Distributed Information-Theoretic Biclustering

Georg Pichler*, Pablo Piantanida†, and Gerald Matz*

*Institute of Telecommunications, Vienna University of Technology, Vienna, Austria
†Laboratoire des Signaux et Systèmes (L2S), CentraleSupélec-CNRS-Université Paris-Sud, Gif-sur-Yvette, France
Email: {georg.pichler,gerald.matz}@nt.tuwien.ac.at; pablo.piantanida@centralesupelec.fr

***Abstract*—This paper investigates the problem of distributed biclustering of memoryless sources and extends previous work [1] to the general case with more than two sources. Given a set of distributed stationary memoryless sources, the encoders' goal is to find rate-limited representations of these sources such that the mutual information between two selected subsets of descriptions (each of them generated by distinct encoders) is maximized. This formulation is fundamentally different from conventional distributed source coding problems since here redundancy among descriptions should actually be maximally preserved. We derive non-trivial outer and inner bounds to the achievable region for this problem and further connect them to the CEO problem under logarithmic loss distortion. Since information-theoretic biclustering is closely related to distributed hypothesis testing against independence, our results are also expected to apply to that problem.**

I. Introduction

Clustering unstructured data is desirable to understand the nature of unstructured data with the goal of making information more accessible to humans. Clustering algorithms have been very successfully used in wide range of applications in areas like life sciences (gene expression analysis [2], segmentation of PET images [3]) and marketing research [4].

In this paper, we study an information-theoretic formulation of the biclustering problem that extends our recent work [1] to the case of multiple (usually dependent) sources. Each source is assumed to be stationary and memoryless and is observed by a distinct encoder. The encoders aim at extracting lossy (rate-limited) descriptions of the associated source such that these descriptions are maximally informative about each other. In this setting, we investigate the optimal tradeoff between *relevance*, the normalized multi-letter mutual information between the descriptions, and *complexity*, the encoding rates. We refer to this scenario as the distributed multi-terminal *information-theoretic biclustering* (ITB) problem and derive an inner bound to the sets of achievable rates of relevance and complexity.

The problem at hand is fundamentally different from conventional distributed source coding problems, which aim at discarding redundant information while guaranteeing correct decoding. In contrast to the ITB problem, where redundant information should actually be maximally preserved. Furthermore, the ITB problem is closely related to testing against independence with multi-terminal data compression [5], for which a general solution remains elusive [6].

While the outer bound follows directly from the case of two sources, the straightforward generalization of the achievability result [1, Theorem 6] yields a very poor inner bound. This bound can be improved by correctly accounting for the possibility of encoders that act as helpers, which can be achieved using a binning strategy, that is, a variant of Berger-Tung coding. Nevertheless, the inner bound cannot be tight, as the well-known Körner-Marton problem [7] becomes a special case of the considered problem when considering more than two sources. Thus, in general Berger-Tung coding is suboptimal.

We further investigate the CEO problem under a mutual information constraint. This is another special case of ITB, and we show that it becomes equivalent to classical multiterminal lossy source coding under logarithmic loss distortion. By leveraging this equivalence, we obtain tight bounds for this special case using methods from [8].

Notation

We denote random quantities and their realizations by capital, sans-serif and lowercase letters, respectively. Furthermore, vectors are indicated by bold-face type and sets by calligraphic type. We use subscript to denote slices of vectors, i.e., $\boldsymbol{x}_{\mathcal{A}} \triangleq (\boldsymbol{x}_i)_{i\in\mathcal{A}}$ as well as the usual notation $\boldsymbol{x}_i^j \triangleq \boldsymbol{x}_{\{i,\dots,j\}}$, $\boldsymbol{x}^j \triangleq \boldsymbol{x}_1^j$. If a vector is already carrying a subscript, it will be separated by a comma, e.g., $\boldsymbol{x}_{3,1}^5 = (\boldsymbol{x}_3)_1^5 = (\boldsymbol{x}_3)^5$. Random variables are assumed to be supported on finite sets. We use the same letter for the random variable and for its support set, e.g., Y takes values in $\mathcal{Y}$ and X_3 takes values in $\mathcal{X}_3$. For a set $\mathcal{X}$, let $\mathcal{X}^n$ denote the n-th Cartesian power of $\mathcal{X}$. Given a random variable X, we write $\mathrm{p}_{\mathsf{X}} \in \mathcal{P}(\mathcal{X})$ for its probability mass function (pmf), where $\mathcal{P}(\mathcal{X})$ is the set of all pmfs on $\mathcal{X}$. We write $\mathsf{X} \sim \mathrm{p}$ to indicate that X is distributed according to $\mathrm{p} \in \mathcal{P}(\mathcal{X})$. We use the notation of [9, Chapter 2] for information-theoretic quantities, however, all logarithms in this paper are to base e and information theoretic quantities are measured in nats. The notation $\mathrm{h}_0(p) \triangleq -p\log p - (1-p)\log(1-p)$ is used for the binary entropy function and $\mathsf{X}\multimap\mathsf{Y}\multimap\mathsf{Z}$ indicates that X, Y, and Z form a Markov chain in this order. For convenience, we define $\mathcal{K} \triangleq \{1,2,\dots,K\}$ and we use the symbol $x_{2^{\mathcal{K}},2^{\mathcal{K}}}$ for a tuple, indexed by disjoint unordered pairs $\{\mathcal{A},\mathcal{B}\}$ of nonempty sets $\mathcal{A},\mathcal{B}\subset\mathcal{K}$ and write $x_{\mathcal{A},\mathcal{B}}$ for its components. For a random variable X, we denote the set of (strongly) δ-typical n-sequences [10, Section 2.4] as $\mathcal{T}^n_{[\mathsf{X}]\delta}$. Regarding the method of types, which is used heavily in Section IV, the reader is referred to [11].

II. Problem Definition and Main Results

A. Problem Statement

Let $\mathsf{X}_{\mathcal{K}}$ be K random variables. The random vectors $\mathbf{X}_{\mathcal{K}}$ consist of n i.i.d. copies of $\mathsf{X}_{\mathcal{K}}$. For $n\in\mathbb{N}$ and $R_{\mathcal{K}}\in\mathbb{R}^K$, an $(n,R_{\mathcal{K}})$ code $f_{\mathcal{K}}$ consists of K functions $f_k\colon \mathcal{X}_k^n \to \mathcal{M}_k$, where $\mathcal{M}_k$ is an arbitrary finite set with $\log|\mathcal{M}_k| \le nR_k$ for each $k\in\mathcal{K}$.

Definition 1 (Relevance)**.** *Consider an $(n,R_{\mathcal{K}})$ code $f_{\mathcal{K}}$ and let $\mathsf{W}_k \triangleq f_k(\mathbf{X}_k)$. For any pair of disjoint nonempty sets $\mathcal{A},\mathcal{B}\subset\mathcal{K}$, we define the co-information of $f_{\mathcal{A}}$ and $f_{\mathcal{B}}$ as*

$$\Theta(f_{\mathcal{A}};f_{\mathcal{B}}) \triangleq \frac{1}{n}\mathrm{I}(\mathsf{W}_{\mathcal{A}};\mathsf{W}_{\mathcal{B}}).$$

Definition 2 (Achievability and relevance-rates region)**.** *A point $(\mu_{2^{\mathcal{K}},2^{\mathcal{K}}}, R_{\mathcal{K}})$ is achievable if there exists an $(n, R_{\mathcal{K}})$ code $f_{\mathcal{K}}$ for some $n \in \mathbb{N}$ such that*

$$\Theta(f_{\mathcal{A}}; f_{\mathcal{B}}) \geq \mu_{\mathcal{A},\mathcal{B}} \ , \quad \forall \mathcal{A}, \mathcal{B} \subset \mathcal{K} \text{ disjoint, nonempty.}$$

The achievable relevance-rates region $\overline{\mathcal{R}}$ is the closure of the set $\mathcal{R}$ of achievable points.

