The Gaussian Many-Help-One Distributed Source Coding Problem

Saurabha Tavildar, Pramod Viswanath, *Member, IEEE*, and Aaron B. Wagner, *Member, IEEE*

***Abstract*—Jointly Gaussian memoryless sources are observed at N distinct terminals. The goal is to efficiently encode the observations in a distributed fashion so as to enable reconstruction of any one of the observations, say the first one, at the decoder subject to a quadratic fidelity criterion. Our main result is a *precise* characterization of the rate–distortion region when the covariance matrix of the sources satisfies a "tree-structure" condition. In this situation, a natural analog–digital separation scheme optimally trades off the distributed quantization rate tuples and the distortion in the reconstruction: each encoder consists of a point-to-point Gaussian vector quantizer followed by a Slepian–Wolf binning encoder. We also provide a partial converse that suggests that the tree-structure condition is fundamental.**

***Index Terms*—Entropy power inequality, Gaussian sources, many-help-one problem, network source coding, rate distortion, tree sources.**

I. Introduction

THE focus of this study is the problem of distributed source coding of memoryless Gaussian sources with quadratic distortion constraints. The rate–distortion region of this problem with two terminals has been recently characterized [14]. Our focus, hence, is on the case when there are at least three terminals. In this paper, we study a special case of this general problem: the so-called "many-help-one" situation depicted in Fig. 1. The setup is the following.

- *Sources*: Each of the N encoders observes a memoryless discrete-time source: encoder i observes, over n discrete time instants, the memoryless source x_i^n. The observations across the encoders are correlated, however. Specifically, the joint observations at time m $(x_1(m), \dots, x_N(m))$ are jointly Gaussian. Further, the joint observations are memoryless over time m.
- *Encoders*: Each encoder i maps the vector of analog observations (over n time instants, say) into a vector of bits (of length $R_i n$, say) that is then communicated without loss to a single decoder (on a link with rate R_i).

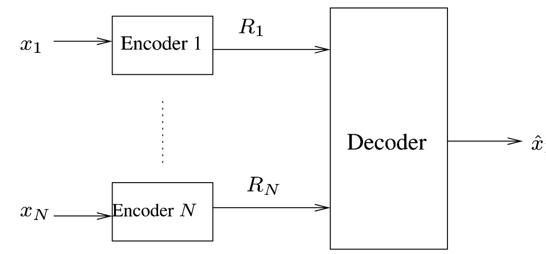

Fig. 1. The many-help-one problem.

- *Decoder*: The decoder is only interested in reconstructing one of the sources (say, x_1^n). The fidelity criterion considered here is a quadratic one: the average (over the statistics of the sources) l_2 distance between the original source vector and the reconstructed vector is required to be no more than Dn.
- *Problem statement*: The problem is to characterize the minimum set of rates at which the encoders can communicate with the decoder while still conveying enough information to satisfy the quadratic distortion constraint on the reconstruction.

In this paper, we precisely characterize the rate–distortion region of a class of many-help-one problems: those for which the sources can be embedded in a Gauss–Markov tree. A crucial step towards solving this problem involves the introduction of a related distributed source coding problem where the source has a "binary tree" structure; this is done in Section II. We show that the natural analog–digital separation strategy of point-to-point Gaussian vector quantization followed by a distributed Slepian–Wolf binning scheme is optimal for this problem (this is done in Sections II-C and II-D). Next, we show how this result can be used to solve various instances of the many-help-one problem of interest; this is done in Section III. Finally, ancillary aspects of the problem at hand are discussed in Section IV: specifically, the worst case property of the Gaussian distribution with respect to the analog–digital separation architecture is demonstrated and a partial converse for the necessity of the tree-structure condition is provided.

II. The Binary Tree Structure Problem

In this section, we take a short detour away from the many-help-one problem of interest (cf. Fig. 1). Specifically, we introduce a related distributed source coding problem that we call the "binary tree-structure problem." We show that the natural analog–digital separation architecture is optimal in terms of the rate–distortion tradeoff for this problem. The connection between the original many-help-one problem and this binary tree-structure problem is made in the next section.

Manuscript received January 19, 2008; revised May 28, 2009. Current version published December 23, 2009. This work was supported by the National Science Foundation under Grants CCF 07-29075 and CCF 06-42925 (CAREER) and by the Air Force Office of Scientific Research under Grant FA9550-08-1-0060.

S. Tavildar is with Qualcomm Flarion Technologies, Bridgewater, NJ 08807 USA (e-mail: tavildar@gmail.com).

P. Viswanath is with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: pramodv@uiuc.edu).

A. B. Wagner is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: wagner@ece.cornell.edu).

Communicated by E.-h. Yang, Associate Editor for Source Coding.

Color versions of Figures 6–8 in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2009.2034791

The outline of this section is as follows:

- we introduce the source variables and their statistical relationships first (Section II-A);
- next we specify precisely the binary tree-structure problem (Section II-B);
- we evaluate the performance of the natural analog–digital architecture in terms of the rate–distortion tradeoff for the binary tree-structure problem (Section II-C);
- under the assumption that certain variables have positive variance, we derive a novel outer bound to the rate–distortion region—this involves a careful use of the entropy-power inequality (extracting critical ideas from [8], [10]) (Section II-D);
- again under the positive variance assumption, we show that the outer bound to the rate–distortion region indeed matches the inner bound derived by evaluating the natural analog–digital separation architecture—this result and the previous one are the most important technical contributions of the paper (Section II-D);
- using a continuity argument, we relax the positive variance assumption and show that the separation architecture is optimal for all binary tree-structure problems (Section II-E);
- finally, we show that Gaussian sources are the worst case in the sense that a non-Gaussian source has a larger rate–distortion region than a Gaussian source with the same covariance matrix, so long as the Gaussian source satisfies the tree structure (Section II-F).

A. Binary Gauss–Markov Trees

Consider the Markov binary tree structure of Gaussian random variables depicted in Fig. 2. Formally, the Gauss–Markov tree structure represents the following Markov chain conditions: consider the node denoted by the random variable $x_i^{(k)}$. We define the set of left descendants, the set of right descendants, and the tree of $x_i^{(k)}$ to be

$$\mathcal{L}\left(x_i^{(k)}\right) = \left\{x_j^{(l)} : l > k, \frac{2^l(i-1)}{2^k} < j \le \frac{2^l(i-0.5)}{2^k}\right\}$$
$$\mathcal{R}\left(x_i^{(k)}\right) = \left\{x_j^{(l)} : l > k, \frac{2^l(i-0.5)}{2^k} < j \le \frac{2^l i}{2^k}\right\}$$
$$\mathcal{T}\left(x_i^{(k)}\right) = \left\{x_i^{(k)}\right\} \cup \mathcal{R}\left(x_i^{(k)}\right) \cup \mathcal{L}\left(x_i^{(k)}\right)$$

respectively. We define the set of nodes $\mathcal{P}(x_i^{(k)})$ to be

$$\left\{x_j^{(l)} : \forall j, l\right\} \backslash \mathcal{T}\left(x_i^{(k)}\right).$$

Then, by definition, the Markov chain condition given by Fig. 2 says that conditioned on the random variable $x_i^{(k)}$, the sets of random variables $\mathcal{P}(x_i^{(k)})$, $\mathcal{L}(x_i^{(k)})$, and $\mathcal{R}(x_i^{(k)})$ are independent; further, this is true for all pairs (i, k).

1) A Specific Construction: Now consider the following specific construction of $x_i^{(k)}$'s that satisfies the Markov chain structure in Fig. 2. Let m, k, and i denote the time index, the tree depth index, and the node within the tree depth index, respectively. Then define

$$x_{2i-1}^{(k+1)}(m) = \alpha_{2i-1}^{(k+1)} x_i^{(k)}(m) + n_{2i-1}^{(k+1)}(m) \qquad (1)$$
$$x_{2i}^{(k+1)}(m) = \alpha_{2i}^{(k+1)} x_i^{(k)}(m) + n_{2i}^{(k+1)}(m) \qquad (2)$$

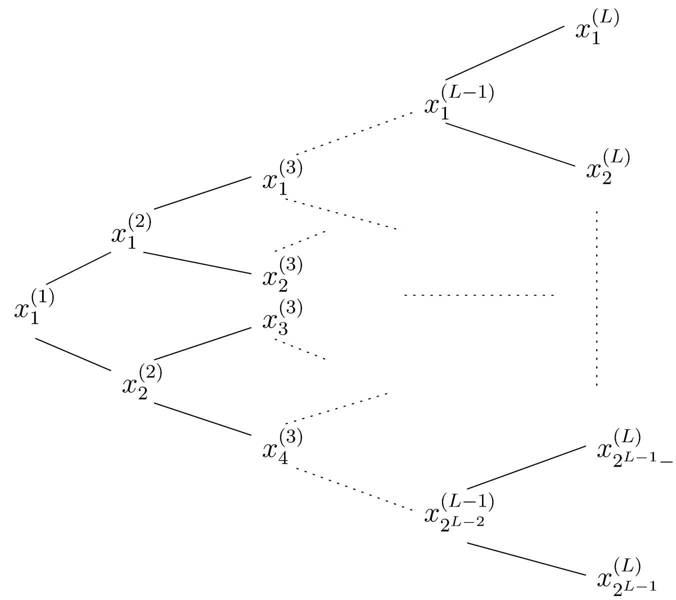

Fig. 2. The binary tree structure.

where the indices vary as

$$m = 1, \ldots, n, \qquad (3)$$
$$k = 1, \ldots, L-1 \qquad (4)$$
$$i = 1, \ldots, 2^{k-1}. \qquad (5)$$

Here $\alpha_{2i-1}^{(k+1)}$ and $\alpha_{2i}^{(k+1)}$ are real numbers. Also define

$$x_1^{(1)}(m) = n_1^{(1)}(m).$$

The random variables

$$\left\{n_i^{(k)}(m), \quad k = 1, \ldots, L, i = 1, \ldots, 2^{k-1}, \quad m = 1, \ldots, n\right\} \qquad (6)$$

are independent Gaussian random variables (with zero mean and variance $\sigma^2_{n_i^{(k)}}$ for the index pair (k, i) and any m).

For the special case of $L = 3$, these formulas read

$$x_1^{(1)} = n_1^{(1)}$$
$$\Longrightarrow \begin{cases} x_1^{(2)} = \alpha_1^{(2)} x_1^{(1)} + n_1^{(2)} \Longrightarrow \begin{cases} x_1^{(3)} = \alpha_1^{(3)} x_1^{(2)} + n_1^{(3)} \\ x_2^{(3)} = \alpha_2^{(3)} x_1^{(2)} + n_2^{(3)} \end{cases} \\ x_2^{(2)} = \alpha_2^{(2)} x_1^{(1)} + n_2^{(2)} \Longrightarrow \begin{cases} x_3^{(3)} = \alpha_3^{(3)} x_2^{(2)} + n_3^{(3)} \\ x_4^{(3)} = \alpha_4^{(3)} x_2^{(2)} + n_4^{(3)}. \end{cases} \end{cases}$$

For any L, the source variables satisfy the binary tree structure in Fig. 2 if and only if they can be written in this form.

Proposition 1: For this construction, the $x_i^{(k)}$ satisfy the Markov chain conditions in Fig. 2. Conversely, any zero-mean, jointly Gaussian $\{x_i^{(k)}, \quad k = 1, \ldots, L, i = 1, \ldots, 2^{k-1}\}$ that satisfy the Markov tree structure in Fig. 2 can be represented using the above construction (cf. (1) and (2)).

B. Binary-Tree Problem Statement

Denote the vector

$$x_{1,n}^{(1)} \stackrel{\text{def}}{=} \left(x_1^{(1)}(1), \ldots, x_1^{(1)}(n)\right). \qquad (7)$$

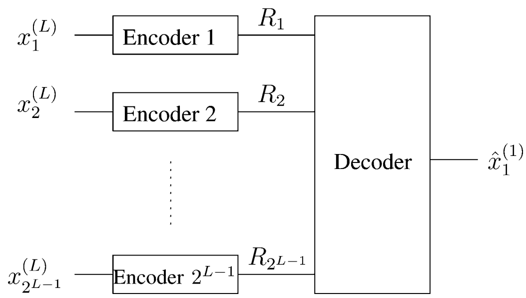

Fig. 3. The problem setup.

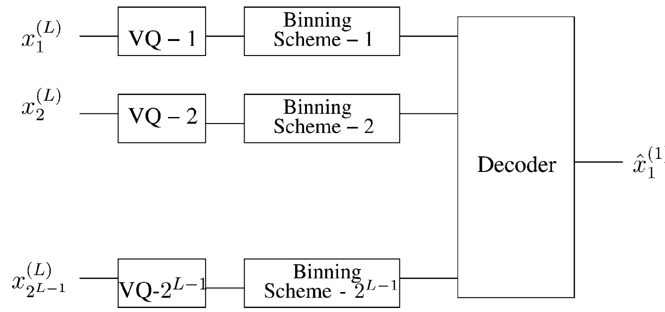

Fig. 4. The natural separation scheme.

Similar notation will be used for other vectors to be introduced later. Consider the following distributed source coding problem depicted in Fig. 3: There are 2^{L-1} distributed encoders each having access to a memoryless observation sequence: encoder i observes the memoryless random process $x_{i,n}^{(L)}$. The goal of each encoder is to map the observation into a *discrete* set (encoder i maps its length-n observation into a discrete set C_i). The encoded observation is then conveyed to the central decoder on rate-constrained links. The rate of communication from encoder i to the decoder is

$$\frac{1}{n}\log|C_i|.$$

The decoder forms an estimate $\hat{x}_{1,n}^{(1)}$ of the *root* of the binary tree, $x_{1,n}^{(1)}$, based on the messages $C_1,\dots,C_{2^{L-1}}$. The average distortion of the reconstruction is

$$\frac{1}{n}\sum_{m=1}^{n}\mathbb{E}\left[\left(x_1^{(1)}(m)-\hat{x}_1^{(1)}(m)\right)^2\right].$$

The goal is to characterize the set of achievable rates and distortions $(R_1,\dots,R_{2^{L-1}},d)$, i.e., those such that there exists an encoder and decoder such that

$$R_i \geq \frac{1}{n}\log|C_i| \quad \text{for all } i$$

and

$$d \geq \frac{1}{n}\sum_{m=1}^{n}\mathbb{E}\left[\left(x_1^{(1)}(m)-\hat{x}_1^{(1)}(m)\right)^2\right].$$

We denote the closure of this set by $\mathcal{RD}^*$.

We note that two special cases of this problem have been resolved in the literature:

- $L=1$ is the single-user Gaussian source coding problem with quadratic distortion;
- $L=2$ is the Gaussian CEO problem solved in [8], [7], [10].

The recent work in [9] studies a special case of the general tree structure depicted in Fig. 2.[1] While a general outer bound is derived in [9] for that special case of the tree structure, it is shown to be tight only for a certain range of the parameters in the problem (the distortion constraint and the covariance matrix of the Gaussian sources).

Our main result is that a natural strategy of point-to-point Gaussian vector quantization followed by Slepian–Wolf binning is optimal for any L. In the next section, we formally present the natural achievable strategy and then state our main result. In the subsequent section, we prove a novel outer bound and use it to establish the main result.

