose inputs are the descriptions indexed by \mathcal{K} , namely $\{J_k : k \in \mathcal{K}\}$. Let $X_{\mathcal{K}}^N = (X_{\mathcal{K}}^{(1)}, \dots, X_{\mathcal{K}}^{(N)})$ denote the output of $D_{\mathcal{K}}$. Then we can write

$$X_{\mathcal{K}}^N = f_{\mathcal{K}}(\{J_k : k \in \mathcal{K}\}), \quad (2)$$

for some functions $f_{\mathcal{K}}(\cdot)$. Next let $d_{\mathcal{K}}$ denote the expected distortion per source symbol associated with the output $X_{\mathcal{K}}^N$:

$$d_{\mathcal{K}} = E \left[\frac{1}{N} \sum_{n=1}^N d(X^{(n)}, X_{\mathcal{K}}^{(n)}) \right], \quad (3)$$

where $d(\cdot, \cdot)$ is a distortion measure. The set of rates $\{R_1, \dots, R_L\}$ and distortions $\{d_{\mathcal{K}} : \mathcal{K} \in 2^{\mathcal{L}} - \{\emptyset\}\}$ that are achievable in the limit $N \rightarrow \infty$ for a given source and distortion measure is called the *rate-distortion (RD) region*. The RD-region is unknown except for some simple sources for $L = 2$.

3 An Achievable Region for MD Coding

We use the following notation to simplify our presentation. Let \mathcal{C} and \mathcal{D} be collections of sets. The set difference between \mathcal{C} and \mathcal{D} is defined in the usual manner as $\mathcal{C} - \mathcal{D} \stackrel{\text{def}}{=} \{\mathcal{M} \in \mathcal{C} : \mathcal{M} \notin \mathcal{D}\}$. If $\mathcal{D} = \{\mathcal{K}\}$ is a singleton set, then we denote $\mathcal{C} - \{\mathcal{K}\}$ simply by $\mathcal{C} - \mathcal{K}$. We denote the set of random variables $\{X_{\mathcal{K}} : \mathcal{K} \in \mathcal{C}\}$ by $X_{(\mathcal{C})}$. We write $R_{\mathcal{K}}$ as shorthand for the sum of rates $\sum_{k \in \mathcal{K}} R_k$. Finally, we will often drop the braces and write $X_1, X_{12} \dots$ in place of $X_{\{1\}}, X_{\{1,2\}}$ etc. We also write X_0 to denote the central decoder output $X_{\mathcal{L}}$ and set X_{\emptyset} equal to a constant, e.g. 0.

The following is an information theoretic characterization of an achievable region for MD coding of an i.i.d. source. It is detailed in a separate publication [7].

Theorem 1 ([7]). *Consider a source that emits i.i.d. finite-alphabet random variables whose distributions are P_X . Then the RD region contains the rates and distortions satisfying*

$$d_{\mathcal{K}} \geq Ed(X, X_{\mathcal{K}}) \tag{4}$$

$$R_{\mathcal{K}} \geq -H(X_{(2^{\mathcal{K}})}|X) + \sum_{\mathcal{M} \subseteq \mathcal{K}} H(X_{\mathcal{M}}|X_{(2^{\mathcal{M}}-\mathcal{M})}) \tag{5}$$

for every $\mathcal{K} \in 2^{\mathcal{L}} - \emptyset$, where $X_{\mathcal{K}}$ is a finite-alphabet random variable.

For $L = 2$, Theorem 1 gives the achievable region of El Gamal and Cover [5]. This generalization suffices for proving the SR results in the following section. A different achievable region for $L = 2$ was determined by Zhang and Berger [8, Thm. 1].