B. Main Results

Theorem 3 (Outer bound)**.** *We have $\mathcal{R} \subseteq \mathcal{R}_{\mathrm{o}}$, where the outer bound $\mathcal{R}_{\mathrm{o}}$ is the set of all points $(\mu_{2^{\mathcal{K}},2^{\mathcal{K}}}, R_{\mathcal{K}})$ such that there exist random variables $\mathsf{U}_{\mathcal{K}}$ with $\mathsf{U}_{\mathcal{A}} \multimap \mathsf{X}_{\mathcal{A}} \multimap \mathsf{X}_{\mathcal{K}}$ for any $\mathcal{A} \subseteq \mathcal{K}$ and*

$$\sum_{k\in\mathcal{A}} R_k \geq \mathrm{I}(\mathsf{U}_{\mathcal{A}}; \mathsf{X}_{\mathcal{K}}|\mathsf{U}_{\mathcal{C}}) \quad \forall \mathcal{C} \subseteq \mathcal{K},$$
$$\mu_{\mathcal{A},\mathcal{B}} \leq \mathrm{I}(\mathsf{U}_{\mathcal{A}}; \mathsf{X}_{\mathcal{A}}) + \mathrm{I}(\mathsf{U}_{\mathcal{B}}; \mathsf{X}_{\mathcal{B}}) - \mathrm{I}(\mathsf{U}_{\mathcal{A}}\mathsf{U}_{\mathcal{B}}; \mathsf{X}_{\mathcal{A}}\mathsf{X}_{\mathcal{B}}),$$

for all $k \in \mathcal{K}$ and all disjoint nonempty sets $\mathcal{A}, \mathcal{B} \subset \mathcal{K}$.

Proof: The proof of this theorem follows from standard information-theoretic arguments using the auxiliary random variable identification $\mathsf{U}_{k,i} \triangleq (f(\mathbf{X}_k), \mathbf{X}_{k,1}^{i-1})$. ∎

Definition 4. *A point $(\mu_{2^{\mathcal{K}},2^{\mathcal{K}}}, R_{\mathcal{K}})$ is in the region $\mathcal{R}_{\mathrm{i}}$ if there exist random variables $\mathsf{U}_{\mathcal{K}}$ satisfying $\mathsf{U}_k \multimap \mathsf{X}_k \multimap (\mathsf{X}_{\mathcal{K}\setminus k}, \mathsf{U}_{\mathcal{K}\setminus k})$ for all $k \in \mathcal{K}$, and if for every pair of disjoint nonempty index sets $\mathcal{A}, \mathcal{B} \subset \mathcal{K}$ there exist subsets $\mathcal{A}_b \subseteq \mathcal{A}_a \subseteq \mathcal{A}$ and $\mathcal{B}_b \subseteq \mathcal{B}_a \subseteq \mathcal{B}$ such that*

$$\sum_{k\in\mathcal{A}'} R_k \geq \mathrm{I}\big(\mathsf{X}_{\mathcal{A}'}; \mathsf{U}_{\mathcal{A}'} \big| \mathsf{U}_{\mathcal{A}_a\setminus\mathcal{A}'}\big), \tag{1}$$
$$\sum_{k\in\mathcal{B}'} R_k \geq \mathrm{I}\big(\mathsf{X}_{\mathcal{B}'}; \mathsf{U}_{\mathcal{B}'} \big| \mathsf{U}_{\mathcal{B}_a\setminus\mathcal{B}'}\big), \tag{2}$$
$$\mu_{\mathcal{A},\mathcal{B}} \leq \mathrm{I}(\mathsf{U}_{\mathcal{A}_b}; \mathsf{U}_{\mathcal{B}_b}), \tag{3}$$

for all subsets $\mathcal{A}' \subseteq \mathcal{A}_a$ with $\mathcal{A}' \cap \mathcal{A}_b \neq \varnothing$ and $\mathcal{B}' \subseteq \mathcal{B}_a$ with $\mathcal{B}' \cap \mathcal{B}_b \neq \varnothing$.

We use typicality coding and binning to show that $\mathcal{R}_{\mathrm{i}}$ is indeed achievable. The conditions (1) and (2) ensure that $\mathsf{U}_{\mathcal{A}_b}$ and $\mathsf{U}_{\mathcal{B}_b}$ can be correctly decoded from the output of the encoders $\mathcal{A}_a$ and $\mathcal{B}_a$, respectively. By (3), $\mathsf{U}_{\mathcal{A}_b}$ and $\mathsf{U}_{\mathcal{B}_b}$ are enough to ensure that $\mu_{\mathcal{A},\mathcal{B}}$ is achievable. Intuitively, one can say that the encoders $\mathcal{A}_a\setminus\mathcal{A}_b$ and $\mathcal{B}_a\setminus\mathcal{B}_b$ act as helpers for decoding $\mathsf{U}_{\mathcal{A}_b}$ and $\mathsf{U}_{\mathcal{B}_b}$, respectively. The special case $\mathcal{A}_b = \mathcal{A}$, $\mathcal{B}_b = \mathcal{B}$ for every disjoint pair of nonempty sets $\mathcal{A}, \mathcal{B} \subset \mathcal{K}$ corresponds to no binning at all, as (1) and (2) then imply $R_k \geq \mathrm{I}(\mathsf{X}_k; \mathsf{U}_k)$ for all $k \in \mathcal{K}$.

Theorem 5 (Inner bound)**.** $\mathcal{R}_{\mathrm{i}} \subseteq \overline{\mathcal{R}}$.

Proof: See Section IV. ∎

This achievability result cannot be shown by merely applying Berger-Tung coding. In contrast to averaged per-letter distortion, joint typicality alone is not sufficient to ensure that relevance is high enough. Therefore, the proof of this achievable region utilizes more sophisticated tools developed for analyzing hypothesis testing problems [5].

Remark. Note that the inner bound in Theorem 5 cannot be tight in general as Definition 2 contains the Körner-Marton problem [7] as a special case. Choose for $K = 3$, $\mathsf{X}_1 \sim \mathcal{B}(\frac{1}{2})$ and $\mathsf{X}_3 \sim \mathcal{B}(p)$ with $p \in (0,1)$ and $p \neq \frac{1}{2}$. Then define $\mathsf{X}_2 \triangleq \mathsf{X}_1 \oplus \mathsf{X}_3$. The point $(\mu_{2^{\mathcal{A}},2^{\mathcal{B}}}, R_{\mathcal{K}})$ where $R_3 = \log(2)$, $R_1 = R_2 = \mathrm{H}(\mathsf{X}_3) = \mathrm{h}_0(p)$, and $\mu_{\mathcal{A},\mathcal{B}} = 0$ except for $\mu_{\{1,2\},\{3\}} = \mathrm{H}(\mathsf{X}_3) = \mathrm{h}_0(p)$ is achievable [7, Theorem 1]. However, the quantize-and-bin scheme cannot achieve this point [7, Proposition 1].

$\mathcal{R}_{\mathrm{i}}$ is not convex in general, thus Theorem 5 can be strengthened to $\mathrm{conv}(\mathcal{R}_{\mathrm{i}}) \subseteq \overline{\mathcal{R}}$. However, characterizing $\mathrm{conv}(\mathcal{R}_{\mathrm{i}})$ using a time-sharing random variable is tedious, due to the choice of index sets $\mathcal{A}_a$, $\mathcal{A}_b$, $\mathcal{B}_a$, and $\mathcal{B}_b$ in Definition 4.

The following result establishes cardinality bounds that render $\mathcal{R}_{\mathrm{i}}$ computable.

Proposition 6. *The region $\mathcal{R}_{\mathrm{i}}$ remains the same if the cardinality bound $|\mathcal{U}_k| \leq |\mathcal{X}_k| + 4^K$ is imposed for every $k \in \mathcal{K}$.*

Proof: The proof is a straightforward application of the support lemma [10, Appendix C]. ∎

III. A Special Case: The CEO Problem

In this section we will analyze a special case of the biclustering problem, a variant of the CEO problem [12], where the quality is measured in terms of mutual information (MI) instead of a distortion criterion. Consider random variables $(\mathsf{X}_{\mathcal{J}}, \mathsf{Y}_{\mathcal{L}})$ with $\mathcal{J} = \{1, 2, \ldots, J\}$ and $\mathcal{L} = \{1, 2, \ldots, L\}$. Slightly abusing notation, we write $x_{2^{\mathcal{J}},2^{\mathcal{L}}}$ for a tuple indexed by a pair of nonempty sets $(\mathcal{A}, \mathcal{B})$ where $\mathcal{A} \subseteq \mathcal{J}$ and $\mathcal{B} \subseteq \mathcal{L}$.

Definition 7. *A point $(\mu_{2^{\mathcal{J}},2^{\mathcal{L}}}, R_{\mathcal{J}})$ is MI-achievable, if for some $n \in \mathbb{N}$ there exists an $(n, R_{\mathcal{J}})$ code $f_{\mathcal{J}}$ for $\mathsf{X}_{\mathcal{J}}$ with $\mathsf{U}_j \triangleq f_j(\mathbf{X}_j)$ such that for all $\mathcal{A} \subseteq \mathcal{J}$ and $\mathcal{B} \subseteq \mathcal{L}$*

$$\frac{1}{n}\mathrm{I}(\mathsf{U}_{\mathcal{A}}; \mathbf{Y}_{\mathcal{B}}) \geq \mu_{\mathcal{A},\mathcal{B}}.$$

Let $\mathcal{R}_{\mathrm{MI}}$ denote the set of all MI-achievable points.