[1]As an aside, we note that the material in [9] along with our own previous work [14] provided the impetus to the present work.

C. Analog–Digital Separation Strategy

The natural achievable analog–digital separation strategy is depicted in Fig. 4: each encoder first vector quantizes the observation as in point-to-point Gaussian rate distortion theory, and then codes the quantizer outputs using a Slepian–Wolf binning scheme. The rate tuples needed by this architecture to satisfy the distortion constraint can be calculated by the so-called Berger–Tung inner bound [1]: let

$$\boldsymbol{u} \stackrel{\text{def}}{=} (u_1,u_2,\dots,u_{2^{L-1}}) \tag{8}$$

denote a vector of 2^{L-1} jointly Gaussian random variables. Consider the set $\mathcal{U}(d)$ of $\boldsymbol{u}$ such that

- for each $i=1,\dots,2^{L-1}$, u_i satisfies
$$u_i = \alpha_i x_i^{(L)} + w_i \tag{9}$$
where $\alpha_1,\dots,\alpha_{2^{L-1}}$ are constants and $w_1,\dots,w_{2^{L-1}}$ are independent zero-mean Gaussian random variables that are also independent of the $x_i^{(k)}$'s;
- $\boldsymbol{u}$ satisfies
$$\mathbb{E}\left[\left(x_1^{(1)}-\mathbb{E}\left[x_1^{(1)}\middle|\boldsymbol{u}\right]\right)^2\right]\leq d. \tag{10}$$

Now, consider

$$\mathcal{A}\subseteq\{1,\dots,2^{L-1}\}. \tag{11}$$

Denote the set

$$\{u_i : i\in\mathcal{A}\} \stackrel{\text{def}}{=} \boldsymbol{u}_{\mathcal{A}}. \tag{12}$$

Similar notation will be used for other vectors introduced later. We now have the following.

Lemma 1: [Berger–Tung inner bound [1]] The analog–digital separation architecture achieves the convex hull of the region

$$\mathcal{RD}_{\text{in}} \stackrel{\text{def}}{=} \Bigg\{ (R_1,\dots,R_{2^{L-1}},d) : \exists \boldsymbol{u}\in\mathcal{U}(d) : \forall\mathcal{A}\subseteq\{1,\dots,2^{L-1}\}, \sum_{i\in\mathcal{A}} R_i \geq I\left(\boldsymbol{x}_{\mathcal{A}}^{(L)};\boldsymbol{u}_{\mathcal{A}}\middle|\boldsymbol{u}_{\mathcal{A}^c}\right)\Bigg\}. \tag{13}$$

In particular, $\mathcal{RD}^*$ contains $\mathrm{Co}(\mathcal{RD}_{\text{in}})$, where $\mathrm{Co}(\cdot)$ denotes the closure of the convex hull.

The region $\mathcal{RD}_{\text{in}}$ can be explicitly computed for a given covariance matrix for the observed Gaussian sources. This computation is aided by the following combinatorial structure of the set $\mathcal{RD}_{\text{in}}$.

1) Combinatorial Structure of $\mathcal{RD}_{\text{in}}$: Consider a specific $\boldsymbol{u} \in \mathcal{U}(d)$ (this parameterizes a specific choice of the analog–digital separation architecture) and the rate tuples $(R_1, \dots, R_{2^{L-1}})$ that satisfy the conditions

$$\sum_{i \in \mathcal{A}} R_i \geq f(\mathcal{A}), \quad \forall \mathcal{A} \subseteq \{1, \dots, 2^{L-1}\} \tag{14}$$

where the function f is defined by

$$f(\mathcal{A}) \stackrel{\text{def}}{=} I\left(\boldsymbol{x}_{\mathcal{A}}^{(L)}; \boldsymbol{u}_{\mathcal{A}} \middle| \boldsymbol{u}_{\mathcal{A}^c}\right) \tag{15}$$

$$f(\Phi) \stackrel{\text{def}}{=} 0 \tag{16}$$

where Φ denotes the empty set and f satisfies the following properties [3]:

$$f(\mathcal{A}_1) \geq 0, \tag{17}$$

$$f(\mathcal{A}_1 \cup \{t\}) \geq f(\mathcal{A}_1), \quad \forall t \in \{1, \dots, 2^{L-1}\} \tag{18}$$

$$f(\mathcal{A}_1 \cup \mathcal{A}_2) + f(\mathcal{A}_1 \cap \mathcal{A}_2) \geq f(\mathcal{A}_1) + f(\mathcal{A}_2). \tag{19}$$

A polyhedron such as the one in (14) with the *rank* function f satisfying (17)–(19) is called a *contra-polymatroid*. A generic reference to the class of polyhedrons called *matroids* is [16] and applications to information theory are in [12] where natural achievable regions of the multiple-access channel are shown to be *polymatroids* and in [3], [15] where natural achievable regions for source-coding problems are shown to be contra-polymatroids. An important property of contra-polymatroids is summarized in [12, Lemma 3.3]: the characterization of its vertices. For π, a permutation on the set $\{1, \dots, 2^{L-1}\}$, let

$$b_{\pi_i}^{(\pi)} \stackrel{\text{def}}{=} f(\{\pi_1, \pi_2, \dots, \pi_i\}) - f(\{\pi_1, \pi_2, \dots, \pi_{i-1}\}), \quad i = 1 \dots 2^{L-1}$$

and $\boldsymbol{b}^{(\pi)} = (b_{\pi_1}^{(\pi)}, \dots, b_{\pi_{2^{L-1}}}^{(\pi)})$. Then the $2^{L-1}!$ points $\{\boldsymbol{b}^{(\pi)}, \pi \text{ a permutation}\}$ are the vertices of (and hence belong to) the contra-polymatroid (14). We use this result to conclude that all of the constraints in (14) are tight for some rate tuple and there is a computationally simple way to find the vertex that leads to a minimal linear functional of the rates [12].

D. An Outer Bound for a Special Case

We first focus on the case in which $\sigma^2_{n_i^{(k)}} > 0$ for all i and k. We abbreviate this condition by saying that "all of the noise variances are positive." To derive our outer bound, we need the following definitions.

- To each node $x_i^{(k)}$ in the binary tree of Fig. 2 we associate a nonnegative number $r_i^{(k)}$. When analyzing a particular code, we shall set

$$r_i^{(k)} = \frac{1}{n} I\left(x_{i,n}^{(k)}; \boldsymbol{C} \middle| x^{(k-1)}_{\lfloor \frac{i+1}{2} \rfloor, n}\right) \tag{20}$$

$$= \frac{1}{n} I\left(n_{i,n}^{(k)}; \boldsymbol{C} \middle| x^{(k-1)}_{\lfloor \frac{i+1}{2} \rfloor, n}\right) \tag{21}$$

where $\boldsymbol{C}$ is the tuple of messages sent to the decoder (throughout we adopt the convention $x_{1,n}^{(0)} \equiv 0$). Then $r_i^{(k)}$ can be physically interpreted as a *noise-quantization rate*.

- Fix $1 \leq k \leq L-1$ and $1 \leq i \leq 2^{k-1}$ and define the function

$$f_{x_i^{(k)}}(r_1, r_2) \stackrel{\text{def}}{=} \frac{1}{2} \log \left(1 + \frac{\left\{\alpha_{2i-1}^{(k+1)}\right\}^2 \sigma^2_{n_i^{(k)}}}{\sigma^2_{n_{2i-1}^{(k+1)}}} (1 - e^{-2r_1}) + \frac{\left\{\alpha_{2i}^{(k+1)}\right\}^2 \sigma^2_{n_i^{(k)}}}{\sigma^2_{n_{2i}^{(k+1)}}} \left(1 - e^{-2r_2}\right)\right), \quad r_1, r_2 \geq 0. \tag{22}$$

This complicated formula admits the following simple interpretation: if each C_i in (21) is the output of an independent and identically distributed (i.i.d.) test channel, then we have

$$r_i^{(k)} = f_{x_i^{(k)}}\left(r_{2i-1}^{(k+1)}, r_{2i}^{(k+1)}\right) \tag{23}$$

(see Appendix A). When $L = 3$, there are three such f functions to consider

$$f_{x_1^{(1)}}\left(r_1^{(2)}, r_2^{(2)}\right) = \frac{1}{2} \log \left(1 + \frac{\left\{\alpha_1^{(2)}\right\}^2 \sigma^2_{n_1^{(1)}}}{\sigma^2_{n_1^{(2)}}} \left(1 - e^{-2r_1^{(2)}}\right) + \frac{\left\{\alpha_2^{(2)}\right\}^2 \sigma^2_{n_1^{(1)}}}{\sigma^2_{n_2^{(2)}}} \left(1 - e^{-2r_2^{(2)}}\right)\right)$$

$$f_{x_1^{(2)}}\left(r_1^{(3)}, r_2^{(3)}\right) = \frac{1}{2} \log \left(1 + \frac{\left\{\alpha_1^{(3)}\right\}^2 \sigma^2_{n_1^{(2)}}}{\sigma^2_{n_1^{(3)}}} \left(1 - e^{-2r_1^{(3)}}\right) + \frac{\left\{\alpha_2^{(3)}\right\}^2 \sigma^2_{n_1^{(2)}}}{\sigma^2_{n_2^{(3)}}} \left(1 - e^{-2r_2^{(3)}}\right)\right)$$

$$f_{x_2^{(2)}}\left(r_3^{(3)}, r_4^{(3)}\right) = \frac{1}{2} \log \left(1 + \frac{\left\{\alpha_3^{(3)}\right\}^2 \sigma^2_{n_2^{(2)}}}{\sigma^2_{n_3^{(3)}}} \left(1 - e^{-2r_3^{(3)}}\right) + \frac{\left\{\alpha_4^{(3)}\right\}^2 \sigma^2_{n_2^{(2)}}}{\sigma^2_{n_4^{(3)}}} \left(1 - e^{-2r_4^{(3)}}\right)\right).$$

- For node $x_i^{(k)}$, we define the set of associated *observations* to be

$$\mathcal{O}\left(x_i^{(k)}\right) = \left\{j : \frac{2^L(i-1)}{2^k} < j \leq \frac{2^L i}{2^k}\right\}. \tag{24}$$

For $L = 3$, this reads

$$\mathcal{O}\left(x_1^{(1)}\right) = \{1,2,3,4\}$$
$$\mathcal{O}\left(x_1^{(2)}\right) = \{1,2\}$$
$$\mathcal{O}\left(x_2^{(2)}\right) = \{3,4\}$$
$$\mathcal{O}\left(x_j^{(3)}\right) = \{j\}, \quad j = 1,\dots,4.$$

- For each node $x_i^{(k)}$ define the set $\boldsymbol{r}_{\mathcal{A},\mathcal{A}^c}(x_i^{(k)})$ to be the set of noise-quantization rates (say, $r_j^{(l)}$) of the variables (say $x_j^{(l)}$) in the tree of $x_i^{(k)}$ whose associated observations are entirely in $\mathcal{A}$ or $\mathcal{A}^c$ and are such that none of the ancestors of $x_j^{(l)}$ have this property. Formally

$$\begin{aligned}\boldsymbol{r}_{\mathcal{A},\mathcal{A}^c}&\left(x_i^{(k)}\right)\\ &= \Big\{ r_j^{(l)} : x_j^{(l)} \in \mathcal{T}\left(x_i^{(k)}\right), \mathcal{O}\left(x_j^{(l)}\right) \subset \mathcal{A} \\ &\qquad \text{or } \mathcal{O}\left(x_j^{(l)}\right) \subset \mathcal{A}^c \\ &\qquad \nexists x_a^{(b)} \in \mathcal{T}\left(x_i^{(k)}\right) \text{ with } \mathcal{O}\left(x_a^{(b)}\right) \subset \mathcal{A} \\ &\qquad \text{or } \mathcal{O}\left(x_a^{(b)}\right) \subset \mathcal{A}^c, \\ &\qquad \text{and } x_j^{(l)} \in \mathcal{R}\left(x_a^{(b)}\right) \cup \mathcal{L}(x_a^{(b)}) \Big\}.\end{aligned}$$

Likewise, we let $\boldsymbol{r}_{\mathcal{A}}(x_i^{(k)})$ denote the set of noise-quantization rates of variables in the tree of $x_i^{(k)}$ whose associated observations are entirely in $\mathcal{A}$ and are such that none of the ancestors have this property. Formally

$$\begin{aligned}\boldsymbol{r}_{\mathcal{A}}\left(x_i^{(k)}\right) &= \Big\{ r_j^{(l)} : x_j^{(l)} \in \mathcal{T}\left(x_i^{(k)}\right), \mathcal{O}\left(x_j^{(l)}\right) \subset \mathcal{A} \\ &\qquad \nexists x_a^{(b)} \in \mathcal{T}\left(x_i^{(k)}\right) \text{ with } \mathcal{O}\left(x_a^{(b)}\right) \subset \mathcal{A}, \\ &\qquad \text{and } x_j^{(l)} \in \mathcal{R}\left(x_a^{(b)}\right) \cup \mathcal{L}\left(x_a^{(b)}\right) \Big\}. \quad (25)\end{aligned}$$

- Define the following set of noise-quantization rates $(r_i^{(k)}, 1 \le k \le L, 1 \le i \le 2^{k-1})$:

$$\begin{aligned}\mathcal{F}_r(d) = \Bigg\{ & r_i^{(k)} \ge 0, r_1^{(1)} \ge \frac{1}{2}\log\frac{\sigma^2_{x_1^{(1)}}}{d}, \\ & r_i^{(k)} \le f_{x_i^{(k)}}\left(r_{2i-1}^{(k+1)}, r_{2i}^{(k+1)}\right) \Bigg\}. \quad (26)\end{aligned}$$

For $L = 3$, this defines the following set:

$$\mathcal{F}_r(d) = \Bigg\{ \left(r_1^{(1)}, r_1^{(2)}, r_2^{(2)}, r_1^{(3)}, r_2^{(3)}, r_3^{(3)}, r_4^{(3)}\right) : \quad (27)$$
$$r_1^{(1)} \le f_{x_1^{(1)}}\left(r_1^{(2)}, r_2^{(2)}\right) \quad (28)$$
$$r_1^{(2)} \le f_{x_1^{(2)}}\left(r_1^{(3)}, r_2^{(3)}\right) \quad (29)$$
$$r_2^{(2)} \le f_{x_2^{(2)}}\left(r_3^{(3)}, r_4^{(3)}\right) \quad (30)$$
$$r_1^{(1)} \ge \frac{1}{2}\log\left(\frac{\sigma^2_{x_1^{(1)}}}{d}\right) \Bigg\}. \quad (31)$$

- We next implicitly define a collection of functions of the noise-quantization rates. Consider a set of noise-quantization rates $(r_i^{(k)}, 1 \le k \le L, 1 \le i \le 2^{k-1})$ in $\mathcal{F}_r(d)$. Then for any i and k, we have

$$r_i^{(k)} \le f_{x_i^{(k)}}\left(r_{2i-1}^{(k+1)}, r_{2i}^{(k+1)}\right).$$

Since $f_{x_i^{(k)}}$ is increasing in both arguments, this implies

$$r_i^{(k)} \le f_{x_i^{(k)}}\left(r_{2i-1}^{(k+1)}, f_{x_{2i}^{(k+1)}}\left(r_{4i-1}^{(k+2)}, r_{4i}^{(k+2)}\right)\right)$$
$$r_i^{(k)} \le f_{x_i^{(k)}}\left(f_{x_{2i-1}^{(k+1)}}\left(r_{4i-3}^{(k+2)}, r_{4i-2}^{(k+2)}\right), r_{2i}^{(k+1)}\right)$$
$$\begin{aligned}r_i^{(k)} \le f_{x_i^{(k)}}\Big(& f_{x_{2i-1}^{(k+1)}}\left(r_{4i-3}^{(k+2)}, r_{4i-2}^{(k+2)}\right), \\ & f_{x_{2i}^{(k+1)}}\left(r_{4i-1}^{(k+2)}, r_{4i}^{(k+2)}\right)\Big).\end{aligned}$$

By repeating this substitution process, we may obtain an upper bound on $r_i^{(k)}$ in terms of the noise-quantization rates in $\boldsymbol{r}_{\mathcal{A},\mathcal{A}^c}\left(x_i^{(k)}\right)$. We implicitly define

$$f_{x_i^{(k)}}^{\mathcal{A},\mathcal{A}^c}\left(\boldsymbol{r}_{\mathcal{A},\mathcal{A}^c}\left(x_i^{(k)}\right)\right) \quad (32)$$

to be this upper bound. (By convention, if

$$\boldsymbol{r}_{\mathcal{A},\mathcal{A}^c}\left(x_i^{(k)}\right) = \{r_i^{(k)}\},$$

then we define this upper bound to be $r_i^{(k)}$ itself.) We then let

$$f_{x_i^{(k)}}^{\mathcal{A}}\left(\boldsymbol{r}_{\mathcal{A}}\left(x_i^{(k)}\right)\right)$$

denote the function of $\boldsymbol{r}_{\mathcal{A}}\left(x_i^{(k)}\right)$ obtained by evaluating the function in (32) with all of the noise quantization rates in

$$\boldsymbol{r}_{\mathcal{A},\mathcal{A}^c}\left(x_i^{(k)}\right) \setminus \boldsymbol{r}_{\mathcal{A}}\left(x_i^{(k)}\right)$$

set equal to zero. The significance of this function will be apparent in the proof of the outer bound.