4 Successive Refinement

The successive refinement (SR) problem is a special case of the MD coding problem of the source where the only decoder outputs we care about are $X_1, X_{12}, \dots, X_{12\dots L}$. For any source, there is an RD region, defined analogously to the RD region for MD coding. When the RD region is limited only by Shannon's single-description rate-distortion bound applied to each output, the source is called successively refinable. More explicitly, we say that X is *successively refinable on a chain* or *in L stages*, if

the RD region for the SR problem is given by

$$\begin{aligned}
d_1 &\geq D(R_1), \\
d_{12} &\geq D(R_1 + R_2), \\
&\vdots \\
R(d_{1\dots L}) &\geq D(R_1 + R_2 + \dots R_L), \\
R_l &\geq 0, \quad l \in \mathcal{L},
\end{aligned} \tag{6}$$

where $D(\cdot)$ is the distortion-rate function.

Successive refinement was first studied by Koshelev. He gave a Markovian sufficient condition in [9, 10, 11]; necessity was shown by Equitz and Cover [4].

Theorem 2. *The Gaussian source is successively refinable on a chain.*

Proof. Let X denote a unit-variance Gaussian random variable. Then the following trivial outer bound on the RD region for the SR problem follows directly from Shannon's rate-distortion theorem [1]:

$$\begin{aligned}
d_{1\dots k} &\geq e^{-2(R_1 + \dots + R_k)}, \quad k \in \mathcal{L}, \\
R_l &\geq 0, \quad l \in \mathcal{L}.
\end{aligned} \tag{7}$$

We now show that this region is achievable. For given rates $R_l \geq 0$, it suffices to demonstrate, using Theorem 1, that

$$d_{1\dots k} = e^{-2(R_1 + \dots + R_k)}, \quad k \in \mathcal{L},$$

are achievable. Notice also that $d_1 \geq d_{12} \geq \dots \geq d_{1\dots L}$. So let $X_{1\dots k}$, $k \in \mathcal{L}$ be Gaussian random variables such that

$$X \rightarrow X_{1\dots L} \rightarrow X_{1\dots(L-1)} \cdots \rightarrow X_1, \tag{8}$$

is a Markov chain and the distortion constraints (4) are satisfied: $Ed(X, X_{1\dots k}) = d_{1\dots k}$. For example let $\sigma_k^2 = d_{1\dots k}/(1 - d_{1\dots k})$ and

$$X_{1\dots k} = \frac{1}{1 + \sigma_k^2} \left(X + \sum_{l=k}^L Z_l \right),$$

for $k \in \mathcal{L}$, where Z_l 's are zero-mean Gaussian random variables independent of X and each other with

$$\text{var } Z_k = \begin{cases} \sigma_k^2 - \sigma_{k+1}^2 & \text{if } 1 \leq k < L, \\ \sigma_L^2 & \text{if } k = L. \end{cases}$$

We take all other decoder outputs to be identically zero: $X_{\mathcal{K}} = 0$ if $\mathcal{K} \neq \{1, \dots, k\}$ for any $k \in \mathcal{L}$. Consider the right hand side of the rate constraint (5) for any \mathcal{K} of

the form $\mathcal{K} = \{1, \dots, k\}$:

$$\begin{aligned}
& -h(X_{(2^\mathcal{K})}|X) + \sum_{\mathcal{M} \subseteq \mathcal{K}} h(X_{\mathcal{M}}|X_{(2^{\mathcal{M}-\mathcal{M}})}) \\
&= -h(X_1 \dots X_{1\dots k}|X) + \sum_{r=1}^k h(X_{1\dots r}|X_1 X_2 \dots X_{1\dots(r-1)}) \\
&= -\sum_{r=1}^{k-1} h(X_{1\dots r}|X_{1\dots(r+1)}) - h(X_{1\dots k}|X) + h(X_1) + \sum_{r=1}^{k-1} h(X_{1\dots(r+1)}|X_{1\dots r}) \\
&= I(X_{1\dots k}; X) = \frac{1}{2} \log(d_{1\dots k}^{-1}) = R_{\mathcal{K}},
\end{aligned}$$