We can obtain $\mathcal{R}_{\mathrm{MI}}$ as a special case from $\mathcal{R}$. Consider $\mathsf{X}_{\mathcal{K}} = (\mathsf{X}_{\mathcal{J}}, \mathsf{Y}_{\mathcal{L}})$ with $J + L = K$. We want to find achievable points $(\mu_{2^{\mathcal{K}},2^{\mathcal{K}}}, R_{\mathcal{K}}) \in \mathcal{R}$ that correspond to the CEO problem. Let $\mathsf{Y}_{\mathcal{L}}$ be transmitted without compression, i.e., $R_{J+l} = \log|\mathcal{Y}_l|$ for $l \in \mathcal{L}$. Furthermore, we are only interested in the information the encodings of $\mathbf{X}_{\mathcal{J}}$ provide about $\mathbf{Y}_{\mathcal{L}}$, i.e., $\mu_{\mathcal{A},\mathcal{B}} \neq 0$ only if $\mathcal{A} \subseteq \mathcal{J}$ and $\mathcal{B} \subseteq J + \mathcal{L}$.

We want to show that the above CEO problem is equivalent to the logarithmic loss (log-loss) distortion approach from [8]. For $\mathcal{A} \subseteq \mathcal{J}$ and $\mathcal{B} \subseteq \mathcal{L}$, we define a decoding function $g_{\mathcal{A},\mathcal{B}} \colon \mathcal{M}_{\mathcal{A}} \to \mathcal{P}(\mathcal{Y}_{\mathcal{B}}^n)$, which produces a probabilistic estimate of $\mathsf{Y}_{\mathcal{B}}$ given the output of the encoders $\mathcal{A}$. We use log-loss fidelity $\zeta_{\mathcal{B}}$, $\mathcal{B} \subseteq \mathcal{L}$, defined as the negative log-loss distortion $\mathrm{d}_{\mathcal{B}}(\boldsymbol{y}_{\mathcal{B}}, \mathrm{p}) \triangleq -\frac{1}{n}\log \mathrm{p}(\boldsymbol{y}_{\mathcal{B}})$ on $\mathcal{Y}_{\mathcal{B}}^n$ augmented by entropy, i.e.,

$$\zeta_{\mathcal{B}} \colon \mathcal{Y}_{\mathcal{B}}^n \times \mathcal{P}(\mathcal{Y}_{\mathcal{B}}^n) \to \mathbb{R},$$
$$(\boldsymbol{y}_{\mathcal{B}}, \mathrm{p}) \mapsto \mathrm{H}(\mathsf{Y}_{\mathcal{B}}) - \mathrm{d}_{\mathcal{B}}(\boldsymbol{y}_{\mathcal{B}}, \mathrm{p}).$$

Definition 8 (Rate-distortion region for log-loss distortion)**.** *A point $(\mu_{2^{\mathcal{J}},2^{\mathcal{L}}}, R_{\mathcal{J}})$ is log-loss achievable if for some $n \in \mathbb{N}$ there exists an $(n, R_{\mathcal{J}})$-code $f_{\mathcal{J}}$ yielding $\mathsf{W}_j \triangleq f_j(\mathbf{X}_j)$ and decoding functions $g_{\mathcal{A},\mathcal{B}} \colon \mathcal{M}_{\mathcal{A}} \to \mathcal{P}(\mathcal{Y}_{\mathcal{B}}^n)$ for all nonempty $\mathcal{A} \subseteq \mathcal{J}$ and $\mathcal{B} \subseteq \mathcal{L}$ such that*

$$\mathbb{E}[\zeta_{\mathcal{B}}(\mathbf{Y}_{\mathcal{B}}, g_{\mathcal{A},\mathcal{B}}(\mathsf{W}_{\mathcal{A}}))] \geq \mu_{\mathcal{A},\mathcal{B}}.$$

Let $\mathcal{R}_{LL}$ be the set of all log-loss achievable points.

To show the equivalence with the ITB problem, we state an auxiliary lemma similar to [8, Lemma 1].

Lemma 9. *For any decoding function $g_{\mathcal{A},\mathcal{B}}$ and code $f_{\mathcal{J}}$,*

$$\mathbb{E}[\zeta_{\mathcal{B}}(\mathbf{Y}_{\mathcal{B}}, g_{\mathcal{A},\mathcal{B}}(\mathsf{U}_{\mathcal{A}}))] \leq \frac{1}{n}\mathrm{I}(\mathbf{Y}_{\mathcal{B}};\mathsf{U}_{\mathcal{A}}),$$

with equality if and only if $g_{\mathcal{A},\mathcal{B}}(u_{\mathcal{A}}) = \mathrm{p}_{\mathbf{Y}_{\mathcal{B}}|\mathsf{U}_{\mathcal{A}}}(\,\cdot\,|u_{\mathcal{A}})$.

Proof: The proof follows with suitable modifications directly from the proof of [8, Lemma 1] . ∎

As a corollary of Lemma 9, we have equality between the two achievable regions.

Corollary 10 (Equivalence of MI and log-loss)**.** $\mathcal{R}_{LL} = \mathcal{R}_{MI}$.

Proof: Immediate from Lemma 9. ∎

Corollary 10 implies that the results in [8] directly apply to the CEO problem with mutual information constraint. For instance, the CEO problem with J encoders under logarithmic loss distortion [8, Appendix B] can be obtained in Definition 8 by setting $L = 1$ (i.e., the CEO is only interested in $\mathbf{Y}_1$) and $\mu_{\mathcal{A},1} = 0$, whenever $\mathcal{A} \neq \mathcal{J}$ (i.e., the CEO listens to all her agents). Then the region $\mathcal{R}_{\mathrm{MI}} = \mathcal{R}_{\mathrm{LL}}$ collapses to the set of all $(\mu, R_{\mathcal{J}})$ such that there exists an $(n, \mathcal{R}_{\mathcal{J}})$-code $f_{\mathcal{J}}$ with $\mathrm{I}(\mathsf{U}_{\mathcal{J}};\mathbf{Y}_1) \geq n\mu$, where $\mathsf{U}_j \triangleq f_j(\mathbf{X}_j)$. The resulting achievable region $\mathcal{R}^i_{\mathrm{MI}} \subseteq \mathcal{R}_{\mathrm{MI}}$, that can be obtained from $\mathcal{R}_{\mathrm{i}}$ is shown to be tight by [8, Lemma 5] if $\mathsf{X}_{\mathcal{J}}$ are mutually independent given Y_1. A similar argument can be made for the multiterminal source coding problem under logarithmic loss, introduced in [8, Section II], where the inner bound, obtained from $\mathcal{R}_{\mathrm{i}}$, is also tight.

IV. Proof of Theorem 5

This proof extends ideas in [5, Section VI] to more than two sources and incorporates a binning strategy.

Lemma 11 (Existence of a code)**.** *Let $\varepsilon > 0$, $\mathsf{U}_k \multimap \mathsf{X}_k \multimap (\mathsf{X}_{\mathcal{K}\setminus k}, \mathsf{U}_{\mathcal{K}\setminus k})$ for all $k \in \mathcal{K}$, and $R_{\mathcal{K}} \in \mathbb{R}_+^K$. Then, for some suitably small $\delta > 0$ and suitably large $n \in \mathbb{N}$, we can obtain an $(n, R_{\mathcal{K}} + \varepsilon)$ code $f_{\mathcal{K}}$ and decoding functions $g_{\mathcal{A}_a,\mathcal{A}_b} \colon \mathcal{M}_{\mathcal{A}_a} \to \mathcal{U}^n_{\mathcal{A}_b}$ for each pair of nonempty index sets $\mathcal{A}_b \subseteq \mathcal{A}_a \subseteq \mathcal{K}$, such that the following two properties hold:*