- For any set

$$\mathcal{A} \subseteq \{1, 2, \dots, 2^{L-1}\} \quad (33)$$

we define the *ancestors set at level* k to be

$$\mathcal{A}^{(k)} \overset{\text{def}}{=} \left\{ i : \mathcal{O}\left(x_i^{(k)}\right) \cap \mathcal{A} \neq \Phi \right\}, \quad (34)$$

where Φ denotes the empty set.

Consider the following region, $\mathcal{RD}_{\text{out}}$, defined as

$$\begin{aligned}\mathcal{RD}_{\text{out}} = \Bigg\{ & (R_1, \dots, R_{2^{L-1}}, d) : \exists \left\{ r_i^{(k)} \right\} \in \mathcal{F}_r(d) : \\ & \forall \mathcal{A} \subseteq \{1, \dots, 2^{L-1}\}, \\ & \sum_{i \in \mathcal{A}} R_i \ge \sum_{k=1}^{L} \sum_{i \in \mathcal{A}^{(k)}} \left(r_i^{(k)} - f_{x_i^{(k)}}^{\mathcal{A}^c}\left(\boldsymbol{r}_{\mathcal{A}^c}\left(x_i^{(k)}\right)\right)\right) \Bigg\}. \quad (35)\end{aligned}$$

This constitutes an outer bound to the rate–distortion region of the binary tree-structure problem.

Lemma 2: For the binary tree-structure problem in which all of the noise variances are positive

$$\mathcal{RD}^* \subset \mathcal{RD}_{\text{out}}. \tag{36}$$

Proof: See Appendix A.

We next show that the outer bound just derived matches the inner bound derived from the analog–digital separation architecture (cf. Lemma 1). Recall that we use $\text{Co}(\cdot)$ to denote the closure of the convex hull of a given set.

Lemma 3: For the binary tree-structure problem in which all of the noise variances are positive

$$\mathcal{RD}_{\text{out}} = \text{co}(\mathcal{RD}_{\text{in}}). \tag{37}$$

Proof: See Appendix B.

E. Main Result

Using a continuity argument, one can relax the assumption that all of the noise variables have positive variance. This allows us to conclude our first main result of this paper: the optimality of the analog–digital separation architecture in achieving the rate–distortion region of the binary tree-structure problem.

Theorem 1: For the binary tree-structure problem, the optimal rate–distortion region is achieved by the analog–digital separation architecture

$$\mathcal{RD}^* = \text{co}(\mathcal{RD}_{\text{in}}).$$

Proof: See Appendix C. □

F. Worst Case Property

Up to this point we have assumed that the source variables are jointly Gaussian. In this subsection, we justify this assumption by noting that the rate–distortion regions for other distributions with the same covariance are only larger.

Let $(x_i^{(k)})$ be a Gaussian source satisfying the tree structure as before. Let

$$\left(\tilde{x}_1^{(1)}, \tilde{x}_1^{(L)}, \ldots, \tilde{x}_{2^{L-1}}^{(L)}\right)$$

be an alternate source with the same covariance of

$$\left(x_1^{(1)}, x_1^{(L)}, \ldots, x_{2^{L-1}}^{(L)}\right).$$

Note that the alternate source need not be part of a Markov tree. Let $\widetilde{\mathcal{RD}}^*$ denote the rate–distortion region of the alternate source.

The separation-based architecture yields an inner bound on the rate–distortion region of the alternate source. Specifically, let $\widetilde{\mathcal{RD}}_{in}$ denote the region obtained by replacing $(x_1^{(1)}, x_1^{(L)}, \ldots, x_{2^{L-1}}^{(L)})$ with $(\tilde{x}_1^{(1)}, \tilde{x}_1^{(L)}, \ldots, \tilde{x}_{2^{L-1}}^{(L)})$ in the discussion in Section II-C. Then

$$\text{co}\left(\widetilde{\mathcal{RD}}_{\text{in}}\right) \subset \widetilde{\mathcal{RD}}^*.$$

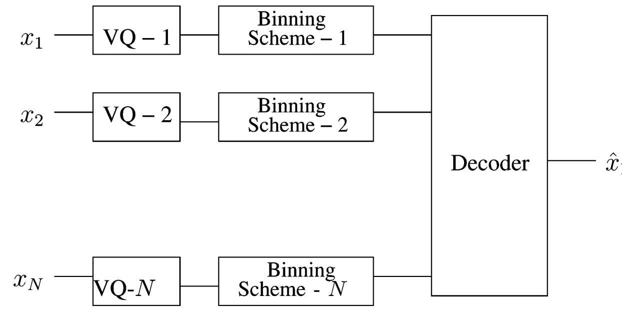

Fig. 5. The natural analog–digital separation architecture.

Theorem 2: A Gaussian source satisfying the binary tree structure has the smallest rate–distortion region for its covariance

$$\mathcal{RD}^* \subset \widetilde{\mathcal{RD}}^*.$$

In fact, the separation-based architecture has the most difficulty compressing a Gaussian source in the sense that

$$\mathcal{RD}^* = \text{co}(\mathcal{RD}_{\text{in}}) \subset \text{co}(\widetilde{\mathcal{RD}}_{\text{in}}) \subset \widetilde{\mathcal{RD}}^*. \tag{38}$$

The proof is similar to Proposition 2 in [14] and is omitted.

III. Tree Structure and the Many-Help-One Problem

We now turn to the main problem of interest: the many-help-one distributed source coding problem. As in the tree-structure problem, there is a natural analog–digital separation architecture that is a candidate solution. This is illustrated in Fig. 5.

A. Main Result

Our main result is a sufficient condition under which the analog–digital separation architecture is optimal. To state it, we first define a *general* Gauss–Markov tree: it is made up of jointly Gaussian random variables and respects the Markov conditions implied by the tree structure. The only extra feature compared to the *binary* Gauss–Markov tree (cf. Fig. 2) is that each node can have any number of descendants (not just two).

Theorem 3: Consider the many-help-one distributed source coding problem illustrated in Fig. 1. Suppose the observations $x_1, \ldots, x_N$ can be embedded in a general Gauss–Markov tree of size $M \geq N$. Then the natural analog–digital separation architecture (cf. Fig. 5) achieves the entire rate–distortion region.

Proof: The proof is elementary and builds heavily on Theorem 1. We outline the steps as follows.

- A general Gauss–Markov tree can be recast as a (potentially larger) binary Gauss–Markov tree with the root being identified with any specified node in the original tree. To see this, we only need to observe that the Markov chain relations are the same no matter which node is identified as the root.
- Next, by potentially increasing the height of the binary tree (to $\tilde{L} \geq L$) we can ensure that the observations $x_1, \ldots, x_N$ are a subset of the $2^{\tilde{L}-1}$ leaves of the binary Gauss–Markov tree. If one observation of interest, say x_i, is an intermediate node of the binary Gauss–Markov tree we can effec-

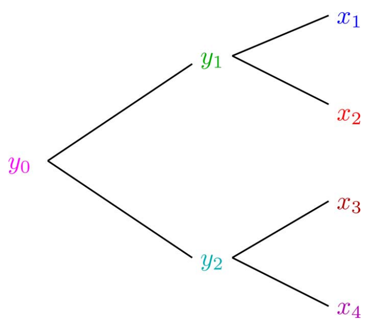

Fig. 6. Four observations are embedded in a (binary) Gauss–Markov tree.

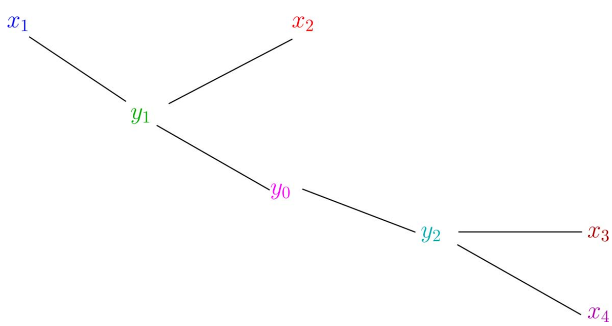

Fig. 7. The tree rewritten with x_1 as the root.

tively make it a leaf by adding descendants that are identical (almost surely) to x_i.

This allows us to convert the many-help-one problem into a binary tree-structure problem (with potentially more observations than we started out with). The analog–digital separation architecture is optimal for this problem (cf. Theorem 1). By restricting the corresponding rate–distortion region to the instance when the rates of the encoders corresponding to the observations that were not part of the original N are zero, we still have the optimality of the analog–digital separation architecture. This latter rate–distortion region simply corresponds to the many-help-one problem studied in Fig. 1. This completes the proof. □

We illustrate the two key steps outlined above with an example with $N = 4$. Suppose that $x_1, \ldots, x_4$ can be embedded in the tree depicted in Fig. 6. This tree happens to be binary, but unfortunately the root is not the source of interest, x_1. Fig. 7 shows how to construct a new Gauss–Markov tree that still preserves the Markov conditions but has x_1 as its root. Finally, a binary Gauss–Markov tree of height 5 is constructed that has the original four observations as a subset of its 16 leaf nodes; this is done in Fig. 8—here any node indicated by a dot is simply identically equal (almost surely) to its parent node. Finally, we can set to zero the rates of all the encoders except those numbered 1, 9, 13, and 14. This allows us to capture the rate–distortion region of the original three-help-one problem.

B. Worst Case Property

As with our earlier result for the binary tree-structure problem, the Gaussian assumption in Theorem 3 can be justified on the grounds that it is the worst case distribution. Specifically, as in Section II-F, let $\tilde{x}_1, \ldots, \tilde{x}_N$ denote an alternate source with the same covariances as $x_1, \ldots, x_N$. Let $\widetilde{\mathcal{RD}}^*$ denote the rate–distortion region of the source, and let $\widetilde{\mathcal{RD}}_{\text{in}}$ denote the inner bound obtained by replacing the source

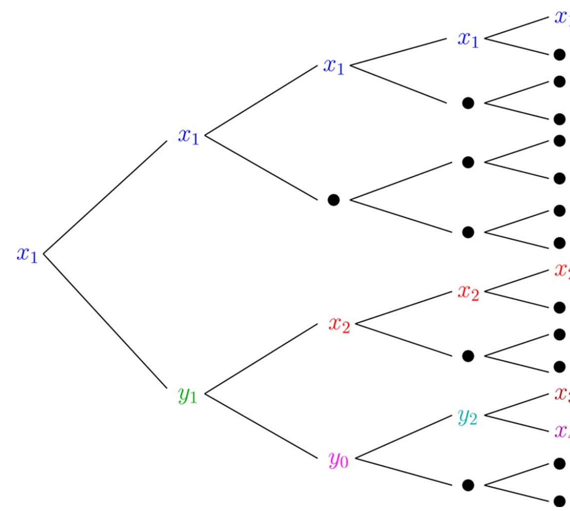

Fig. 8. The many-help-one problem rewritten as a binary tree-structure problem.

variables in the discussion in Section II-C with the alternate source $\tilde{x}_1, \ldots, \tilde{x}_N$.

Theorem 4: A Gaussian source that can be embedded in a Gauss–Markov tree has the smallest rate–distortion region for its covariance:

$$\mathcal{RD}^* \subset \widetilde{\mathcal{RD}}^*.$$

In fact, the separation-based architecture has the most difficulty compressing a Gaussian source in the sense that

$$\mathcal{RD}^* = \text{co}(\mathcal{RD}_{\text{in}}) \subset \text{co}(\widetilde{\mathcal{RD}}_{\text{in}}) \subset \widetilde{\mathcal{RD}}^*.$$

The Proof of Theorem 2 applies verbatim here.

C. Tree Structure Condition and Computational Verification

If $N = 2$, then x_1 and x_2 can always be placed in the trivial Gauss–Markov tree consisting of these two variables; no embedding is needed in this case. We note that $N = 2$ corresponds to the "one-help-one" problem, whose rate–distortion region has been determined by Oohama [6]. With $N \geq 3$, embedding is not always possible. We see an example of this next, where we also see a simple test for when N linearly independent variables can themselves be arranged in a tree, without adding additional variables. We then derive a condition on the covariance matrix of $x_1, \ldots, x_N$ that is necessary for these variables to be embedded as the nodes of a general Gauss–Markov tree. Finally, we show that this condition is also sufficient when $N = 3$.

1) Trees Without Embedding: We next demonstrate a simple test for when N linearly independent, jointly Gaussian random variables can themselves be arranged in a tree, without adding additional variables. Without loss of generality, we may assume that $x_1, \ldots, x_N$ each has unit variance (this can be ensured by normalizing each observation). We shall write

$$\rho_{ij} = \mathbb{E}[x_i x_j].$$

Suppose that $x_1, \ldots, x_N$ are linearly independent, and let $\boldsymbol{K}_x$ denote their (invertible) covariance matrix. We will use the following fact from the literature (Speed and Kiiveri [11]):

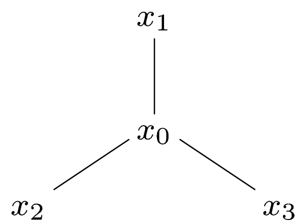

Fig. 9. Tree embedding for x_1, x_2, and x_3.