where the second equality follows from (8) and the third follows by combining the two summations and using the identity $h(U|V) - h(V|U) = h(U) - h(V)$ followed by simplification. Hence the rates R_l satisfy (5) for any \mathcal{K} of the form $\{1, \dots, k\}$. To check the remaining bounds in (5), suppose that $\mathcal{K} \neq \{1, \dots, k\}$ for any $k \in \mathcal{L}$. Then let $k' \geq 1$ be the largest integer, if it exists, such that $\mathcal{K}' = \{1, \dots, k'\} \subseteq \mathcal{K}$. Otherwise take $\mathcal{K}' = \emptyset$. If \mathcal{K}' is not empty, then it is easy to verify that the right hand side of (5) for \mathcal{K} reduces to the same quantity with \mathcal{K} replaced by \mathcal{K}' , which was earlier shown to equal $R_{\mathcal{K}'}$. Since $R_{\mathcal{K}} \geq R_{\mathcal{K}'}$, (5) holds. Finally if $\mathcal{K}' = \emptyset$, the right hand side of (5) reduces to 0. Thus all the rate constraints (5) are satisfied as well, and hence the RD region is given by (7). \square

4.1 Successive Refinement on Trees

We now consider a generalization of the successive refinement problem although it is still a special case of the MD coding problem.

Definition 1. *A finite collection \mathcal{C} of sets of positive integers is said to define a tree structure if (a) for any nonempty $\mathcal{K} \in \mathcal{C}$, there is a unique $\mathcal{K}' \in \mathcal{C}$ called the parent node of \mathcal{K} and denoted by $p(\mathcal{K})$, such that $\mathcal{K}' = p(\mathcal{K}) \subset \mathcal{K}$ and $|\mathcal{K}'| = |\mathcal{K}| - 1$ where $|\cdot|$ denotes cardinality and (b) for distinct $\mathcal{K}_1, \mathcal{K}_2 \in \mathcal{C} - \emptyset$, $\mathcal{V}(\mathcal{K}_1) \neq \mathcal{V}(\mathcal{K}_2)$ where $\mathcal{V}(\mathcal{K}) = \mathcal{K} - p(\mathcal{K})$.*

The following is another way of stating the second condition. Suppose we assign the value $\mathcal{V}(\mathcal{K})$ to a branch from $\mathcal{K}' = p(\mathcal{K})$ to \mathcal{K} , then distinct branches in the tree have different values. An example of a tree structure is shown in Figure 4 for the case

$$\mathcal{C} = \{\emptyset, \{1\}, \{2\}, \{1, 3\}, \{1, 4\}, \{2, 5\}\}.$$

The values of the function $\mathcal{V}(\mathcal{K}) = \mathcal{K} - p(\mathcal{K})$ are indicated on the branches from $p(\mathcal{K})$ to \mathcal{K} . Without loss of generality, we assume that

$$\bigcup_{\mathcal{K} \in \mathcal{C}} \mathcal{K} = \{1, \dots, L\} = \mathcal{L}.$$

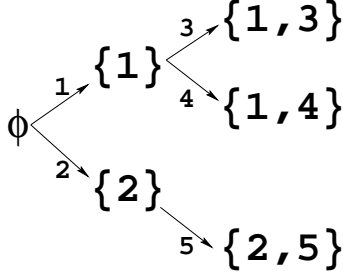


Figure 4: Example of a tree structure

Theorem 3. *Suppose that X is a successively refinable on a chain and \mathcal{C} defines a tree-structure. Then the RD region for $X_{\mathcal{K}}$, $\mathcal{K} \in \mathcal{C}$ is given by the set of constraints*

$$d_{\mathcal{K}} \geq D(R_{\mathcal{K}}), \quad \mathcal{K} \in \mathcal{C}, \quad (9)$$

where $D(\cdot)$ is the distortion-rate function, i.e. X is successively refinable on trees.