1) *Let $\mathcal{A}_a, \mathcal{A}_b, \mathcal{B}_a, \mathcal{B}_b \subseteq \mathcal{K}$ be arbitrary nonempty subsets of indices with $\mathcal{A}_b \subseteq \mathcal{A}_a$, $\mathcal{B}_b \subseteq \mathcal{B}_a$, and $\mathcal{A}_a \cap \mathcal{B}_a = \varnothing$. If (1) and (2) hold, then, using $\mathsf{W}_k \triangleq f_k(\mathbf{X}_k)$ and the abbreviations $\hat{\mathbf{U}}_{\mathcal{A}_b} \triangleq g_{\mathcal{A}_a,\mathcal{A}_b}(\mathsf{W}_{\mathcal{A}_a})$ and $\hat{\mathbf{U}}_{\mathcal{B}_b} \triangleq g_{\mathcal{B}_a,\mathcal{B}_b}(\mathsf{W}_{\mathcal{B}_a})$,*

$$\mathrm{P}\Big\{\big(\hat{\mathbf{U}}_{\mathcal{A}_b}, \mathbf{X}_{\mathcal{A}_a}, \mathbf{X}_{\mathcal{B}_a}, \hat{\mathbf{U}}_{\mathcal{B}_b}\big) \notin \mathcal{T}^n_{[\mathsf{U}_{\mathcal{A}_b}\mathsf{X}_{\mathcal{A}_a}\mathsf{X}_{\mathcal{B}_a}\mathsf{U}_{\mathcal{B}_b}]\delta}\Big\} \leq \varepsilon\ . \tag{4}$$

2) *For any pair $\mathcal{A}_a, \mathcal{B}_a \subset \mathcal{K}$ of disjoint nonempty index sets,*

$$\Big|\big[g_{\mathcal{A}_a,\mathcal{A}_b}(\mathcal{M}_{\mathcal{A}_a}) \times g_{\mathcal{B}_a,\mathcal{B}_b}(\mathcal{M}_{\mathcal{B}_a})\big] \cap \mathcal{T}^n_{[\mathsf{U}_{\mathcal{A}_b}\mathsf{U}_{\mathcal{B}_b}]\delta}\Big| \leq \exp\big(n(\mathrm{I}(\mathsf{U}_{\mathcal{A}_b}\mathsf{U}_{\mathcal{B}_b};\mathsf{X}_{\mathcal{A}_b}\mathsf{X}_{\mathcal{B}_b}) + \varepsilon)\big) \tag{5}$$

for any $\mathcal{A}_b \subseteq \mathcal{A}_a$ and $\mathcal{B}_b \subseteq \mathcal{B}_a$.

Proof: See Appendix A. ∎

We will further need the following set of random variables.

Definition 12. *For random variables $(\mathsf{U}_1, \mathsf{X}_1, \mathsf{X}_2, \mathsf{U}_2)$ and $\delta \geq 0$, define the set of random variables $\mathcal{L}_\delta(\mathsf{U}_1, \mathsf{X}_1, \mathsf{X}_2, \mathsf{U}_2)$, containing $(\mathsf{U}_1, \mathsf{X}_1, \mathsf{X}_2, \mathsf{U}_2)$ as*

$$\begin{aligned}\mathcal{L}_\delta(\mathsf{U}_1, \mathsf{X}_1, \mathsf{X}_2, \mathsf{U}_2) \triangleq \Big\{&(\widetilde{\mathsf{U}}_1, \widetilde{\mathsf{X}}_1, \widetilde{\mathsf{X}}_2, \widetilde{\mathsf{U}}_2) : \forall (u_1, x_1, x_2, u_2) :\\ &|\mathrm{p}_{\widetilde{\mathsf{U}}_1,\widetilde{\mathsf{X}}_1}(u_1, x_1) - \mathrm{p}_{\mathsf{U}_1\mathsf{X}_1}(u_1, x_1)| \leq \delta\ \mathrm{p}_{\mathsf{U}_1\mathsf{X}_1}(u_1, x_1),\\ &|\mathrm{p}_{\widetilde{\mathsf{U}}_2,\widetilde{\mathsf{X}}_2}(u_2, x_2) - \mathrm{p}_{\mathsf{U}_2\mathsf{X}_2}(u_2, x_2)| \leq \delta\ \mathrm{p}_{\mathsf{U}_2\mathsf{X}_2}(u_2, x_2),\\ &|\mathrm{p}_{\widetilde{\mathsf{U}}_1,\widetilde{\mathsf{U}}_2}(u_1, u_2) - \mathrm{p}_{\mathsf{U}_1\mathsf{U}_2}(u_1, u_2)| \leq \delta\ \mathrm{p}_{\mathsf{U}_1\mathsf{U}_2}(u_1, u_2)\Big\}\ .\end{aligned}$$

Note that $\mathcal{L}_\delta(\mathsf{U}_1, \mathsf{X}_1, \mathsf{X}_2, \mathsf{U}_2) \subseteq \mathcal{L}_{\delta'}(\mathsf{U}_1, \mathsf{X}_1, \mathsf{X}_2, \mathsf{U}_2)$ for $\delta \leq \delta'$. Let $(\mu_{2^\kappa,2^\kappa}, R_{\mathcal{K}}) \in \mathcal{R}_{\mathrm{i}}$ and choose $\mathsf{U}_{\mathcal{K}}$ as given in Definition 4. Fix $\varepsilon > 0$ and apply Lemma 11 to obtain encoding functions $f_{\mathcal{K}}$ and decoding functions $g_{\mathcal{A}_a,\mathcal{A}_b}$ for each pair of nonempty index sets $\mathcal{A}_b \subseteq \mathcal{A}_a \subseteq \mathcal{K}$. For an arbitrary pair of disjoint nonempty index sets $\mathcal{A}, \mathcal{B} \subset \mathcal{K}$, find the subsets $\mathcal{A}_b \subseteq \mathcal{A}_a \subseteq \mathcal{A}$ and $\mathcal{B}_b \subseteq \mathcal{B}_a \subseteq \mathcal{B}$, such that (1) to (3) hold. If either $\mathcal{A}_b = \varnothing$ or $\mathcal{B}_b = \varnothing$, then $\mu_{\mathcal{A},\mathcal{B}} \leq 0$, which is achieved by any code. We thus assume $\mathcal{A}_b, \mathcal{B}_b \neq \varnothing$. Define the functions $h_1 \triangleq g_{\mathcal{A}_a,\mathcal{A}_b} \circ f_{\mathcal{A}_a}$ and $h_2 \triangleq g_{\mathcal{B}_a,\mathcal{B}_b} \circ f_{\mathcal{B}_a}$. We want to analyze $\Theta(f_{\mathcal{A}}; f_{\mathcal{B}})$. To this end, define $\mathcal{D}_1 \triangleq h_1(\mathcal{X}^n_{\mathcal{A}_a})$. Naturally we can partition $\mathcal{X}^n_{\mathcal{A}_a}$ as $\mathcal{X}^n_{\mathcal{A}_a} = \bigcup_{\boldsymbol{u}_{\mathcal{A}_b} \in \mathcal{D}_1} h_1^{-1}(\boldsymbol{u}_{\mathcal{A}_b})$. We may assume without loss of generality that $h_1^{-1}(\boldsymbol{u}_{\mathcal{A}_b}) \subseteq \mathcal{T}^n_{[\mathsf{X}_{\mathcal{A}_a}|\mathsf{U}_{\mathcal{A}_b}]\delta}(\boldsymbol{u}_{\mathcal{A}_b})$ whenever $\boldsymbol{u}_{\mathcal{A}_b} \in \mathcal{D}_1 \cap \mathcal{T}^n_{[\mathsf{U}_{\mathcal{A}_b}]\delta}$ as this does not interfere with the properties of the code. Defining $\mathcal{D}_2$ in the same manner, we set $\mathcal{F} \triangleq (\mathcal{D}_1 \times \mathcal{D}_2) \cap \mathcal{T}^n_{[\mathsf{U}_{\mathcal{A}_b}\mathsf{U}_{\mathcal{B}_b}]\delta}$. Using $\hat{\mathbf{U}}_1 \triangleq h_1(\mathbf{X}_{\mathcal{A}_a})$ and $\hat{\mathbf{U}}_2 \triangleq h_2(\mathbf{X}_{\mathcal{B}_a})$ we have