$x_1, \ldots, x_N$ are Markov with respect to a simple, undirected graph G if and only if for all $i \neq j$ such that (i,j) is not an edge in G, the (i,j) entry of $\boldsymbol{K}_x^{-1}$ is zero.

Now let G denote the simple, undirected graph with $x_1, \ldots, x_N$ as the nodes obtained by interpreting $\boldsymbol{K}_x^{-1} - \boldsymbol{I}$ as the adjacency matrix: there is an edge between x_i and x_j if and only if the (i,j) element of $\boldsymbol{K}_x^{-1} - \boldsymbol{I}$ is nonzero. It follows that $x_1, \ldots, x_N$ can be arranged in a Gauss–Markov tree if and only if G is a tree, or more generally, a forest (i.e., a collection of unconnected trees).

This fact can be illustrated with the following example. Suppose that $N = 3$ and

$$\boldsymbol{K}_x = \begin{bmatrix} 1 & 1/4 & 1/4 \\ 1/4 & 1 & 1/4 \\ 1/4 & 1/4 & 1 \end{bmatrix}. \tag{39}$$

Then

$$\boldsymbol{K}_x^{-1} = \frac{1}{9}\begin{bmatrix} 10 & -2 & -2 \\ -2 & 10 & -2 \\ -2 & -2 & 10 \end{bmatrix}$$

which yields a fully connected graph. Hence x_1, x_2, and x_3 cannot be arranged in a Gauss–Markov tree.

Nevertheless, it is possible that x_1, x_2, x_3 can be *embedded* in a larger Gauss–Markov tree. Indeed, in this case it turns out that it is possible to embed the variables in a tree of size 4. We offer the following specific construction to demonstrate this fact. Let x_0 be a standard Normal random variable and let

$$x_1 = \frac{1}{2} \cdot x_0 + z_1$$
$$x_2 = \frac{1}{2} \cdot x_0 + z_2$$
$$x_3 = \frac{1}{2} \cdot x_0 + z_3$$

where z_1, z_2, and z_3 are i.i.d. Gaussian with variance $3/4$, and are independent of x_0. The covariance matrix for this quadruple of variables is

$$\boldsymbol{K}_x = \begin{bmatrix} 1 & 1/2 & 1/2 & 1/2 \\ 1/2 & 1 & 1/4 & 1/4 \\ 1/2 & 1/4 & 1 & 1/4 \\ 1/2 & 1/4 & 1/4 & 1 \end{bmatrix}.$$

The inverse of this matrix is

$$\boldsymbol{K}_x^{-1} = \frac{2}{3}\begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 2 & 0 & 0 \\ -1 & 0 & 2 & 0 \\ -1 & 0 & 0 & 2 \end{bmatrix}$$

with the resulting G being the tree depicted in Fig. 9.

2) Necessary Condition for Tree Embedding: Even allowing additional variables in the Gauss–Markov tree, it can turn out that embedding is impossible. Towards understanding the situation better, we derive a necessary condition for $x_1, \ldots, x_N$ to be embeddable. It turns out that this condition is also sufficient when $N = 3$.

Proposition 2: Let $N \geq 3$. If $x_1, \ldots, x_N$ can be embedded in a Gauss–Markov tree, then

$$|\rho_{ik}| \geq |\rho_{ij}\rho_{jk}| \tag{40}$$

and

$$\rho_{ik}\rho_{ij}\rho_{jk} \geq 0 \tag{41}$$

for all distinct i, j, and k. Conversely, if $N = 3$ and conditions (40) and (41) hold for all distinct i j, and k, then $x_1, \ldots, x_N$ can be embedded in a Gauss–Markov tree.

Proof: See Appendix D. □

IV. A Partial Converse

We have shown that if the source can be embedded in a Gauss–Markov tree, then the separation-based scheme achieves the entire rate–distortion region for the many-help-one problem. This raises the question of whether the tree-embeddability condition can be relaxed, or whether it is necessary in order for the separation-based scheme to achieve the entire rate–distortion region. We next show that it is reasonable to conjecture that tree-embeddability, or a similar condition, is a necessary and sufficient condition for separation to achieve the entire rate–distortion region. Our argument consists of two parts.

- First, we provide an example that shows that separation does not always achieve the entire rate–distortion region for the many-help-one problem, which establishes that some added condition is required.
- We then establish a connection between this counterexample and the tree embeddability condition.

A. Suboptimality of Separation

Recall that the separation-based bound achieves exactly the Berger–Tung inner bound with Gaussian auxiliary random variables. We next show that this inner bound can be strictly smaller than the true rate–distortion region for the many-help-one problem. Consider the special case of three sources ($N = 3$), where x_1 and x_2 have covariance matrix

$$\begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}, \quad 0 < \rho < 1$$

and where $x_3 = x_1 - x_2$. We shall assume that the goal is to reproduce x_3 at the decoder and that $R_3 = 0$, i.e., the helpers completely shoulder the communication burden.

We shall focus in particular on the asymptotic regime in which σ^2 is large and ρ is near one. Specifically, let

$$\rho = 1 - \frac{1}{2\sigma^2}$$

and consider the behavior of the rate–distortion region as σ^2 tends to infinity. Note that the variance of x_3 does not tend to infinity, and in fact equals one for any positive value of σ^2, due to our choice of ρ. In this regime, the separation-based scheme performs quite poorly.

Proposition 3: Let $0 < d < 1$ and let $R(\sigma^2, d)$ denote the minimum value of $R_1 + R_2$ such that $(R_1, R_2, 0, d)$ is in the rate–distortion region for the separation-based scheme. Then

$$\lim_{\sigma^2 \to \infty} R(\sigma^2, d) = \infty.$$

Proof: Please see Appendix E. □

We now exhibit a scheme whose sum rate is bounded as σ^2 tends to infinity. This scheme is simple in the sense that it operates on individual samples, not long blocks. Consider two lattices in $\mathbb{R}$

$$\Lambda_i = \{k \cdot 2^{-n} : k \in \mathbb{Z}\}$$
$$\Lambda_o = \{k \cdot 2^{m} : k \in \mathbb{Z}\}.$$

Let $Q_i(x)$ denote the lattice point in Λ_i that is closest to x; ties are broken arbitrarily. Let

$$x \bmod \Lambda_i = x - Q_i(x).$$

Analogous definitions for Λ_o are also in effect.

Let

$$\tilde{x}_1(\ell) = Q_i(x_1(\ell)).$$

For each time ℓ, the first encoder communicates

$$u_1(\ell) = \tilde{x}_1(\ell) \bmod \Lambda_o$$

to the decoder. This requires sending $n + m$ bits per sample. The second encoder operates analogously, yielding a sum rate of $2(n + m)$ bits per sample.

The decoder uses

$$\hat{x}_3(\ell) = [u_1(\ell) - u_2(\ell)] \bmod \Lambda_o$$

as its estimate for $x_3(\ell)$.

Proposition 4: For any $d > 0$, if m and n are sufficiently large, then

$$\mathbb{E}[(x_3(\ell) - \hat{x}_3(\ell))^2] \le d$$

all ℓ and all σ^2.

Proof: Please see Appendix F. □

Since n and m need not tend to infinity as σ^2 grows, this simple scheme beats the separation-based approach by an arbitrarily large amount as σ^2 tends to infinity. The scheme can be improved by using higher dimensional lattices for Λ_i and Λ_o. This has been explored by Krithivasan and Pradhan [5].

Conceptually, the difference between the two schemes can be understood as follows. Consider the binary expansion of x_1. The quantity

$$Q_i(x_1) \bmod \Lambda_o$$

can be computed from the sign of x_1 and the m bits to the left of the binary point and the $n+1$ bits to the right of the binary point. Thus, Proposition 4 shows that only these $n+m+2$ bits are necessary for the purpose of reproducing the difference $x_1 - x_2$. In particular, it is not necessary to send the bits that are more significant than the block of m to the left of the binary point. As a result of using a standard vector quantizer, however, the separation-based scheme effectively sends these most significant bits. If the variances of x_1 and x_2 are large, this is inefficient.

B. On the Necessity of the Tree Condition

The previous section shows that the separation-based architecture does not achieve the complete rate–distortion region when x_1 and x_2 are positively correlated and $x_3 = x_1 - x_2$, at least when the variances of x_1 and x_2 are large and their correlation coefficient is near one. This is also true of the problem in which x_1 and x_2 are negatively correlated and $x_3 = x_1 + x_2$. The defining feature of these two examples is that if $\mathbb{E}[x_3 \,|\, x_1, x_2] = a_1 x_1 + a_2 x_2$, then

$$a_1 \cdot a_2 \cdot \mathbb{E}[x_1 x_2] < 0. \tag{42}$$

We next show that for $N = 3$, if the sources cannot be embedded in a Gauss–Markov tree, then this condition holds, except for a possible relabeling.

Proposition 5: For $N = 3$, if x_1, x_2, and x_3 cannot be embedded in a Gauss–Markov tree, then (42) holds for some relabeling of x_1, x_2, and x_3.

Proof: Please see Appendix G. □

Appendix A
Proof of Lemma 2

Consider any encoding–decoding procedure that achieves the rate–distortion tuple

$$(R_1, R_2, \ldots, R_{2^{L-1}}, d)$$

for the binary tree-structure problem over a block of time of length n. Let the discrete set C_i denote the output of encoder i (for $i = 1, \ldots, 2^{L-1}$). We have that

$$R_i \ge \frac{1}{n} \log |C_i|, \quad i = 1, \ldots, 2^{L-1} \tag{43}$$

$$d \ge \frac{1}{n} \sum_{m=1}^{n} \mathrm{Var}\left(x_1^{(1)}(m) \middle| \boldsymbol{C} \right). \tag{44}$$

Here we have denoted

$$\boldsymbol{C} \stackrel{\text{def}}{=} \{C_1, \ldots, C_{2^{L-1}}\}, \tag{45}$$

the set of all the encoder outputs. Further, the distributed nature of encoding imposes natural Markov chain conditions on the encoder outputs with respect to the observations. These Markov chain conditions are described in Fig. 10.

Recall our earlier definition of the *ancestors* set $\mathcal{A}^{(k)}$ (cf. (34))

$$\mathcal{A}^{(k)} \stackrel{\text{def}}{=} \left\{ i : \mathcal{O}(x_i^{(k)}) \cap \mathcal{A} \ne \Phi \right\}. \tag{46}$$

Now define

$$\boldsymbol{x}_{\mathcal{A},n}^{(k)} \stackrel{\text{def}}{=} \left\{ x_{i,n}^{(k)} : i \in \mathcal{A}^{(k)} \right\}. \tag{47}$$

Our outer bound will consider arbitrary subsets $\mathcal{A}$ of $\{1, \ldots, 2^{L-1}\}$. Denote the set

$$\boldsymbol{C}_{\mathcal{A}} \stackrel{\text{def}}{=} \{C_i : i \in \mathcal{A}\}. \tag{48}$$

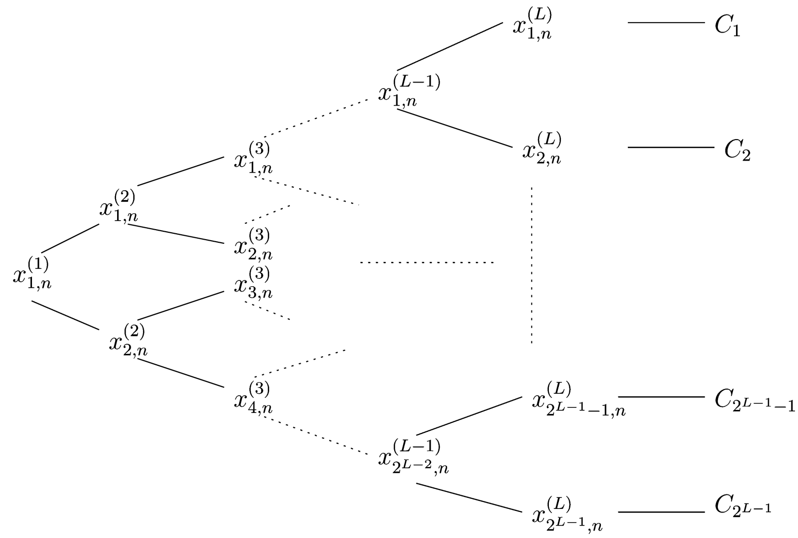

Fig. 10. The tree structure with the encoder outputs over a block of length n.

The sum of any subset $\mathcal{A}$ of the encoder rates satisfies

$$
\begin{aligned}
n\sum_{i\in\mathcal{A}} R_i &= \sum_{i\in\mathcal{A}} \log|C_i| \\
&\geq \sum_{i\in\mathcal{A}} H(C_i) \\
&\geq H\left(\boldsymbol{C}_{\mathcal{A}}\right) \\
&\geq H\left(\boldsymbol{C}_{\mathcal{A}} \,|\, \boldsymbol{C}_{\mathcal{A}^c}\right) \\
&= I\left(\boldsymbol{x}_{\mathcal{A},n}^{(L)}; \boldsymbol{C}_{\mathcal{A}} \,|\, \boldsymbol{C}_{\mathcal{A}^c}\right) \\
&\stackrel{(a)}{=} I\left(\boldsymbol{x}_{\mathcal{A},n}^{(1)}, \ldots \boldsymbol{x}_{\mathcal{A},n}^{(L-1)}, \boldsymbol{x}_{\mathcal{A},n}^{(L)}; \boldsymbol{C}_{\mathcal{A}} \,|\, \boldsymbol{C}_{\mathcal{A}^c}\right) &(49)\\
&\stackrel{(b)}{=} \sum_{k=1}^{L} I\left(\boldsymbol{x}_{\mathcal{A},n}^{(k)}; \boldsymbol{C}_{\mathcal{A}} \middle| \boldsymbol{x}_{\mathcal{A},n}^{(k-1)}, \boldsymbol{C}_{\mathcal{A}^c}\right) &(50)\\
&\stackrel{(c)}{=} \sum_{k=1}^{L} \left(I\left(\boldsymbol{x}_{\mathcal{A},n}^{(k)}; \boldsymbol{C} \middle| \boldsymbol{x}_{\mathcal{A},n}^{(k-1)}\right) \right.\\
&\qquad \left. - I\left(\boldsymbol{x}_{\mathcal{A},n}^{(k)}; \boldsymbol{C}_{\mathcal{A}^c} \middle| \boldsymbol{x}_{\mathcal{A},n}^{(k-1)}\right)\right). &(51)
\end{aligned}
$$

Here each of the steps (a), (b), and (c) follow from the Markov chain conditions described in Fig. 10. We use the chain rule to expand each of the mutual information terms in the lower bound of (51)

$$
\begin{aligned}
&I\left(\boldsymbol{x}_{\mathcal{A},n}^{(k)}; \boldsymbol{C} \,\middle|\, \boldsymbol{x}_{\mathcal{A},n}^{(k-1)}\right)\\
&\quad= \sum_{i\in\mathcal{A}^{(k)}} I\left(x_{i,n}^{(k)}; \boldsymbol{C} \,\middle|\, \boldsymbol{x}_{\mathcal{A},n}^{(k-1)}, x_{j,n}^{(k)}, j<i, j\in\mathcal{A}^{(k)}\right) &(52)\\
&\quad= \sum_{i\in\mathcal{A}^{(k)}} I\left(x_{i,n}^{(k)}; \boldsymbol{C} \,\middle|\, x_{\lfloor\frac{i+1}{2}\rfloor,n}^{(k-1)}\right) &(53)
\end{aligned}
$$

and

$$
\begin{aligned}
&I\left(\boldsymbol{x}_{\mathcal{A},n}^{(k)}; \boldsymbol{C}_{\mathcal{A}^c} \,\middle|\, \boldsymbol{x}_{\mathcal{A},n}^{(k-1)}\right)\\
&\quad= \sum_{i\in\mathcal{A}^{(k)}} I\left(x_{i,n}^{(k)}; \boldsymbol{C}_{\mathcal{A}^c} \,\middle|\, \boldsymbol{x}_{\mathcal{A},n}^{(k-1)}, x_{j,n}^{(k)}, j<i, j\in\mathcal{A}^{(k)}\right)\\
& &(54)\\
&\quad= \sum_{i\in\mathcal{A}^{(k)}} I\left(x_{i,n}^{(k)}; \boldsymbol{C}_{\mathcal{A}^c} \,\middle|\, x_{\lfloor\frac{i+1}{2}\rfloor,n}^{(k-1)}\right) &(55)
\end{aligned}
$$

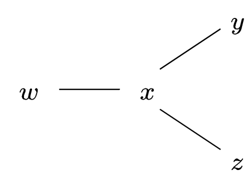

Fig. 11. The Markov chain conditions.