Proof. First observe that (9) is an outer bound by Shannon's rate-distortion theorem. Therefore, it suffices to show that, for given rates R_l , we can achieve $d_{\mathcal{K}}$ that satisfy $R_{\mathcal{K}} = R(d_{\mathcal{K}})$. Let $\mathcal{K}_0, \mathcal{K}_1, \dots, \mathcal{K}_M$ be an ordering of the elements of \mathcal{C} such that $\mathcal{K}_0 = \emptyset$ and

$$d_{\mathcal{K}_1} \geq d_{\mathcal{K}_2} \geq \dots \geq d_{\mathcal{K}_M} \iff R_{\mathcal{K}_1} \leq R_{\mathcal{K}_2} \leq \dots \leq R_{\mathcal{K}_M},$$

with the additional constraint that all subsets of \mathcal{K} in \mathcal{C} precede \mathcal{K} . We write $\mathcal{K} \prec \mathcal{M}$, if \mathcal{K} precedes \mathcal{M} in the list. Then $p(\mathcal{K}) \prec \mathcal{K}$. Now clearly

$$\bigcup_{\mathcal{K} \in \mathcal{C} - \emptyset} \mathcal{V}(\mathcal{K}) = \bigcup_{\mathcal{K} \in \mathcal{C}} \mathcal{K} = \mathcal{L}. \quad (10)$$

Therefore for every $l \in \mathcal{L}$, we can find a unique \mathcal{K} such that $\mathcal{V}(\mathcal{K}) = \{l\}$ because the left hand side of (10) is a disjoint union. We denote this set by $\mathcal{K} = \mathcal{V}^{-1}(\{l\})$. Next for $1 \leq m \leq M$ let

$$\tilde{R}_m = R_{\mathcal{K}_m} - R_{\mathcal{K}_{m-1}} \geq 0. \quad (11)$$

Now X is successively refinable. Therefore, for any $\epsilon > 0$, we can find an integer N , descriptions \tilde{J}_m , $m = 1, \dots, M$ of X^N and functions $\tilde{f}_m(\cdot)$ such that

$$H(\tilde{J}_m) \leq N\tilde{R}_m \quad (12)$$

$$\frac{1}{N}Ed(X^N, \tilde{X}_{1\dots m}^N) \leq D(\tilde{R}_1 + \dots + \tilde{R}_m) + \epsilon \quad (13)$$

$$\text{where } \tilde{X}_{1\dots m}^N = \tilde{f}_m(\tilde{J}_1, \dots, \tilde{J}_m). \quad (14)$$

For any $l \in \mathcal{L}$, define

$$J_l = \{\tilde{J}_m, \forall m : p(\mathcal{K}) \prec \mathcal{K}_m \preceq \mathcal{K}\} \quad \text{where } \mathcal{K} = \mathcal{V}^{-1}(\{l\}). \quad (15)$$

Then we claim that the following two sets of random variables are identical:

$$\{J_l : l \in \mathcal{K}\} = \{\tilde{J}_m, \forall m : \mathcal{K}_m \preceq \mathcal{K}\}. \quad (16)$$

We prove this by induction. The result holds trivially for the first element in the list $\mathcal{K}_0 = \emptyset$. Suppose that the result holds for every $\mathcal{K}' \prec \mathcal{K}$. Then

$$\{J_l : l \in p(\mathcal{K})\} = \{\tilde{J}_m, \forall m : \mathcal{K}_m \preceq p(\mathcal{K})\} \quad (17)$$

because $p(\mathcal{K}) \prec \mathcal{K}$. Now $\mathcal{V}(\mathcal{K}) = \{l_0\}$ for some l_0 . Then $\mathcal{K} = p(\mathcal{K}) \cup \{l_0\}$ and hence

$$\begin{aligned} (J_l : l \in \mathcal{K}) &= (J_{l_0} \text{ and } J_l : l \in p(\mathcal{K})) \\ &= (\tilde{J}_m, \forall m : p(\mathcal{K}) \prec \mathcal{K}_m \preceq \mathcal{K} \text{ and } \tilde{J}_m, \forall m : \mathcal{K}_m \preceq p(\mathcal{K})) \\ &\equiv (\tilde{J}_m, \forall m : \mathcal{K}_m \preceq \mathcal{K}), \end{aligned}$$

where the second equality follows from (15) and (17). This result (16) implies that

$$\{J_l : l \in \mathcal{K}_m\} = \{\tilde{J}_{m'}, \forall m' : \mathcal{K}_{m'} \preceq \mathcal{K}_m\} = \{\tilde{J}_1, \dots, \tilde{J}_m\}, \quad 1 \leq m \leq M. \quad (18)$$