$$\begin{aligned}
&n \cdot \Theta(f_{\mathcal{A}}(\mathbf{X}_{\mathcal{A}}); f_{\mathcal{B}}(\mathbf{X}_{\mathcal{B}}))\\
&\overset{(a)}{\geq} n \cdot \Theta(f_{\mathcal{A}_a}(\mathbf{X}_{\mathcal{A}_a}); f_{\mathcal{B}_a}(\mathbf{X}_{\mathcal{B}_a})) \overset{(b)}{\geq} n \cdot \Theta(\hat{\mathbf{U}}_1; \hat{\mathbf{U}}_2)\\
&\overset{(c)}{\geq} \mathrm{P}\{(\hat{\mathbf{U}}_1, \hat{\mathbf{U}}_2) \in \mathcal{F}\} \log \frac{\mathrm{P}\{(\hat{\mathbf{U}}_1, \hat{\mathbf{U}}_2) \in \mathcal{F}\}}{\mathrm{P}\{(\overline{\mathbf{U}}_1, \overline{\mathbf{U}}_2) \in \mathcal{F}\}}\\
&\qquad + \mathrm{P}\{(\hat{\mathbf{U}}_1, \hat{\mathbf{U}}_2) \in \mathcal{F}^{\mathrm{c}}\} \log \frac{\mathrm{P}\{(\hat{\mathbf{U}}_1, \hat{\mathbf{U}}_2) \in \mathcal{F}^{\mathrm{c}}\}}{\mathrm{P}\{(\overline{\mathbf{U}}_1, \overline{\mathbf{U}}_2) \in \mathcal{F}^{\mathrm{c}}\}}\\
&\geq -\mathrm{h}_0(\mathrm{P}\{(\hat{\mathbf{U}}_1, \hat{\mathbf{U}}_2) \in \mathcal{F}\})\\
&\qquad - \mathrm{P}\{(\hat{\mathbf{U}}_1, \hat{\mathbf{U}}_2) \in \mathcal{F}\} \log \mathrm{P}\{(\overline{\mathbf{U}}_1, \overline{\mathbf{U}}_2) \in \mathcal{F}\}\\
&\overset{(d)}{\geq} -\log(2) - (1-\varepsilon) \log \mathrm{P}\{(\overline{\mathbf{U}}_1, \overline{\mathbf{U}}_2) \in \mathcal{F}\}\ ,
\end{aligned} \tag{6}$$

where (a) and (b) follow from the data processing inequality [9, Theorem 2.8.1], (c) is a consequence of the log-sum inequality [9, Theorem 2.7.1], and (d) follows from (4). We further used $\overline{\mathbf{U}}_1 \triangleq h_1(\overline{\mathbf{X}}_{\mathcal{A}_a})$, $\overline{\mathbf{U}}_2 \triangleq h_2(\overline{\mathbf{X}}_{\mathcal{B}_a})$, where $(\overline{\mathbf{X}}_{\mathcal{A}_a}, \overline{\mathbf{X}}_{\mathcal{B}_a})$ are i.i.d. copies of $(\overline{\mathsf{X}}_{\mathcal{A}_a}, \overline{\mathsf{X}}_{\mathcal{B}_a}) \sim \mathrm{p}_{\mathsf{X}_{\mathcal{A}_a}}\mathrm{p}_{\mathsf{X}_{\mathcal{B}_a}}$.

For each $\boldsymbol{u}_{\mathcal{A}_b} \in \mathcal{D}_1$ and $\boldsymbol{u}_{\mathcal{B}_b} \in \mathcal{D}_2$, define:

$$\mathcal{S}(\boldsymbol{u}_{\mathcal{A}_b}, \boldsymbol{u}_{\mathcal{B}_b}) \triangleq \{\boldsymbol{u}_{\mathcal{A}_b}\} \times h_1^{-1}(\boldsymbol{u}_{\mathcal{A}_b}) \times h_2^{-1}(\boldsymbol{u}_{\mathcal{B}_b}) \times \{\boldsymbol{u}_{\mathcal{B}_b}\}$$

and

$$\mathfrak{S} \triangleq \bigcup_{(\boldsymbol{u}_{\mathcal{A}_b}, \boldsymbol{u}_{\mathcal{B}_b}) \in \mathcal{F}} \mathcal{S}(\boldsymbol{u}_{\mathcal{A}_b}, \boldsymbol{u}_{\mathcal{B}_b})\ .$$

Let $(\hat{\mathsf{U}}_{\mathcal{A}_b}, \hat{\mathsf{X}}_{\mathcal{A}_a}, \hat{\mathsf{X}}_{\mathcal{B}_a}, \hat{\mathsf{U}}_{\mathcal{B}_b})$ be the type variables corresponding to some fixed $(\hat{\boldsymbol{u}}_{\mathcal{A}_b}, \hat{\boldsymbol{x}}_{\mathcal{A}_a}, \hat{\boldsymbol{x}}_{\mathcal{B}_a}, \hat{\boldsymbol{u}}_{\mathcal{B}_b}) \in \mathfrak{S}$. From [11, Lemma 2.6], we know

$$\begin{aligned}
&\mathrm{P}\big\{\overline{\mathbf{X}}_{\mathcal{A}_a} = \hat{\boldsymbol{x}}_{\mathcal{A}_a}, \overline{\mathbf{X}}_{\mathcal{B}_a} = \hat{\boldsymbol{x}}_{\mathcal{B}_a}\big\}\\
&\quad = \exp\Big[-n\Big(\mathrm{H}(\hat{\mathsf{X}}_{\mathcal{A}_a}\hat{\mathsf{X}}_{\mathcal{B}_a}) + \mathrm{D}_{\mathrm{KL}}(\hat{\mathsf{X}}_{\mathcal{A}_a}\hat{\mathsf{X}}_{\mathcal{B}_a} \| \overline{\mathsf{X}}_{\mathcal{A}_a}\overline{\mathsf{X}}_{\mathcal{B}_a})\Big)\Big]\ .
\end{aligned} \tag{7}$$

Letting $K(\boldsymbol{u}_{\mathcal{A}_b}, \boldsymbol{u}_{\mathcal{B}_b})$ denote the number of elements in $\mathcal{S}(\boldsymbol{u}_{\mathcal{A}_b}, \boldsymbol{u}_{\mathcal{B}_b})$ with type $(\hat{\mathsf{U}}_{\mathcal{A}_b}, \hat{\mathsf{X}}_{\mathcal{A}_a}, \hat{\mathsf{X}}_{\mathcal{B}_a}, \hat{\mathsf{U}}_{\mathcal{B}_b})$,

$$K(\boldsymbol{u}_{\mathcal{A}_b}, \boldsymbol{u}_{\mathcal{B}_b}) \leq \exp\Big(n\mathrm{H}(\hat{\mathsf{X}}_{\mathcal{A}_a}\hat{\mathsf{X}}_{\mathcal{B}_a}|\hat{\mathsf{U}}_{\mathcal{A}_b}\hat{\mathsf{U}}_{\mathcal{B}_b})\Big) \tag{8}$$

by [11, Lemma 2.5]. Letting $K(\hat{\mathsf{U}}_{\mathcal{A}_b}, \hat{\mathsf{X}}_{\mathcal{A}_a}, \hat{\mathsf{X}}_{\mathcal{B}_a}, \hat{\mathsf{U}}_{\mathcal{B}_b})$ be the

number of elements of $\mathfrak{S}$ with type $(\hat{\mathsf{U}}_{\mathcal{A}_b}, \hat{\mathsf{X}}_{\mathcal{A}_a}, \hat{\mathsf{X}}_{\mathcal{B}_a}, \hat{\mathsf{U}}_{\mathcal{B}_b})$,

$$\begin{aligned} K(\hat{\mathsf{U}}_{\mathcal{A}_b}, \hat{\mathsf{X}}_{\mathcal{A}_a}, \hat{\mathsf{X}}_{\mathcal{B}_a}, \hat{\mathsf{U}}_{\mathcal{B}_b}) &= \sum_{(\boldsymbol{u}_{\mathcal{A}_b}, \boldsymbol{u}_{\mathcal{B}_b}) \in \mathcal{F}} K(\boldsymbol{u}_{\mathcal{A}_b}, \boldsymbol{u}_{\mathcal{B}_b}) \\ &\overset{(e)}{\leq} \sum_{(\boldsymbol{u}_{\mathcal{A}_b}, \boldsymbol{u}_{\mathcal{B}_b}) \in \mathcal{F}} \exp\Big(n\mathrm{H}(\hat{\mathsf{X}}_{\mathcal{A}_a}\hat{\mathsf{X}}_{\mathcal{B}_a}|\hat{\mathsf{U}}_{\mathcal{A}_b}\hat{\mathsf{U}}_{\mathcal{B}_b})\Big) \\ &\overset{(f)}{\leq} \exp\Big[n\Big(\mathrm{I}(\mathsf{U}_{\mathcal{A}_b}\mathsf{U}_{\mathcal{B}_b}; \mathsf{X}_{\mathcal{A}_a}\mathsf{X}_{\mathcal{B}_a}) \\ &\qquad + \mathrm{H}(\hat{\mathsf{X}}_{\mathcal{A}_a}\hat{\mathsf{X}}_{\mathcal{B}_a}|\hat{\mathsf{U}}_{\mathcal{A}_b}\hat{\mathsf{U}}_{\mathcal{B}_b}) + \varepsilon\Big)\Big] \ , \end{aligned} \tag{9}$$