Here both (53) and (55) follow from the Markov chain conditions described in Fig. 10. Denote by

$$
r_i^{(k)} \stackrel{\text{def}}{=} \frac{1}{n} I\left(x_{i,n}^{(k)}; \boldsymbol{C} \,\middle|\, x_{\lfloor\frac{i+1}{2}\rfloor,n}^{(k-1)}\right) \tag{56}
$$

the term inside the summation in (53). Then $r_1^{(1)}$ is the number of bits per sample that the encoders send about the root of the tree and $r_i^{(k)}$ for $k>1$ can be interpreted as the number of bits per sample that the encoders use to represent the noise introduced at node $x_i^{(k)}$. We will upper-bound the terms inside the summation in (55) in terms of these quantities. To do this, we start with a central preliminary lemma.

A. A Preliminary Lemma

Consider four memoryless jointly Gaussian random processes $w(m), x(m), y(m), z(m), m = 1, \ldots, n$. They are identically jointly distributed in the (time) index m. At any given time index m, their joint distribution satisfies the Markov chain conditions implied in Fig. 11. Then we can write, for all $m = 1, \ldots, n$

$$
\begin{aligned}
x(m) &= \alpha_{xw} w(m) + n_0(m)\\
y(m) &= \alpha_{yx} x(m) + n_1(m)\\
z(m) &= \alpha_{zx} x(m) + n_2(m)
\end{aligned}
$$

for some real $\alpha_{xw}, \alpha_{yx}, \alpha_{zx}$. Here $n_0(m), n_1(m), n_2(m), m = 1\ldots, n$, are i.i.d. in time and independent of each other and independent of the process $w(m), m = 1, \ldots, n$. Further, the random variables $n_0(m), n_1(m), n_2(m), w(m)$ at any time index n are $\mathcal{N}(0,\sigma_{n_0}^2), \mathcal{N}(0,\sigma_{n_1}^2), \mathcal{N}(0,\sigma_{n_2}^2)$ and $\mathcal{N}(0,\sigma_w^2)$, respectively.

Write the vectors

$$w_n = [w(1), \ldots, w(n)] \tag{57}$$
$$x_n = [x(1), \ldots, x(n)] \tag{58}$$
$$y_n = [y(1), \ldots, y(n)] \tag{59}$$
$$z_n = [z(1), \ldots, z(n)]. \tag{60}$$

Consider two random variables C_1, C_2 that satisfy the following two Markov chain conditions:

$$(w_n, x_n, z_n, C_2) \leftrightarrow y_n \leftrightarrow C_1 \tag{61}$$
$$(w_n, x_n, y_n, C_1) \leftrightarrow z_n \leftrightarrow C_2. \tag{62}$$

Our first inequality concerns this Markov chain condition. We intentionally use notation similar to that introduced in Section II-D.

Lemma 4: Define

$$r_1 \stackrel{\text{def}}{=} \frac{1}{n} I(y_n; C_1 \,|\, x_n)$$
$$r_2 \stackrel{\text{def}}{=} \frac{1}{n} I(z_n; C_2 \,|\, x_n)$$
$$f_x(r_1, r_2) \stackrel{\text{def}}{=} \frac{1}{2} \log \left(1 + \frac{\alpha_{yx}^2 \sigma_{n_0}^2}{\sigma_{n_1}^2} (1 - e^{-2r_1}) + \frac{\alpha_{zx}^2 \sigma_{n_0}^2}{\sigma_{n_2}^2} (1 - e^{-2r_2}) \right).$$

Then

$$\frac{1}{n} I(x_n; C_1, C_2 \,|\, w_n) \le f_x(r_1, r_2) \tag{63}$$
$$\frac{1}{n} I(x_n; C_1 \,|\, w_n) \le f_x(r_1, 0) \tag{64}$$
$$\frac{1}{n} I(x_n; C_2 \,|\, w_n) \le f_x(0, r_2). \tag{65}$$

Proof: This lemma is a conditional version (conditioned on w_n) of Lemma 3 in [8]. The proof follows "mutatis mutandis" that of Lemma 3 in [8]; the only extra fact needed is that conditioned on *any* realization of w_n, (x_n, y_n, z_n) are jointly Gaussian with their original variances and $(x_n, y_n, z_n, C_1, C_2)$ satisfies the Markov condition

$$C_1 \leftrightarrow y_n \leftrightarrow x_n \leftrightarrow z_n \leftrightarrow C_2.$$

Specifically, suppose first that α_{yx} and α_{zx} are nonzero. For any realization of w_n, say $\tilde{w}_n$, Oohama [8, Lemma 3] has shown that

$$\frac{1}{n} I(x_n; C_1, C_2 \,|\, w_n = \tilde{w}_n) \le f_x(r_1, r_2).$$

By averaging the left-hand side over $\tilde{w}_n$, we obtain (63). The proofs of (64) and (65) are similar. If both α_{yx} and α_{zx} are zero, then the result is trivial. If, say, only α_{yx} is zero, then $I(x_n; C_1, C_2 \,|\, w_n) = I(x_n; C_2 \,|\, w_n)$ and (63) follows from (65). □

1) Sufficient Conditions for Equality: It is useful to observe the conditions for equality in (63), (64), and (65): suppose

$$C_k = [u_k(1), \ldots, u_k(n)], \quad k = 1, 2. \tag{66}$$

Here

$$u_1(m) = \alpha_1 y(m) + v_1(m), \quad m = 1, \ldots, n$$
$$u_2(m) = \alpha_2 z(m) + v_2(m), \quad m = 1, \ldots, n$$

where $v_1(m)$ and $v_2(m)$ are Gaussian and independent of each other and of w_n, x_n, y_n, z_n and are i.i.d. in the time index m. Then it is verified directly that with this choice of C_1, C_2 (cf. (66)) the inequalities in (63), (64), and (65) are all simultaneously met with equality (this verification is also done in [8], [10]). This fact will be used later to show that the achievable region of the separation-based inner bound coincides with the outer bound.

2) An Important Instance: Of specific interest to us will be the following association of the random variables in Fig. 11 to the binary tree structure in Fig. 2: fix $1 \le k \le L-1$ and $1 \le i \le 2^{k-1}$. Then let

$$x = x_i^{(k)} \tag{67}$$
$$y = x_{2i-1}^{(k+1)} \tag{68}$$
$$z = x_{2i}^{(k+1)} \tag{69}$$
$$w = x_{\lfloor \frac{i+1}{2} \rfloor}^{(k-1)}. \tag{70}$$

With this association, denote the function corresponding to f_x in (63) by $f_{x_i^{(k)}}$

$$f_{x_i^{(k)}}(r_1, r_2) \stackrel{\text{def}}{=} \frac{1}{2} \log \left(1 + \frac{\left\{\alpha_{2i-1}^{(k+1)}\right\}^2 \sigma_{n_i^{(k)}}^2}{\sigma_{n_{2i-1}^{(k+1)}}^2} (1 - e^{-2r_1}) + \frac{\left\{\alpha_{2i}^{(k+1)}\right\}^2 \sigma_{n_i^{(k)}}^2}{\sigma_{n_{2i}^{(k+1)}}^2} \left(1 - e^{-2r_2}\right) \right), \quad r_1, r_2 \ge 0. \tag{71}$$

Indeed, this is the same notation as that introduced in Section II-D (cf. (22)).

B. An Iteration Lemma

As an immediate application of the preliminary lemma derived in the previous subsection, consider any subset $\mathcal{A} \subseteq \{1, \ldots, 2^{L-1}\}$. Fix $1 \le k \le L-1$ and $1 \le i \le 2^{k-1}$.

Lemma 5:

$$\frac{1}{n} I\left(x_{i,n}^{(k)}; \boldsymbol{C}_{\mathcal{A}} \,\middle|\, x_{\lfloor \frac{i+1}{2} \rfloor, n}^{(k-1)} \right) \le f_{x_i^{(k)}} \left(\frac{1}{n} I\left(x_{2i-1,n}^{(k+1)}; \boldsymbol{C}_{\mathcal{A}} \,\middle|\, x_{i,n}^{(k)} \right), \frac{1}{n} I\left(x_{2i,n}^{(k+1)}; \boldsymbol{C}_{\mathcal{A}} \,\middle|\, x_{i,n}^{(k)} \right) \right). \tag{72}$$

Proof: For any node $x_i^{(k)}$, recall the set of associated *observations* defined as (cf. (24))

$$\mathcal{O}\left(x_i^{(k)}\right) = \left\{ j : \frac{2^L (i-1)}{2^k} < j \le \frac{2^L i}{2^k} \right\}.$$

With this definition, we observe that

$$\mathcal{O}\left(x_i^{(k)}\right)=\mathcal{O}\left(x_{2i-1}^{(k+1)}\right)\cup\mathcal{O}\left(x_{2i}^{(k+1)}\right)$$

$$I\left(x_{i,n}^{(k)};\boldsymbol{C}_{\mathcal{A}}\,\middle|\,x_{\lfloor\frac{i+1}{2}\rfloor,n}^{(k-1)}\right)=I\left(x_{i,n}^{(k)};\boldsymbol{C}_{\mathcal{A}\cap\mathcal{O}\left(x_i^{(k)}\right)}\,|\,x_{\lfloor\frac{i+1}{2}\rfloor,n}^{(k-1)}\right).$$

Then we only need to invoke Lemma 4 with the following random variables:

$$\begin{aligned} w_n &= x_{\lfloor\frac{i+1}{2}\rfloor,n}^{(k-1)}\\ x_n &= x_{i,n}^{(k)}\\ y_n &= x_{2i-1,n}^{(k+1)}\\ z_n &= x_{2i,n}^{(k+1)}\\ C_1 &= \boldsymbol{C}_{\mathcal{A}\cap\mathcal{O}\left(x_{2i-1}^{(k+1)}\right)}\\ C_2 &= \boldsymbol{C}_{\mathcal{A}\cap\mathcal{O}\left(x_{2i}^{(k+1)}\right)}. \end{aligned}$$

This completes the proof. □

Observe that the parameters inside the function $f_{x_i^{(k)}}(\cdot,\cdot)$ are themselves of the type of the term in the left-hand side of (72). Then, we can repeatedly apply Lemma 5. As an example, we have for $k\le L-2$ and $1\le i\le 2^{k-1}$, the two parameters of $f_{x_i^{(k)}}$ in (72) are upper-bounded by

$$\begin{aligned} &\frac{1}{n}I\left(x_{2i-1,n}^{(k+1)};\boldsymbol{C}_{\mathcal{A}}\,\middle|\,x_{i,n}^{(k)}\right)\\ &\quad\le f_{x_{2i-1}^{(k+1)}}\left(\frac{1}{n}I\left(x_{4i-3,n}^{(k+2)};\boldsymbol{C}_{\mathcal{A}}\,\middle|\,x_{2i-1,n}^{(k+1)}\right),\right.\\ &\qquad\qquad\left.\frac{1}{n}I\left(x_{4i-2,n}^{(k+2)};\boldsymbol{C}_{\mathcal{A}}\,\middle|\,x_{2i-1,n}^{(k+1)}\right)\right). \qquad (73)\\ &\frac{1}{n}I\left(x_{2i,n}^{(k+1)};\boldsymbol{C}_{\mathcal{A}}\,\middle|\,x_{i,n}^{(k)}\right)\\ &\quad\le f_{x_{2i}^{(k+1)}}\left(\frac{1}{n}I\left(x_{4i-1,n}^{(k+2)};\boldsymbol{C}_{\mathcal{A}}\,\middle|\,x_{2i,n}^{(k+1)}\right),\right.\\ &\qquad\qquad\left.\frac{1}{n}I\left(x_{4i,n}^{(k+2)};\boldsymbol{C}_{\mathcal{A}}\,\middle|\,x_{2i,n}^{(k+1)}\right)\right). \qquad (74) \end{aligned}$$

Now the function $f_{x_i^{(k)}}(\cdot,\cdot)$ is monotonically increasing in both of its parameters (this is true for each $1\le k\le L-1$ and $1\le i\le 2^{k-1}$). So, we can combine (72), (74) and (73) to get

$$\begin{aligned} &\frac{1}{n}I\left(x_{i,n}^{(k)};\boldsymbol{C}_{\mathcal{A}}\,\middle|\,x_{\lfloor\frac{i+1}{2}\rfloor,n}^{(k-1)}\right) \qquad (75)\\ &\le f_{x_i^{(k)}}\left(f_{x_{2i-1}^{(k+1)}}\left(\frac{1}{n}I\left(x_{4i-3,n}^{(k+2)};\boldsymbol{C}_{\mathcal{A}}\,\middle|\,x_{2i-1,n}^{(k+1)}\right),\right.\right.\\ &\qquad\left.\frac{1}{n}I\left(x_{4i-2,n}^{(k+2)};\boldsymbol{C}_{\mathcal{A}}\,\middle|\,x_{2i-1,n}^{(k+1)}\right)\right),\\ &\qquad f_{x_{2i}^{(k+1)}}\left(\frac{1}{n}I\left(x_{4i-1,n}^{(k+2)};\boldsymbol{C}_{\mathcal{A}}\,\middle|\,x_{2i,n}^{(k+1)}\right),\right.\\ &\qquad\left.\left.\frac{1}{n}I\left(x_{4i,n}^{(k+2)};\boldsymbol{C}_{\mathcal{A}}\,\middle|\,x_{2i,n}^{(k+1)}\right)\right)\right). \qquad (76) \end{aligned}$$

The stage is now set to recursively apply Lemma 5. Continuing this process until the boundary conditions are met, we arrive at

$$\frac{1}{n}I\left(x_{i,n}^{(k)};\boldsymbol{C}_{\mathcal{A}}\,\middle|\,x_{\lfloor\frac{i+1}{2}\rfloor,n}^{(k-1)}\right)\le f_{x_i^{(k)}}^{\mathcal{A}}\left(\boldsymbol{r}_{\mathcal{A}}\left(x_i^{(k)}\right)\right). \qquad (77)$$

Here, the set $\boldsymbol{r}_{\mathcal{A}}\left(x_i^{(k)}\right)$ is defined as in (25)

$$\begin{aligned} \boldsymbol{r}_{\mathcal{A}}\left(x_i^{(k)}\right)=\Big\{r_j^{(l)}:\ &x_j^{(l)}\in\mathcal{T}\left(x_i^{(k)}\right),\mathcal{O}(x_j^{(l)})\subset\mathcal{A},\\ &\nexists x_a^{(b)}\in\mathcal{T}\left(x_i^{(k)}\right)\text{ with }\mathcal{O}\left(x_a^{(b)}\right)\subset\mathcal{A},\\ &\text{and } x_j^{(l)}\in\mathcal{R}\left(x_a^{(b)}\right)\cup\mathcal{L}\left(x_a^{(b)}\right)\Big\}. \qquad (78) \end{aligned}$$

The function $f_{x_i^{(k)}}^{\mathcal{A}}(\,\cdot\,)$ was also defined in Section II-D.