We take $X_{\mathcal{K}_m}^N = \tilde{X}_{1\dots m}$ so that, in view of (14) and (18), $X_{\mathcal{K}_m}^N$ is some function of the descriptions $J_l, l \in \mathcal{K}_m$:

$$X_{\mathcal{K}_m}^N = f_m(J_l, l \in \mathcal{K}_m). \quad (19)$$

We use (11), (12) and (15) to argue, for $l \in \mathcal{L}$, that

$$\begin{aligned} H(J_l) &\leq \sum_{m:p(\mathcal{K}) \prec \mathcal{K}_m \preceq \mathcal{K}} H(\tilde{J}_m) \quad \text{where } \mathcal{K} = \mathcal{V}^{-1}(\{l\}) \\ &\leq \sum_{m:p(\mathcal{K}) \prec \mathcal{K}_m \preceq \mathcal{K}} N \tilde{R}_m \\ &= N(R_{\mathcal{K}(l)} - R_{p(\mathcal{K}(l))}) = NR_l. \end{aligned} \quad (20)$$

Next we use (11) and (13) to deduce that

$$\begin{aligned} \frac{1}{N} Ed(X^N, X_{\mathcal{K}_m}^N) &\leq D(\tilde{R}_1 + \dots + \tilde{R}_m) + \epsilon \\ &= D(R_{\mathcal{K}_m}) + \epsilon \\ &= d_{\mathcal{K}_m} + \epsilon, \end{aligned} \quad (21)$$

for $1 \leq m \leq M$. Finally (19), (20) and (21) imply that the rates R_l and distortions $d_{\mathcal{K}}, \mathcal{K} \in \mathcal{C}$ are achievable. \square

Corollary 1. *The Gaussian source is successively refinable on trees.*

References

- [1] C. E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:379–423, July 1948. Continued 27:623–656, October 1948.
- [2] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Inform. Th.*, IT-19(4):471–480, July 1973.
- [3] T. M. Cover. Broadcast channels. *IEEE Trans. Inform. Th.*, IT-18(1):2–14, January 1972.
- [4] W. H. R. Equitz and T. M. Cover. Successive refinement of information. *IEEE Trans. Inform. Th.*, 37(2):269–275, March 1991. See also [12].
- [5] A. A. El Gamal and T. M. Cover. Achievable rates for multiple descriptions. *IEEE Trans. Inform. Th.*, IT-28(6):851–857, November 1982.
- [6] B. Rimoldi. Successive refinement of information: Characterization of the achievable rates. *IEEE Trans. Inform. Th.*, 40(1):253–259, January 1994.
- [7] R. Venkataramani, G. Kramer, and V. K Goyal. Bounds on the achievable region for certain multiple description coding problems. In *Proc. IEEE Int. Symp. Inform. Th.*, Washington, DC, June 2001. Submitted.
- [8] Z. Zhang and T. Berger. New results in binary multiple descriptions. *IEEE Trans. Inform. Th.*, IT-33(4):502–521, July 1987.
- [9] V. Koshelev. Multilevel source coding and data-transmission theorem. In *Proc. VII All-Union Conf. on Theory of Coding and Data Transm.*, volume 1, pages 85–92, Vilnius, USSR, 1978.
- [10] V. Koshelev. Hierarchical coding of discrete sources. *Probl. Peredach. Inform.*, 16(3):31–49, 1980.
- [11] V. Koshelev. An evaluation of the average distortion for discrete scheme of sequential approximation. *Probl. Peredach. Inform.*, 17(3):20–33, 1981.
- [12] W. H. R. Equitz and T. M. Cover. Addendum to “Successive refinement of information”. *IEEE Trans. Inform. Th.*, 39(4):1465–1466, July 1993.