where (e) follows from (8) and (f) from (5). Thus,

$$\begin{aligned} &\mathrm{P}\big\{(\overline{\mathbf{U}}_{\mathcal{A}_b}, \overline{\mathbf{U}}_{\mathcal{B}_b}) \in \mathcal{F}\big\} \\ &\overset{(g)}{=} \sum_{\hat{\mathsf{U}}_{\mathcal{A}_b}, \hat{\mathsf{X}}_{\mathcal{A}_a}, \hat{\mathsf{X}}_{\mathcal{B}_a}, \hat{\mathsf{U}}_{\mathcal{B}_b}} K(\hat{\mathsf{U}}_{\mathcal{A}_b}, \hat{\mathsf{X}}_{\mathcal{A}_a}, \hat{\mathsf{X}}_{\mathcal{B}_a}, \hat{\mathsf{U}}_{\mathcal{B}_b}) \\ &\qquad \cdot \exp\Big[-n\Big(\mathrm{H}(\hat{\mathsf{X}}_{\mathcal{A}_a}\hat{\mathsf{X}}_{\mathcal{B}_a}) + \mathrm{D}_{\mathrm{KL}}(\hat{\mathsf{X}}_{\mathcal{A}_a}\hat{\mathsf{X}}_{\mathcal{B}_a}\|\overline{\mathsf{X}}_{\mathcal{A}_a}\overline{\mathsf{X}}_{\mathcal{B}_a})\Big)\Big] \\ &\overset{(h)}{\leq} \sum_{\hat{\mathsf{U}}_{\mathcal{A}_b}, \hat{\mathsf{X}}_{\mathcal{A}_a}, \hat{\mathsf{X}}_{\mathcal{B}_a}, \hat{\mathsf{U}}_{\mathcal{B}_b}} \exp\Big(-n\big(k(\hat{\mathsf{U}}_{\mathcal{A}_b}, \hat{\mathsf{X}}_{\mathcal{A}_a}, \hat{\mathsf{X}}_{\mathcal{B}_a}, \hat{\mathsf{U}}_{\mathcal{B}_b}) - \varepsilon\big)\Big) \ , \end{aligned}$$

where the sum is over all types that occur in $\mathfrak{S}$. Here, (g) follows from (7), and (h) from (9), and we defined

$$\begin{aligned} k(\hat{\mathsf{U}}_{\mathcal{A}_b}, \hat{\mathsf{X}}_{\mathcal{A}_a}, \hat{\mathsf{X}}_{\mathcal{B}_a}, \hat{\mathsf{U}}_{\mathcal{B}_b}) &\triangleq \mathrm{I}(\hat{\mathsf{U}}_{\mathcal{A}_b}\hat{\mathsf{U}}_{\mathcal{B}_b}; \hat{\mathsf{X}}_{\mathcal{A}_a}\hat{\mathsf{X}}_{\mathcal{B}_a}) \\ &\quad - \mathrm{I}(\mathsf{U}_{\mathcal{A}_b}\mathsf{U}_{\mathcal{B}_b}; \mathsf{X}_{\mathcal{A}_a}\mathsf{X}_{\mathcal{B}_a}) + \mathrm{D}_{\mathrm{KL}}(\hat{\mathsf{X}}_{\mathcal{A}_a}\hat{\mathsf{X}}_{\mathcal{B}_a}\|\overline{\mathsf{X}}_{\mathcal{A}_a}\overline{\mathsf{X}}_{\mathcal{B}_a}) \ . \end{aligned}$$

Using a type counting argument [11, Lemma 2.2], we can further bound the number of types that occur in $\mathfrak{S}$ by

$$(n+1)^{|\mathcal{U}_{\mathcal{A}_b}||\mathcal{X}_{\mathcal{A}_a}||\mathcal{X}_{\mathcal{B}_a}||\mathcal{U}_{\mathcal{B}_b}|} \leq (n+1)^{|\mathcal{U}_{\mathcal{K}}||\mathcal{X}_{\mathcal{K}}|}$$

and obtain

$$\begin{aligned} &\mathrm{P}\big\{(\overline{\mathbf{U}}_{\mathcal{A}_b}, \overline{\mathbf{U}}_{\mathcal{B}_b}) \in \mathcal{F}\big\} \leq (n+1)^{|\mathcal{U}_{\mathcal{K}}||\mathcal{X}_{\mathcal{K}}|} \\ &\times \max_{\hat{\mathsf{U}}_{\mathcal{A}_b}, \hat{\mathsf{X}}_{\mathcal{A}_a}, \hat{\mathsf{X}}_{\mathcal{B}_a}, \hat{\mathsf{U}}_{\mathcal{B}_b}} \exp\Big[-n\Big(k(\hat{\mathsf{U}}_{\mathcal{A}_b}, \hat{\mathsf{X}}_{\mathcal{A}_a}, \hat{\mathsf{X}}_{\mathcal{B}_a}, \hat{\mathsf{U}}_{\mathcal{B}_b}) - \varepsilon\Big)\Big], \end{aligned} \tag{10}$$

where the maximum is over all types $(\hat{\mathsf{U}}_{\mathcal{A}_b}, \hat{\mathsf{X}}_{\mathcal{A}_a}, \hat{\mathsf{X}}_{\mathcal{B}_a}, \hat{\mathsf{U}}_{\mathcal{B}_b})$ occurring in $\mathfrak{S}$. For any such type, we have by construction $(\hat{\mathsf{U}}_{\mathcal{A}_b}, \hat{\mathsf{X}}_{\mathcal{A}_a}, \hat{\mathsf{X}}_{\mathcal{B}_a}, \hat{\mathsf{U}}_{\mathcal{B}_b}) \in \mathcal{L}_\delta(\mathsf{U}_{\mathcal{A}_b}, \mathsf{X}_{\mathcal{A}_a}, \mathsf{X}_{\mathcal{B}_a}, \mathsf{U}_{\mathcal{B}_b})$. From (10), we can thus conclude:

$$\begin{aligned} &\mathrm{P}\big\{(\overline{\mathbf{U}}_{\mathcal{A}_b}, \overline{\mathbf{U}}_{\mathcal{B}_b}) \in \mathcal{F}\big\} \leq (n+1)^{|\mathcal{U}_{\mathcal{K}}||\mathcal{X}_{\mathcal{K}}|} \\ &\quad \times \max_{\mathcal{L}_\delta} \exp\Big[-n\Big(k(\widetilde{\mathsf{U}}_{\mathcal{A}_b}, \widetilde{\mathsf{X}}_{\mathcal{A}_a}, \widetilde{\mathsf{X}}_{\mathcal{B}_a}, \widetilde{\mathsf{U}}_{\mathcal{B}_b}) - \varepsilon\Big)\Big] \ , \end{aligned} \tag{11}$$

where we used $\max_{\mathcal{L}_\delta}$ to denote maximization over all $(\widetilde{\mathsf{U}}_{\mathcal{A}_b}, \widetilde{\mathsf{X}}_{\mathcal{A}_a}, \widetilde{\mathsf{X}}_{\mathcal{B}_a}, \widetilde{\mathsf{U}}_{\mathcal{B}_b}) \in \mathcal{L}_\delta(\mathsf{U}_{\mathcal{A}_b}, \mathsf{X}_{\mathcal{A}_a}, \mathsf{X}_{\mathcal{B}_a}, \mathsf{U}_{\mathcal{B}_b})$. Combining (6) and (11), we showed that for n large enough and some constant C