C. Putting Them Together

We are now ready to complete the Proof of Lemma 2. First, we substitute (77) in (55) to get

$$I\left(\boldsymbol{x}_{\mathcal{A},n}^{(k)};\boldsymbol{C}_{\mathcal{A}^c}\,\middle|\,\boldsymbol{x}_{\mathcal{A},n}^{(k-1)}\right)\le\sum_{i\in\mathcal{A}^{(k)}}f_{x_i^{(k)}}^{\mathcal{A}^c}\left(\boldsymbol{r}_{\mathcal{A}^c}\left(x_i^{(k)}\right)\right). \qquad (79)$$

Combining (79) with (53) and (56), we can rewrite the inequality in (51) as

$$\sum_{i\in\mathcal{A}}R_i\ge\sum_{k=1}^{L}\sum_{i\in\mathcal{A}^{(k)}}\left(r_i^{(k)}-f_{x_i^{(k)}}^{\mathcal{A}^c}\left(\boldsymbol{r}_{\mathcal{A}^c}\left(x_i^{(k)}\right)\right)\right). \qquad (80)$$

The quantities $r_i^{(k)}$ satisfy other natural inequalities as well.

- Supposing that $\mathcal{A}$ equals the entire set $\{1,2,\ldots,2^{L-1}\}$ and substituting in Lemma 5 we have

$$r_i^{(k)}\le f_{x_i}^{(k+1)}\left(r_{2i-1}^{(k+1)},r_{2i}^{(k+1)}\right). \qquad (81)$$

- By direct calculation we also have

$$\begin{aligned} r_1^{(1)}&=\frac{1}{n}I\left(x_{1,n}^{(1)};\boldsymbol{C}\right) &(82)\\ &=\frac{1}{n}h\left(x_{1,n}^{(1)}\right)-\frac{1}{n}h\left(x_{1,n}^{(1)}\,\middle|\,\boldsymbol{C}\right) &(83)\\ &\ge\frac{1}{2}\log\left(2\pi e\sigma_{x_1^{(1)}}^2\right)-\frac{1}{n}h\left(x_{1,n}^{(1)}-\mathbb{E}\left[x_{1,n}^{(1)}\,\middle|\,\boldsymbol{C}\right]\right) &(84)\\ &\ge\frac{1}{2}\log\left(\sigma_{x_1^{(1)}}^2\right)\\ &\quad-\frac{1}{2n}\log\det\left(\mathrm{Covar}\left(x_{1,n}^{(1)}-\mathbb{E}\left[x_{1,n}^{(1)}\,\middle|\,\boldsymbol{C}\right]\right)\right) &(85)\\ &\ge\frac{1}{2}\log\left(\sigma_{x_1^{(1)}}^2\right)\\ &\quad-\frac{1}{2}\log\left(\frac{1}{n}\mathrm{Trace}\left(\mathrm{Covar}\left(x_{1,n}^{(1)}-\mathbb{E}\left[x_{1,n}^{(1)}\,\middle|\,\boldsymbol{C}\right]\right)\right)\right) &(86)\\ &=\frac{1}{2}\log\left(\sigma_{x_1^{(1)}}^2\right)-\frac{1}{2}\log\left(\frac{1}{n}\sum_{m=1}^{n}\mathrm{Var}\left(x_1^{(1)}(m)\,\middle|\,\boldsymbol{C}\right)\right) &(87)\\ &\ge\frac{1}{2}\log\frac{\sigma_{x_1^{(1)}}^2}{d}. &(88) \end{aligned}$$

where

— equation (84) follows from the fact that conditioning only reduces the differential entropy;

— equation (85) is the usual bound on the differential entropy of a vector by the determinant of its covariance matrix;

— equation (86) follows from the Hadamard inequality on the determinant of a positive definite matrix in terms of its trace;

— equation (88) follows from the fact that the encoder outputs describe the original root node of the tree with sufficiently small quadratic fidelity (cf. (44)).

Based on (81) and (88) we see that the set of $r_i^{(k)}$ indeed belong to the set $\mathcal{F}_r(d)$ defined in (26). Combining this fact with the key inequality in (80), we have completed the proof of the outer bound in Lemma 2. □

Appendix B
Proof of Lemma 3

Since we know that

$$\mathrm{co}(\mathcal{RD}_{\mathrm{in}}) \subset \mathcal{RD}_{\mathrm{out}} \tag{89}$$

it suffices to prove that for any d and any componentwise nonnegative vector $(\alpha_1, \dots, \alpha_{2^{L-1}})$

$$\inf_{\boldsymbol{R}:(\boldsymbol{R},d)\in\mathcal{RD}_{\mathrm{out}}} \sum_{i=1}^{2^{L-1}} \alpha_i R_i \geq \inf_{\boldsymbol{R}:(\boldsymbol{R},d)\in\mathcal{RD}_{\mathrm{in}}} \sum_{i=1}^{2^{L-1}} \alpha_i R_i.$$

We will assume that $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_{2^{L-1}}$. The proof for the other orderings is similar. We will also use the convention $\alpha_0 = 0$. Now for any $R_1, \dots, R_{2^{L-1}}$

$$\begin{aligned}\sum_{i=1}^{2^{L-1}} \alpha_i R_i &= \alpha_1 \sum_{i=1}^{2^{L-1}} R_i + (\alpha_2 - \alpha_1) \sum_{i=2}^{2^{L-1}} R_i + \cdots \\ &\qquad + (\alpha_{2^{L-1}} - \alpha_{2^{L-1}-1}) R_{2^{L-1}} \\ &= \sum_{j=1}^{2^{L-1}} (\alpha_j - \alpha_{j-1}) \sum_{i=j}^{2^{L-1}} R_i.\end{aligned}$$

Thus

$$\inf_{\boldsymbol{R}:(\boldsymbol{R},d)\in\mathcal{RD}_{\mathrm{out}}} \sum_{i=1}^{2^{L-1}} \alpha_i R_i = \inf_{\boldsymbol{R}:(\boldsymbol{R},d)\in\mathcal{RD}_{\mathrm{out}}} \sum_{j=1}^{2^{L-1}} (\alpha_j - \alpha_{j-1}) \sum_{i=j}^{2^{L-1}} R_i.$$

Let $\epsilon > 0$. Then there exists $s \in \mathcal{F}_r(d)$ and $\boldsymbol{R}^*$ such that

$$\sum_{j=1}^{2^{L-1}} (\alpha_j - \alpha_{j-1}) \sum_{i=j}^{2^{L-1}} R_i^* \leq \inf_{\boldsymbol{R}:(\boldsymbol{R},d)\in\mathcal{RD}_{\mathrm{out}}} \sum_{j=1}^{2^{L-1}} (\alpha_j - \alpha_{j-1}) \sum_{i=j}^{2^{L-1}} R_i + \epsilon$$

and

$$\sum_{i\in\mathcal{A}} R_i^* \geq \sum_{k=1}^{L} \sum_{i\in\mathcal{A}^{(k)}} \left(s_i^{(k)} - f_{x_i^{(k)}}^{\mathcal{A}^c} \left(\boldsymbol{s}_{\mathcal{A}^c} \left(x_i^{(k)} \right) \right) \right)$$

for all $\mathcal{A}$. Let

$$\mathcal{A}_j = \{j, \dots, 2^{L-1}\} \cap \{i : s_i^{(L)} > 0\}.$$

Then

$$\begin{aligned}\sum_{j=1}^{2^{L-1}} (\alpha_j - \alpha_{j-1}) \sum_{i=j}^{2^{L-1}} R_i^* &\geq \sum_{j=1}^{2^{L-1}} (\alpha_j - \alpha_{j-1}) \sum_{i\in\mathcal{A}_j} R_i^* \\ &\geq \sum_{j=1}^{2^{L-1}} (\alpha_j - \alpha_{j-1}) \sum_{k=1}^{L} \sum_{i\in\mathcal{A}_j^{(k)}} \left(s_i^{(k)} - f_{x_i^{(k)}}^{\mathcal{A}_j^c} (\boldsymbol{s}_{\mathcal{A}^c}(x_i^{(k)})) \right) \\ &\geq \inf \sum_{j=1}^{2^{L-1}} (\alpha_j - \alpha_{j-1}) \sum_{k=1}^{L} \sum_{i\in\mathcal{A}_j^{(k)}} \left(r_i^{(k)} - f_{x_i^{(k)}}^{\mathcal{A}_j^c} (\boldsymbol{r}_{\mathcal{A}^c}(x_i^{(k)})) \right)\end{aligned}$$

where the infimum is over all $\boldsymbol{r}$ in $\mathcal{F}_r(d)$ such that $r_i^{(L)} = 0$ if $s_i^{(L)} = 0$. Then there exists $\tilde{\boldsymbol{s}} \in \mathcal{F}_r(d)$ such that

1) $\tilde{\boldsymbol{s}}_i^{(L)} = 0$ if $\boldsymbol{s}_i^{(L)} = 0$,

2)
$$\begin{aligned}&\sum_{j=1}^{2^{L-1}} (\alpha_j - \alpha_{j-1}) \sum_{k=1}^{L} \sum_{i\in\mathcal{A}_j^{(k)}} \left(\tilde{s}_i^{(k)} - f_{x_i^{(k)}}^{\mathcal{A}_j^c} \left(\tilde{\boldsymbol{s}}_{\mathcal{A}^c} \left(x_i^{(k)} \right) \right) \right) \\ &\quad \leq \inf \sum_{j=1}^{2^{L-1}} (\alpha_j - \alpha_{j-1}) \sum_{k=1}^{L} \sum_{i\in\mathcal{A}_j^{(k)}} \left(r_i^{(k)} \right. \\ &\qquad \left. - f_{x_i^{(k)}}^{\mathcal{A}_j^c} \left(\boldsymbol{r}_{\mathcal{A}_j^c} \left(x_i^{(k)} \right) \right) \right) + \epsilon, \text{ and}\end{aligned} \tag{90}$$

3) of all the $\tilde{\boldsymbol{s}}$ satisfying the first two conditions, the chosen $\tilde{\boldsymbol{s}}$ minimizes
$$\sum_{k=1}^{L} \sum_{i=1}^{2^{k-1}} \tilde{s}_i^{(k)}. \tag{91}$$

Now since the $\tilde{s}_i^{(k)}$ are in $\mathcal{F}_r(d)$, they must satisfy the inequalities that define this set in (26)

$$\tilde{s}_1^{(1)} \geq \frac{1}{2} \log \frac{\sigma_{x_1^{(1)}}^2}{d} \tag{92}$$

$$\tilde{s}_i^{(k)} \leq f_{x_i^{(k)}} \left(s_{2i-1}^{(k+1)}, s_{2i}^{(k+1)} \right). \tag{93}$$

We will show that both of these inequalities must actually be equalities. Since the left-hand side of (90) is nonincreasing in $s_1^{(1)}$ and the $s_i^{(k)}$ minimize (91), it follows that the $s_1^{(1)}$ inequality must be tight.

Next suppose that

$$\tilde{s}_m^{(n)} < f_{x_m^{(n)}} (\tilde{s}_{2m-1}^{(n+1)}, \tilde{s}_{2m}^{(n+1)}) \tag{94}$$

for some non-leaf node $x_m^{(n)}$. We will show that this is incompatible with the assumption that the $\tilde{s}_i^{(k)}$ minimize (91). Without loss of generality, we may assume that none of the children of $x_m^{(n)}$ have a strict inequality in (93). In order for (94) to hold, $\tilde{s}_j^{(L)}$ must be positive for at least one leaf variable $x_j^{(L)}$ under $x_m^{(n)}$. Consider the leaf variable $x_{\hat{m}}^{(L)}$ under $x_m^{(n)}$ with the largest index $\hat{m}$ such that $\tilde{s}_{\hat{m}}^{(L)}$ is positive

$$\hat{m} = \arg\max \left\{ \frac{2^L(m-1)}{2^n} < j \leq \frac{2^L m}{2^n} : \tilde{s}_j^{(L)} > 0 \right\}.$$

Then consider the descendant of $x_m^{(n)}$, $x_{\tilde{m}}^{(n+1)}$, that leads to the leaf variable $x_{\hat{m}}^{(L)}$. Note that we must have $\tilde{s}_{\tilde{m}}^{(n+1)} > 0$.

Suppose that we decrease $\tilde{s}_{\tilde{m}}^{(n+1)}$ by a slight amount such that (94) still holds. Fix a j in $\{1,\ldots,2^{L-1}\}$ and consider the sum

$$\sum_{k=1}^{L}\sum_{i\in\mathcal{A}_j^{(k)}}\left(\tilde{s}_i^{(k)}-f_{x_i^{(k)}}^{\mathcal{A}_j^c}\left(\tilde{\boldsymbol{s}}_{\mathcal{A}_j^c}\left(x_i^{(k)}\right)\right)\right) \quad (95)$$

and recall that

$$\mathcal{A}_j=\{j,\ldots,2^{L-1}\}\cap\left\{i:\tilde{s}_i^L>0\right\}.$$

Now if $j>\hat{m}$, then all of the observations under $x_m^{(n)}$ are in $\mathcal{A}_j^c$, which implies that the sum in (95) does not depend on $\tilde{s}_{\tilde{m}}^{(n+1)}$. On the other hand, if $j\le\hat{m}$, then not all of the observations under $\tilde{s}_{\tilde{m}}^{(n+1)}$ are in $\mathcal{A}_j^c$, and so

$$\tilde{s}_{\tilde{m}}^{(n+1)}\notin\tilde{\boldsymbol{s}}_{\mathcal{A}_j^c}\left(x_i^{(k)}\right)$$

for all $x_i^{(k)}$. It follows that the objective in (90) is not increased while the sum in (91) is reduced by decreasing $\tilde{s}_{\tilde{m}}^{(n+1)}$, which is a contradiction. Thus, (94) cannot hold at any non-leaf nodes in the tree. We have thus shown that equality must hold in (92) and (93).