$$\begin{aligned} \Theta(f_{\mathcal{A}}; f_{\mathcal{B}}) &\geq -\frac{\log(2)}{n} - \frac{1-\varepsilon}{n}\log \mathrm{P}\big\{(\overline{\mathbf{U}}_{\mathcal{A}_b}, \overline{\mathbf{U}}_{\mathcal{B}_b}) \in \mathcal{F}\big\} \\ &\geq -\varepsilon + (1-\varepsilon)\min_{\mathcal{L}_\delta}\Big(k(\widetilde{\mathsf{U}}_{\mathcal{A}_b}, \widetilde{\mathsf{X}}_{\mathcal{A}_a}, \widetilde{\mathsf{X}}_{\mathcal{B}_a}, \widetilde{\mathsf{U}}_{\mathcal{B}_b}) - \varepsilon\Big) \\ &\geq -2\varepsilon + (1-\varepsilon)\min_{\mathcal{L}_\delta}\Big(k(\widetilde{\mathsf{U}}_{\mathcal{A}_b}, \widetilde{\mathsf{X}}_{\mathcal{A}_a}, \widetilde{\mathsf{X}}_{\mathcal{B}_a}, \widetilde{\mathsf{U}}_{\mathcal{B}_b})\Big) \\ &\geq \min_{\mathcal{L}_\delta}\Big(k(\widetilde{\mathsf{U}}_{\mathcal{A}_b}, \widetilde{\mathsf{X}}_{\mathcal{A}_a}, \widetilde{\mathsf{X}}_{\mathcal{B}_a}, \widetilde{\mathsf{U}}_{\mathcal{B}_b})\Big) - C\varepsilon \ , \end{aligned} \tag{12}$$

where $\min_{\mathcal{L}_\delta}$ is used to denote minimization over all $(\widetilde{\mathsf{U}}_{\mathcal{A}_b}, \widetilde{\mathsf{X}}_{\mathcal{A}_a}, \widetilde{\mathsf{X}}_{\mathcal{B}_a}, \widetilde{\mathsf{U}}_{\mathcal{B}_b}) \in \mathcal{L}_\delta(\mathsf{U}_{\mathcal{A}_b}, \mathsf{X}_{\mathcal{A}_a}, \mathsf{X}_{\mathcal{B}_a}, \mathsf{U}_{\mathcal{B}_b})$. Observing that $k(\widetilde{\mathsf{U}}_{\mathcal{A}_b}, \widetilde{\mathsf{X}}_{\mathcal{A}_a}, \widetilde{\mathsf{X}}_{\mathcal{B}_a}, \widetilde{\mathsf{U}}_{\mathcal{B}_b})$ is a continuous function of $\mathrm{p}_{\widetilde{\mathsf{U}}_{\mathcal{A}_b}, \widetilde{\mathsf{X}}_{\mathcal{A}_a}, \widetilde{\mathsf{X}}_{\mathcal{B}_a}, \widetilde{\mathsf{U}}_{\mathcal{B}_b}}$ and (12) holds for arbitrarily small δ, we obtain for n large enough,

$$\Theta(f_{\mathcal{A}}; f_{\mathcal{B}}) \geq \min_{\mathcal{L}_0} k(\widetilde{\mathsf{U}}_{\mathcal{A}_b}, \widetilde{\mathsf{X}}_{\mathcal{A}_a}, \widetilde{\mathsf{X}}_{\mathcal{B}_a}, \widetilde{\mathsf{U}}_{\mathcal{B}_b}) - C'\varepsilon \tag{13}$$

for some (larger) constant C' by letting $\delta \to 0$. For $(\widetilde{\mathsf{U}}_{\mathcal{A}_b}, \widetilde{\mathsf{X}}_{\mathcal{A}_a}, \widetilde{\mathsf{X}}_{\mathcal{B}_a}, \widetilde{\mathsf{U}}_{\mathcal{B}_b}) \in \mathcal{L}_0(\mathsf{U}_{\mathcal{A}_b}, \mathsf{X}_{\mathcal{A}_a}, \mathsf{X}_{\mathcal{B}_a}, \mathsf{U}_{\mathcal{B}_b})$, observe that we have

$$k(\widetilde{\mathsf{U}}_{\mathcal{A}_b}, \widetilde{\mathsf{X}}_{\mathcal{A}_a}, \widetilde{\mathsf{X}}_{\mathcal{B}_a}, \widetilde{\mathsf{U}}_{\mathcal{B}_b}) = \mathrm{I}(\widetilde{\mathsf{U}}_{\mathcal{A}_b}\widetilde{\mathsf{X}}_{\mathcal{A}_a}; \widetilde{\mathsf{X}}_{\mathcal{B}_a}\widetilde{\mathsf{U}}_{\mathcal{B}_b}) \ . \tag{14}$$

and from the the definition of $\mathcal{L}_0(\mathsf{U}_{\mathcal{A}_b}, \mathsf{X}_{\mathcal{A}_a}, \mathsf{X}_{\mathcal{B}_a}, \mathsf{U}_{\mathcal{B}_b})$ clearly

$$\min_{\mathcal{L}_0} \mathrm{I}(\widetilde{\mathsf{U}}_{\mathcal{A}_b}\widetilde{\mathsf{X}}_{\mathcal{A}_a}; \widetilde{\mathsf{X}}_{\mathcal{B}_a}\widetilde{\mathsf{U}}_{\mathcal{B}_b}) \geq \mathrm{I}(\mathsf{U}_{\mathcal{A}_b}; \mathsf{U}_{\mathcal{B}_b}). \tag{15}$$

Combining (13), (14), and (15), we obtain $\Theta(f_{\mathcal{A}}; f_{\mathcal{B}}) \geq \mathrm{I}(\mathsf{U}_{\mathcal{A}_b}; \mathsf{U}_{\mathcal{B}_b}) - C'\varepsilon \geq \mu_{\mathcal{A},\mathcal{B}} - C'\varepsilon$ where the last inequality follows from (3). Thus, $(\mu_{2^{\mathcal{A}},2^{\mathcal{B}}} - C'\varepsilon, R_{\mathcal{K}} + \varepsilon) \in \mathcal{R}$, completing the proof since ε was arbitrary.

V. Summary and Discussion

We extended the information-theoretic biclustering problem to multiple sources and derived bounds on the achievable region. However, these bounds are not tight since the infamous Körner-Marton problem constitutes a counterexample. For the analog of the CEO problem we showed that our outer bound is tight in a special case, leveraging existing results from multiterminal lossy source coding. The challenge of the biclustering problem lies in the fact that one needs to bound the mutual information between two arbitrary encodings solely based on their rates. This appears to be a difficult task since standard information-theoretic manipulations seem incapable of handling this dependence well.

Appendix

A. Proof of Lemma 11

Fix $0 < \varepsilon', \varepsilon'' < \varepsilon$ and $\widetilde{R}_k = \mathrm{I}(\mathsf{X}_k; \mathsf{U}_k) + \varepsilon''/2$ for $k \in \mathcal{K}$.

Encoding: For $n \in \mathbb{N}$ define $\widetilde{M}_k \triangleq \mathrm{e}^{n\widetilde{R}_k}$ and $\widetilde{\mathcal{M}}_k \triangleq \{1, 2, \ldots, \widetilde{M}_k\}$. We apply [13, Lemma 3.4] and denote the random codebook $\mathcal{C}_k \triangleq (\mathbf{V}_i^{(k)})_{i \in \widetilde{\mathcal{M}}_k}$, which are drawn independently uniform from $\mathcal{T}^n_{[\mathsf{U}_k]\delta}$ for each $k \in \mathcal{K}$. Denote the resulting randomized coding functions as $\widetilde{\mathsf{W}}_k = \widetilde{f}_k(\mathbf{X}_k, \mathcal{C}_k)$ and the corresponding decoded value as $\widetilde{\mathbf{U}}_k \triangleq \mathbf{V}^{(k)}_{\widetilde{\mathsf{W}}_k}$. We have therefore, if n is chosen large enough and δ small enough,

$$P_e \triangleq \mathrm{P}\Big\{(\widetilde{\mathbf{U}}_{\mathcal{K}}, \mathbf{X}_{\mathcal{K}}) \notin \mathcal{T}^n_{[\mathsf{U}_{\mathcal{K}}\mathsf{X}_{\mathcal{K}}]\delta}\Big\} \leq \varepsilon' \ . \tag{16}$$

Next, we introduce (deterministic) binning. If $R_k < \mathrm{I}(\mathsf{X}_k; \mathsf{U}_k)$, partition $\widetilde{\mathcal{M}}_k$ into $M_k \triangleq \mathrm{e}^{n(R_k + \varepsilon'')}$ equally sized, consecutive bins, each of size $\mathrm{e}^{n\Delta_k}$ with $\Delta_k \triangleq \widetilde{R}_k - R_k - \varepsilon'' = \mathrm{I}(\mathsf{X}_k; \mathsf{U}_k) - R_k - \frac{\varepsilon''}{2}$. The deterministic function $\beta_k \colon \widetilde{\mathcal{M}}_k \to \mathcal{M}_k$ maps a codeword onto the bin index in $\mathcal{M}_k \triangleq \{1, 2, \ldots, M_k\}$, it belongs to. Now use the randomized encoding function $f_k \triangleq \beta_k \circ \widetilde{f}_k$. If $R_k \geq \mathrm{I}(\mathsf{X}_k; \mathsf{U}_k)$, let β_k be the identity on $\widetilde{\mathcal{M}}_k$ and $f_k \triangleq \widetilde{f}_k$.