We are now in a position to show that

$$\sum_{j=1}^{2^{L-1}}(\alpha_j-\alpha_{j-1})\sum_{k=1}^{L}\sum_{i\in\mathcal{A}_j^{(k)}}\left(\tilde{s}_i^{(k)}-f_{x_i^{(k)}}^{\mathcal{A}_j^c}\left(\tilde{\boldsymbol{s}}_{\mathcal{A}_j^c}\left(x_i^{(k)}\right)\right)\right)\ge\inf_{\boldsymbol{R}:(\boldsymbol{R},d)\in\mathcal{RD}_{\text{in}}(d)}\sum_{i=1}^{2^{L-1}}\alpha_iR_i. \quad (96)$$

Specifically, choose the auxiliary random variables $\boldsymbol{u}$ in the Berger–Tung inner bound such that

$$I\left(x_i^{(L)};u_i\,\middle|\,x_{\lfloor\frac{i+1}{2}\rfloor}^{(L-1)}\right)=\tilde{s}_i^{(L)}$$

for each observation i. We will first show by induction that

$$I\left(x_i^{(k)};\boldsymbol{u}\,\middle|\,x_{\lfloor(i+1)/2\rfloor}^{(k-1)}\right)=\tilde{s}_i^{(k)} \quad (97)$$

for all variables $x_i^{(k)}$ in the tree. This is true of the leaf variables $x_i^{(L)}$, $i=1,\ldots,2^{L-1}$ by hypothesis. Next consider a variable $x_i^{(k)}$ and suppose the condition holds for $x_{2i-1}^{(k+1)}$ and $x_{2i}^{(k+1)}$. By the observation in Appendix A

$$\begin{aligned}I&\left(x_i^{(k)};\boldsymbol{u}\,\middle|\,x_{\lfloor(i+1)/2\rfloor}^{(k-1)}\right)\\&=f_{x_i^{(k)}}\left(I\left(x_{2i-1}^{(k+1)};\boldsymbol{u}\,\middle|\,x_i^{(k)}\right),I\left(x_{2i}^{(k+1)};\boldsymbol{u}\,\middle|\,x_i^{(k)}\right)\right)\\&=f_{x_i^{(k)}}(\tilde{s}_{2i-1}^{(k+1)},\tilde{s}_{2i}^{(k+1)})\\&=\tilde{s}_i^{(k)}.\end{aligned}$$

This establishes (97). Then

$$\mathbb{E}\left[\left(x_1^{(1)}-\mathbb{E}\left[x_1^{(1)}\,\middle|\,\boldsymbol{u}\right]\right)^2\right]=\sigma_{x_1^{(1)}}^2\exp\left(-2\tilde{s}_1^{(1)}\right)=d.$$

Thus, $\boldsymbol{u}$ is in $\mathcal{U}(d)$. If we let

$$\tilde{R}_i=I(x_i^{(L)};u_i\,|\,u_1,\ldots,u_{i-1})$$

then $(\tilde{\boldsymbol{R}},d)$ is in $\mathcal{RD}_{\text{in}}$. Since u_i is conditionally independent of $\boldsymbol{u}$ and all of the source variables given $x_i^{(L)}$, it follows that $\tilde{s}_i^{(L)}=0$ if and only if u_i is independent of all of the other variables. We will show that

$$\sum_{i=j}^{2^{L-1}}\tilde{R}_i=I\left(x_j^{(L)},\ldots,x_{2^{L-1}}^{(L)};u_j,\ldots,u_{2^{L-1}}\,|\,u_1,\ldots,u_{j-1}\right)$$

by induction. For $j=2^{L-1}$, this condition holds by the definition of $\tilde{R}_j$. Next suppose that the condition holds for j. Then by the tree structure

$$\begin{aligned}\sum_{i=j-1}^{2^{L-1}}\tilde{R}_i&=I\left(x_{j-1}^{(L)};u_{j-1}\,|\,u_1,\ldots,u_{j-2}\right)\\&\quad+I(x_j^{(L)},\ldots,x_{2^{L-1}}^{(L)};u_j,\ldots,u_{2^{L-1}}\,|\,u_1,\ldots,u_{j-1})\\&=I\left(x_{j-1}^{(L)},\ldots,x_{2^{L-1}}^{(L)};u_{j-1}\,|\,u_1,\ldots,u_{j-2}\right)\\&\quad+I\left(x_{j-1}^{(L)},\ldots,x_{2^{L-1}}^{(L)};u_j,\ldots,u_{2^{L-1}}\,|\,u_1,\ldots,u_{j-1}\right)\\&=I\left(x_{j-1}^{(L)},\ldots,x_{2^{L-1}}^{(L)};u_{j-1},\ldots,u_{2^{L-1}}\,|\,u_1,\ldots,u_{j-2}\right).\end{aligned}$$

Thus

$$\begin{aligned}&\inf_{\boldsymbol{R}:(\boldsymbol{R},d)\in\mathcal{RD}_{\text{in}}(d)}\sum_{i=1}^{2^{L-1}}\alpha_iR_i\\&\le\sum_{i=1}^{2^{L-1}}\alpha_i\tilde{R}_i\\&\le\sum_{j=1}^{2^{L-1}}(\alpha_j-\alpha_{j-1})I\left(x_j^{(L)},\ldots,\right.\\&\qquad\left.x_{2^{L-1}}^{(L)};u_j,\ldots,u_{2^{L-1}}\,|\,u_1,\ldots,u_{j-1}\right)\\&=\sum_{j=1}^{2^{L-1}}(\alpha_j-\alpha_{j-1})I\left(\boldsymbol{x}_{\mathcal{A}_j}^{(L)};\boldsymbol{u}_{\mathcal{A}_j}\,|\,\boldsymbol{u}_{\mathcal{A}_j^c}\right).\end{aligned}$$

By mimicking (49) through (55), one can show that

$$\begin{aligned}I&\left(\boldsymbol{x}_{\mathcal{A}_j}^{(L)};\boldsymbol{u}_{\mathcal{A}_j}\,|\,\boldsymbol{u}_{\mathcal{A}_j^c}\right)\\&=\sum_{k=1}^{L}\sum_{i\in\mathcal{A}_j^{(k)}}\left(\tilde{s}_i^{(k)}-I\left(x_i^{(k)};\boldsymbol{u}_{\mathcal{A}_j^c}\,|\,x_{\lfloor(i+1)/2\rfloor}^{(k-1)}\right)\right).\end{aligned}$$

But by Lemma 5 and the observation in Appendix A

$$I(x_i^{(k)};\boldsymbol{u}_{\mathcal{A}_j^c}\,|\,x_{\lfloor(i+1)/2\rfloor}^{(k-1)})=f_{x_i^{(k)}}^{\mathcal{A}_j^c}(\tilde{\boldsymbol{s}}_{\mathcal{A}_j^c}(x_i^{(k)})).$$

This establishes (96). It follows that

$$\inf_{\boldsymbol{R}:(\boldsymbol{R},d)\in\mathcal{RD}_{\text{in}}}\sum_{i=1}^{2^{L-1}}\alpha_iR_i\le\inf_{\boldsymbol{R}:(\boldsymbol{R},d)\in\mathcal{RD}_{\text{out}}}\sum_{i=1}^{2^{L-1}}\alpha_iR_i+2\epsilon.$$

Since ϵ was arbitrary, the proof is complete.

APPENDIX C
PROOF OF THEOREM 1

We must show that $\mathcal{RD}^* \subseteq \mathrm{co}(\mathcal{RD}_{\mathrm{in}})$. Since both sets are convex, it suffices to show that for any componentwise nonnegative vector $(\beta_1, \ldots, \beta_{2^{L-1}}, \beta)$

$$\inf_{(\boldsymbol{R},d)\in\mathcal{RD}^*} \sum_{i=1}^{2^{L-1}} \beta_i R_i + \beta d \geq \inf_{(\boldsymbol{R},d)\in\mathrm{co}(\mathcal{RD}_{\mathrm{in}})} \sum_{i=1}^{2^{L-1}} \beta_i R_i + \beta d = \inf_{(\boldsymbol{R},d)\in\mathcal{RD}_{\mathrm{in}}} \sum_{i=1}^{2^{L-1}} \beta_i R_i + \beta d. \tag{98}$$

We shall assume that $\beta_1 \leq \beta_2 \leq \cdots \leq \beta_{2^{L-1}}$; the other cases are similar. Let us temporarily use $\mathcal{RD}^*(\boldsymbol{K}_x)$ to denote the rate–distortion region for the binary tree-structure problem when the source variables have covariance matrix $\boldsymbol{K}_x$ and similarly for $\mathcal{RD}_{\mathrm{in}}(\boldsymbol{K}_x)$. If $\boldsymbol{K}_x$ is such that all of the noise variances are positive, then (98) follows from Lemma 2.

If some of the noise variances are zero, then let $\boldsymbol{K}_x^{(n)}$ be a sequence of source covariance matrices converging to $\boldsymbol{K}_x$ such that for each n, $\boldsymbol{K}_x^{(n)}$ corresponds to a source satisfying the binary tree structure for which all of the noise variances are positive. Then $\mathcal{RD}^*(\boldsymbol{K}_x^{(n)}) = \mathrm{co}(\mathcal{RD}_{\mathrm{in}}(\boldsymbol{K}_x^{(n)}))$ for each n, so

$$\inf_{(\boldsymbol{R},d)\in\mathcal{RD}^*(\boldsymbol{K}_x^{(n)})} \sum_{i=1}^{2^{L-1}} \beta_i R_i + \beta d = \inf_{(\boldsymbol{R},d)\in\mathcal{RD}_{\mathrm{in}}(\boldsymbol{K}_x^{(n)})} \sum_{i=1}^{2^{L-1}} \beta_i R_i + \beta d.$$

We will first show that

$$\liminf_{n\to\infty} \inf_{(\boldsymbol{R},d)\in\mathcal{RD}_{\mathrm{in}}(\boldsymbol{K}_x^{(n)})} \sum_{i=1}^{2^{L-1}} \beta_i R_i + \beta d \geq \inf_{(\boldsymbol{R},d)\in\mathcal{RD}_{\mathrm{in}}(\boldsymbol{K}_x)} \sum_{i=1}^{2^{L-1}} \beta_i R_i + \beta d. \tag{99}$$

For each n, there exists a set of auxiliary random variables $\boldsymbol{u}^{(n)}$ such that [12, Lemma 3.3]

$$\begin{aligned}\inf_{(\boldsymbol{R},d)\in\mathcal{RD}_{\mathrm{in}}(\boldsymbol{K}_x^{(n)})} & \sum_{i=1}^{2^{L-1}} \beta_i R_i + \beta d \\ &= \sum_{i=1}^{2^{L-1}} \beta_i I\left(u_i^{(n)}; x_i^{(L,n)} \,|\, x_1^{(L,n)}, \ldots, x_{i-1}^{(L,n)}\right) \\ &\quad + \beta \mathbb{E}\left\{\left(x_1^{(1,n)} - \mathbb{E}\left[x_1^{(1,n)} \,|\, \boldsymbol{u}^{(n)}\right]\right)^2\right\}.\end{aligned} \tag{100}$$

Here $x_i^{(L,n)}$ denotes the ith variable at depth L of the tree corresponding to covariance matrix $\boldsymbol{K}_x^{(n)}$. Now the auxiliary random variables $\boldsymbol{u}^{(n)}$ can be parametrized by a compact set, so consider a subsequence of $\boldsymbol{K}_x^{(n)}$ along which $\boldsymbol{u}^{(n)}$ converges in distribution to a limit $\boldsymbol{u}$ and the right-hand side of (100) converges to the $\liminf$. Then

$$\begin{aligned}\liminf_{n\to\infty} \inf_{(\boldsymbol{R},d)\in\mathcal{RD}_{\mathrm{in}}(\boldsymbol{K}_x^{(n)})} & \sum_{i=1}^{2^{L-1}} \beta_i R_i + \beta d \\ &= \sum_{i=1}^{2^{L-1}} \beta_i I\left(u_i; x_i^{(L)} \,|\, x_1^{(L)}, \ldots, x_{i-1}^{(L)}\right) \\ &\quad + \beta \mathbb{E}\left\{\left(\left(x_1^{(1)} - \mathbb{E}\left[x_1^{(1)} \middle| \boldsymbol{u}\right]\right)^2\right\} \\ &\geq \inf_{(\boldsymbol{R},d)\in\mathcal{RD}_{\mathrm{in}}(\boldsymbol{K}_x)} \sum_{i=1}^{2^{L-1}} \beta_i R_i + \beta d.\end{aligned}$$

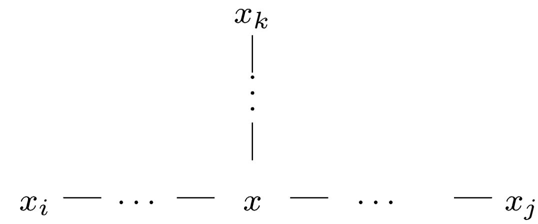

Fig. 12. x is the point at which the two paths split.

This establishes (99). On the other hand, Chen and Wagner [2] have shown that the rate–distortion region is inner-semi-continuous

$$\limsup_{n\to\infty} \inf_{(\boldsymbol{R},d)\in\mathcal{RD}^*(\boldsymbol{K}_x^{(n)})} \sum_{i=1}^{2^{L-1}} \beta_i R_i + \beta d \leq \inf_{(\boldsymbol{R},d)\in\mathcal{RD}^*(\boldsymbol{K}_x)} \sum_{i=1}^{2^{L-1}} \beta_i R_i + \beta d.$$

Together with (99), this establishes (98) and hence Theorem 1.

APPENDIX D
PROOF OF PROPOSITION 2

Suppose that $x_1, \ldots, x_N$ can be embedded in a Gauss–Markov tree and fix distinct indices i, j, and k. Without loss of generality, we may assume that all variables in the tree have mean zero and variance one. Consider two paths (i.e., two sequences of variables), one from x_i to x_j and one from x_i to x_k. Evidently both paths contain x_i; let x denote the last variable in the first path that is contained in the second. This is the point at which the two paths split, as shown in Fig. 12. Note that it is possible for x to equal x_i, x_j, or x_k.

Now since x is along the path from x_i to x_j, it follows from the tree condition that $x_i \leftrightarrow x \leftrightarrow x_j$. Likewise, $x_i \leftrightarrow x \leftrightarrow x_k$. Since all of the variables are standard Normals, this implies [17, eq. (5.13)]

$$\rho_{ij} = \mathbb{E}[x_i x]\mathbb{E}[x x_j] \tag{101}$$

$$\rho_{ik} = \mathbb{E}[x_i x]\mathbb{E}[x x_k]. \tag{102}$$

Next consider the paths from x_j to x_i and from x_j to x_k, and let $\tilde{x}$ denote the last variable in the first path that is contained in the second. Then both x and $\tilde{x}$ lie along the path from x_j to x_i. If $x \neq \tilde{x}$, then the path from x to x_k to $\tilde{x}$ to x would form a loop, which is impossible since the graph is a tree. Thus, $\tilde{x}$ must equal x. Thus, $x_j \leftrightarrow x \leftrightarrow x_k$ and

$$\rho_{jk} = \mathbb{E}[x_j x]\mathbb{E}[x x_k].$$

Combining this equation with (101) and (102) yields conditions (40) and (41).

Now suppose that $N = 3$ and conditions (40) and (41) hold. If ρ_{ij} is nonzero for all $i \neq j$, then

$$0 < \frac{\rho_{ij}\rho_{ik}}{\rho_{jk}} \leq 1$$

for all distinct i, j, and k. This implies that x_1, x_2, and x_3, can be written

$$\begin{aligned} x_1 &= \sqrt{\frac{\rho_{12}\rho_{13}}{\rho_{23}}} \cdot \mathrm{sgn}(\rho_{23}) \cdot x_0 + z_1 \\ x_2 &= \sqrt{\frac{\rho_{12}\rho_{23}}{\rho_{13}}} \cdot \mathrm{sgn}(\rho_{13}) \cdot x_0 + z_2 \\ x_3 &= \sqrt{\frac{\rho_{13}\rho_{23}}{\rho_{12}}} \cdot \mathrm{sgn}(\rho_{12}) \cdot x_0 + z_3, \end{aligned}$$

where $\mathrm{sgn}(\,\cdot\,)$ is the signum function

$$\mathrm{sgn}(\rho) = \begin{cases} 1, & \text{if } \rho > 0 \\ 0, & \text{if } \rho = 0 \\ -1, & \text{if } \rho < 0 \end{cases}$$

and where x_0, z_1, z_2, z_3 are independent Gaussian random variables. Here x is a standard Normal and the variances of the z's are chosen to such that the x's have unit variance. It is readily verified that this construction yields the correct correlation coefficients among the x's. It is then clear that x_0 and the x's can be arranged in the Gauss–Markov tree shown in Fig. 9.

If, say, $\rho_{12} = 0$, then by condition (40), either $\rho_{13} = 0$ or $\rho_{23} = 0$. Suppose that $\rho_{13} = 0$. Then x_1 is uncorrelated, and hence independent, of x_2 and x_3. It follows that the x's can be written

$$\begin{aligned} x_1 &= z_1 \\ x_2 &= \sqrt{|\,\rho_{23}\,|} \cdot x_0 + z_2 \\ x_3 &= \sqrt{|\,\rho_{23}\,|} \cdot \mathrm{sgn}(\rho_{23}) \cdot x_0 + z_3, \end{aligned}$$

so that the x_0 and the x's can again be arranged in the Gauss-Markov tree shown in Fig. 9.

Appendix E
Proof of Proposition 3

Since we are assuming that $R_3 = 0$, the problem effectively reduces to a two-encoder setup. By Lemma 1 and (19), the minimum $R_1 + R_2$ equals

$$\begin{aligned} \inf \quad & I(\boldsymbol{x}; \boldsymbol{u}) \\ \text{subject to} \quad & u_1 \leftrightarrow x_1 \leftrightarrow x_2 \leftrightarrow u_2 \\ & (\boldsymbol{x}, \boldsymbol{u}) \text{ jointly Gaussian} \\ & \mathbb{E}[(x_3 - \mathbb{E}[x_3 \,|\, \boldsymbol{u}])^2] \leq d. \end{aligned}$$

Without loss of generality, we may assume that

$$\begin{aligned} u_1 &= x_1 + z_1 \\ u_2 &= x_2 + z_2 \end{aligned}$$

where the z variables are Gaussian and independent of each other and $\boldsymbol{x}$. Let z_1 have variance $\alpha\sigma^2$ and z_2 have variance $\beta\sigma^2$.

Via straightforward calculations one can show that

$$\begin{aligned} & I(\boldsymbol{x}; \boldsymbol{u}) \\ & \quad = \frac{1}{2} \log\left((1-\rho^2)\alpha^{-1}\beta^{-1} + \alpha^{-1} + \beta^{-1} + 1\right) \end{aligned}$$

and

$$\begin{aligned} & \mathbb{E}[(x_3 - \mathbb{E}[x_3 \,|\, \boldsymbol{u}])^2] \\ & \quad = 1 - \frac{1}{\sigma^2} \frac{2(1+\rho) + \alpha + \beta}{4(1+\alpha)(1+\beta) - 4\rho^2}. \end{aligned} \tag{103}$$

Now

$$\begin{aligned} \frac{2(1+\rho) + \alpha + \beta}{4(1+\alpha)(1+\beta) - 4\rho^2} &\leq \frac{4 + \alpha + \beta}{4\alpha + 4\beta + 4\alpha\beta} \\ &\leq \frac{1 + \alpha + \beta}{\alpha + \beta + \alpha\beta} \\ &\leq \frac{1}{\alpha\beta} + 2. \end{aligned}$$

It follows that as σ^2 tends to infinity, in order to continue to meet the distortion constraint, we require that $\alpha\beta$ tend to zero. But this implies that $I(\boldsymbol{x}; \boldsymbol{u})$ tends to infinity, by (103).

Appendix F
Proof of Proposition 4

Since the average distortion is the same for all ℓ, let us assume that $\ell = 1$ and write x_3 in place of $x_3(1)$ and likewise for the other variables. Then by the triangle inequality

$$\begin{aligned} & \sqrt{\mathbb{E}[(x_3 - \hat{x}_3)^2]} \\ & \quad \leq \sqrt{\mathbb{E}[(x_3 - (\tilde{x}_1 - \tilde{x}_2))^2]} + \sqrt{\mathbb{E}[((\tilde{x}_1 - \tilde{x}_2) - \hat{x}_3)^2]}. \end{aligned}$$

Now

$$|\,x_1 - \tilde{x}_1\,| \leq 2^{-(n+1)}$$

and likewise for $|\,x_2 - \tilde{x}_2\,|$. Thus

$$\mathbb{E}[(x_3 - (\tilde{x}_1 - \tilde{x}_2))^2] \leq 2^{-2n}.$$

Define the event

$$A = \{\,|\,\tilde{x}_1 - \tilde{x}_2\,| < 2^{m-1}\}.$$

Now on A

$$\begin{aligned} \hat{x}_3 &= u_1 - u_2 \bmod \Lambda_o \\ &= \tilde{x}_1 - \tilde{x}_2 \bmod \Lambda_o \\ &= \tilde{x}_1 - \tilde{x}_2 \end{aligned}$$

so

$$\begin{aligned} \mathbb{E}[(\tilde{x}_1 - \tilde{x}_2 - \hat{x}_3)^2] &= \mathbb{E}[(\tilde{x}_1 - \tilde{x}_2 - \hat{x}_3)^2 1_{A^c}] \\ &\leq \sqrt{\mathbb{E}[(\tilde{x}_1 - \tilde{x}_2 - \hat{x}_3)^4]\mathbb{P}(A^c)}. \end{aligned}$$

But

$$\begin{aligned} |\tilde{x}_1 - \tilde{x}_2 - \hat{x}_3| &\leq |x_1 - x_2| + |\tilde{x}_1 - x_1| + |x_2 - \tilde{x}_2| + |\hat{x}_3| \\ &\leq |x_1 - x_2| + 2^{-n} + 2^{m-1} \\ &\leq |x_1 - x_2| + 2^m. \end{aligned}$$

Since $x_1 - x_2$ is a standard Normal random variable, $\mathbb{E}[(x_1 - x_2)^4] = 3$, and Minkowski's inequality implies

$$\mathbb{E}[(\tilde{x}_1 - \tilde{x}_2 - \hat{x}_3)^4] \le 3 + 2^m.$$

It only remains to bound $\mathbb{P}(A^c)$. Using a well-known upper bound on the tail of the Gaussian distribution

$$\mathbb{P}(A^c) \le 2\exp(-2^{2m-3}).$$

Combining these various bounds gives

$$\mathbb{E}[(x_3 - \hat{x}_3)^2] \le (2^{-n} + (2(3+2^m)\exp(-2^{2m-3}))^{1/2})^2$$

Proposition 4 follows.

Appendix G
Proof of Proposition 5

Recall we may assume that all of the variables have unit variance. By Proposition 2, if x_1, x_2, and x_3 cannot be embedded in a Gauss–Markov tree, then either

$$\rho_{12}\rho_{13}\rho_{23} < 0 \tag{104}$$

or

$$|\rho_{ij}| < |\rho_{ik}\rho_{kj}| \tag{105}$$

for some distinct i, j, and k. Suppose first that (104) holds. Then we must have $|\rho_{ij}| < 1$ for all $i \ne j$. Now

$$\begin{aligned}\mathbb{E}&[x_1 \,|\, x_2, x_3]\\ &= \frac{\rho_{12} - \rho_{13}\rho_{23}}{1-\rho_{23}^2}\cdot x_2 + \frac{\rho_{13} - \rho_{12}\rho_{23}}{1-\rho_{23}^2}\cdot x_3\\ &\stackrel{\text{def}}{=} a_2x_2 + a_3x_3.\end{aligned}$$

Then

$$\begin{aligned}a_2\cdot a_3\cdot\rho_{23} = {}& \frac{1}{(1-\rho_{23}^2)^2}\frac{\rho_{23}}{\rho_{13}\rho_{12}}\\ &\times\left(\rho_{12}^2 - \rho_{12}\rho_{13}\rho_{23}\right)\left(\rho_{13}^2 - \rho_{12}\rho_{13}\rho_{23}\right)\end{aligned} \tag{106}$$

which is negative by (104). This establishes the desired conclusion in this case. We will therefore assume throughout the remainder of the proof that $\rho_{12}\rho_{13}\rho_{23} \ge 0$.

Suppose that (105) holds, say, for $i = 1$, $j = 2$, and $k = 3$. Then we must have $|\rho_{12}| < 1$ and $\rho_{13}\cdot\rho_{23} \ne 0$. Furthermore, if $|\rho_{23}| = 1$, then $|\rho_{12}| = |\rho_{13}|$, which would contradict (105). Thus, we may assume that $|\rho_{23}| < 1$. First suppose that $\rho_{12} = 0$. Then

$$a_2\cdot a_3\cdot\rho_{23} = -\frac{\rho_{13}^2\rho_{23}^2}{(1-\rho_{23}^2)^2}$$

which is negative. We will therefore focus on the case in which $\rho_{12}\rho_{13}\rho_{23} > 0$.

Next observe that since we are assuming that (105) holds for $i = 1$, $j = 2$, and $k = 3$, the opposite inequality must hold strictly in the other two cases

$$\begin{aligned}|\rho_{13}| &> |\rho_{12}\rho_{23}|\\ |\rho_{23}| &> |\rho_{12}\rho_{13}|.\end{aligned}$$

This can be seen by contradiction: if, e.g., $|\rho_{13}| \le |\rho_{12}\rho_{23}|$, then combining this fact with (105) yields

$$|\rho_{12}| < |\rho_{13}\rho_{23}| \le |\rho_{12}||\rho_{23}|^2$$

which is evidently false. From (106) and the three assumed conditions, $\rho_{12}\rho_{13}\rho_{23} > 0$, $|\rho_{12}| < |\rho_{13}\rho_{23}|$, and $|\rho_{13}| > |\rho_{12}\rho_{23}|$, it follows that $a_2\cdot a_3\cdot\rho_{23}$ is negative, as desired.

References

[1] T. Berger, *Multiterminal Source Coding*, ser. The Information Theory Approach to Communications. New York: Springer-Verlag, 1978, vol. 229, CISM courses and lectures.

[2] J. Chen and A. B. Wagner, "A semicontinuity theorem and its application to network source coding," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Toronto, ON, Canada, Jul. 2008, pp. 429–433.

[3] J. Chen, X. Zhang, T. Berger, and S. B. Wicker, "An upper bound on the sum-rate distortion function and its corresponding rate allocation schemes for the CEO problem," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 977–987, Aug. 2004.

[4] S. Hanly and D. N. C. Tse, "Multiple-access fading channels: Part II: Delay-limited capacities," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2816–2831, Nov. 1998.

[5] D. Krithivasan and S. S. Pradhan, "Lattices for Distributed Source Coding: Jointly Gaussian Sources and Reconstruction of a Linear Function," arXiv:0707.3461.

[6] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1912–1923, Nov. 1997.

[7] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.

[8] Y. Oohama, "Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2577–2593, Jul. 2005.

[9] Y. Oohama, "Gaussian multiterminal source coding with several side informations at the decoder," in *Proc. IEEE Symp. Information Theory (ISIT)*, Seattle, WA, Jul. 2006, pp. 1409–1413.

[10] V. Prabhakaran, D. Tse, and K. Ramchandran, "Rate region of the quadratic Gaussian CEO problem," in *Proc. IEEE Symp. Information Theory (ISIT)*, Chicago, IL, Jun./Jul. 2004, p. 119.

[11] T. P. Speed and H. T. Kiiveri, "Gaussian Markov distributions over finite graphs," *Ann. Statist.*, vol. 14, no. 1, pp. 138–150, Mar. 1986.

[12] D. N. C. Tse and S. Hanly, "Multiple-access fading channels: Part I: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2796–2815, Nov. 1998.

[13] S.-Y. Tung, "Multiterminal Source Coding," Ph.D. dissertation, Cornell University, Ithaca, NY, 1978.

[14] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-encoder source-coding problem," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1938–1961, May 2008.

[15] H. Wang and P. Viswanath, "Vector Gaussian multiple description for individual and central receivers," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2133–2153, Jun. 2007.

[16] D. J. A. Welsh, *Matroid Theory*. London, U.K.: Academic, 1976.

[17] E. Wong and B. Hajek, *Stochastic Processes in Engineering Systems*. New York: Springer-Verlag, 1985.

Saurabha Tavildar received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Bombay, in 2001 and the M.S. and Ph.D. degrees from the University of Illinois at Urbana-Champaign, Urbana, IL, in 2003 and 2006, respectively.

He is currently working at the Qualcomm Flarion Technologies, Bridgewater, NJ. His research interests are wireless systems design and information theory.

Pramod Viswanath (S'98–M'03) received the Ph.D. degree in electrical engineering and computer science from the University of California at Berkeley in 2000.

He was a Member of Technical Staff at Flarion Technologies, Bridgewater, NJ, until August 2001 before joining the Department or Electrical and Computer Engineering at the University of Illinois, Urbana-Champaign.

Dr. Viswanath is a recipient of the Eliahu Jury Award from the EECS Department of U.C. Berkeley (2000), the Bernard Friedman Award from the Mathematics Department of UC Berkeley (2000), and the NSF CAREER Award (2003). He was an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY for the period 2006–2008.

Aaron B. Wagner (S'00–M'05) received the B.S. degree in electrical engineering from the University of Michigan, Ann Arbor, in 1999 and the M.S. and Ph.D. degrees in electrical engineering and computer sciences from the University of California, Berkeley, in 2002 and 2005, respectively.

During the 2005–2006 academic year, he was a Postdoctoral Research Associate in the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign and a Visiting Assistant Professor in the School of Electrical and Computer Engineering at Cornell University, Ithaca, NY. Since 2006, he has been an Assistant Professor in the School of Electrical and Computer Engineering at Cornell.

Prof. Wagner has received the NSF CAREER award, the David J. Sakrison Memorial Prize from the U.C. Berkeley EECS Department, the Bernard Friedman Memorial Prize in Applied Mathematics from the U.C. Berkeley Department of Mathematics, the Michael Tien '72 Excellence in Teaching Award, and the Distinguished Achievement Award in Electrical Engineering from the University of Michigan College of Engineering.