Decoding: Given the codebooks, the decoding procedure $g_{\mathcal{A}_a,\mathcal{A}_b} \colon \mathcal{M}_{\mathcal{A}_a} \to \mathcal{U}^n_{\mathcal{A}_b}$ for each $\varnothing \neq \mathcal{A}_b \subseteq \mathcal{A}_a \subseteq \mathcal{K}$ is carried out as follows. Given $m_{\mathcal{A}_a} \in \mathcal{M}_{\mathcal{A}_a}$, let $\widetilde{\mathfrak{m}}_{\mathcal{A}_a} \triangleq \beta^{-1}_{\mathcal{A}_a}(m_{\mathcal{A}_a}) \subseteq \widetilde{\mathcal{M}}_{\mathcal{A}_a}$ be all indices that are in the bins $m_{\mathcal{A}_a}$. Consider only the typical sequences $\mathbf{V}^{(\mathcal{A}_a)}_{\widetilde{\mathfrak{m}}_{\mathcal{A}_a}} \cap \mathcal{T}^n_{[\mathsf{U}_{\mathcal{A}_a}]\delta} \triangleq \Phi \subseteq \mathcal{U}^n_{\mathcal{A}_a}$. Denote the restriction of Φ to the coordinates $\mathcal{A}_b$ as $[\Phi]_{\mathcal{A}_b}$. If $\Phi \neq \varnothing$, choose the lexicographically smallest element of $[\Phi]_{\mathcal{A}_b}$, otherwise choose the choose the lexicographically smallest element of $\Big[\mathbf{V}^{(\mathcal{A}_a)}_{\widetilde{\mathfrak{m}}_{\mathcal{A}_a}}\Big]_{\mathcal{A}_b}$.

Let $\mathcal{A}_a, \mathcal{A}_b, , \mathcal{B}_a, \mathcal{B}_b \subset \mathcal{K}$ be sets of indices, such that the conditions of part 1 are satisfied. Using $\mathsf{W}_k \triangleq f_k(\mathbf{X}_k, \mathcal{C}_k)$ and the randomized decodings $\hat{\mathbf{U}}_1 \triangleq g_{\mathcal{A}_a,\mathcal{A}_b}(\mathsf{W}_{\mathcal{A}_a}, \mathcal{C}_{\mathcal{A}_a})$ and $\hat{\mathbf{U}}_2 \triangleq g_{\mathcal{B}_a,\mathcal{B}_b}(\mathsf{W}_{\mathcal{B}_a}, \mathcal{C}_{\mathcal{B}_a})$, consider the error event $\mathcal{E}_0 \triangleq \{(\hat{\mathbf{U}}_1, \mathbf{X}_{\mathcal{A}_a}, \mathbf{X}_{\mathcal{B}_a}, \hat{\mathbf{U}}_2) \notin \mathcal{T}^n_{[\mathsf{U}_{\mathcal{A}_b}\mathsf{X}_{\mathcal{A}_a}\mathsf{X}_{\mathcal{B}_a}\mathsf{U}_{\mathcal{B}_b}]\delta}\}$. Define

$$\mathcal{E}_1 \triangleq \{(\widetilde{\mathbf{U}}_{\mathcal{A}_a}, \mathbf{X}_{\mathcal{A}_a}, \mathbf{X}_{\mathcal{B}_a}, \widetilde{\mathbf{U}}_{\mathcal{B}_a}) \notin \mathcal{T}^n_{[\mathsf{U}_{\mathcal{A}_b}\mathsf{X}_{\mathcal{A}_a}\mathsf{X}_{\mathcal{B}_a}\mathsf{U}_{\mathcal{B}_b}]\delta}\},$$
$$\mathcal{E}_2 \triangleq \left\{\left|\left[\mathbf{V}^{(\mathcal{A}_a)}_{\widetilde{\mathfrak{W}}_{\mathcal{A}_a}} \cap \mathcal{T}^n_{[\mathsf{U}_{\mathcal{A}_a}]\delta}\right]_{\mathcal{A}_b}\right| > 1\right\},$$
$$\mathcal{E}_3 \triangleq \left\{\left|\left[\mathbf{V}^{(\mathcal{B}_a)}_{\widetilde{\mathfrak{W}}_{\mathcal{B}_a}} \cap \mathcal{T}^n_{[\mathsf{U}_{\mathcal{B}_a}]\delta}\right]_{\mathcal{B}_b}\right| > 1\right\},$$

with the random index set $\widetilde{\mathfrak{W}}_{\mathcal{A}} \triangleq \beta^{-1}_{\mathcal{A}}(\mathsf{W}_{\mathcal{A}})$. We clearly have $\mathcal{E}_0 \subseteq \mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3$ and thus

$$\begin{aligned}\mathrm{P}\{\mathcal{E}_0\} &\leq \mathrm{P}\{\mathcal{E}_1\} + \mathrm{P}\{\mathcal{E}_2|\mathcal{E}_1^{\mathrm{c}}\} + \mathrm{P}\{\mathcal{E}_3|\mathcal{E}_1^{\mathrm{c}}\} \\ &\overset{(j)}{\leq} \mathrm{P}\{\mathcal{E}_2|\mathcal{E}_1^{\mathrm{c}}\} + \mathrm{P}\{\mathcal{E}_3|\mathcal{E}_1^{\mathrm{c}}\} + \varepsilon' \ , \end{aligned} \tag{17}$$

where (j) follows from (16). We can partition $\widetilde{\mathfrak{W}}_{\mathcal{A}_a}$ into (random) subsets $\mathcal{D}_{\mathcal{A}'}$, indexed by $\mathcal{A}' \subseteq \mathcal{A}_a$ as

$$\mathcal{D}_{\mathcal{A}'} \triangleq \{\widetilde{w}_{\mathcal{A}_a} \in \widetilde{\mathfrak{W}}_{\mathcal{A}_a} : \widetilde{w}_{\mathcal{A}'^{\mathrm{c}}} = \widetilde{\mathsf{W}}_{\mathcal{A}'^{\mathrm{c}}} \wedge \widetilde{w}_k \neq \widetilde{\mathsf{W}}_k, \forall k \in \mathcal{A}'\} \ ,$$

where we used $\mathcal{A}'^{\mathrm{c}} \triangleq \mathcal{A}_a \setminus \mathcal{A}'$. Observe that $\mathcal{D}_{\varnothing} = \{\widetilde{\mathsf{W}}_{\mathcal{A}_a}\}$. For each set $\varnothing \neq \mathcal{A}' \subseteq \mathcal{A}_a$, we define the error event

$$\mathcal{E}_{\mathcal{A}'} \triangleq \{\mathbf{V}^{(\mathcal{A}_a)}_{\mathcal{D}_{\mathcal{A}'}} \cap \mathcal{T}^n_{[\mathsf{U}_{\mathcal{A}_a}]\delta} \neq \varnothing\}$$

and we have

$$\mathrm{P}\{\mathcal{E}_2|\mathcal{E}_1^{\mathrm{c}}\} \leq \sum_{\mathcal{A}' \subseteq \mathcal{A}_a : \mathcal{A}' \cap \mathcal{A}_b \neq \varnothing} \mathrm{P}\{\mathcal{E}_{\mathcal{A}'}|\mathcal{E}_1^{\mathrm{c}}\} \tag{18}$$

since $\mathcal{E}_2 \subseteq \bigcup_{\mathcal{A}' \subseteq \mathcal{A}_a : \mathcal{A}' \cap \mathcal{A}_b \neq \varnothing} \mathcal{E}_{\mathcal{A}'}$. By the construction of the codebook, $\mathcal{D}_{\mathcal{A}'}$ has $\prod_{k \in \mathcal{A}'}(\exp[n\Delta_k] - 1)$ elements. For $\widetilde{w}_{\mathcal{A}_a} \in \mathcal{D}_{\mathcal{A}'}$, the components $\mathbf{V}^{(\mathcal{A}')}_{\widetilde{w}_{\mathcal{A}'}}$ are uniformly distributed on $\prod_{k \in \mathcal{A}'} \mathcal{T}^n_{[\mathsf{U}_k]\delta}$ and $\widetilde{w}_{\mathcal{A}'^{\mathrm{c}}} = \widetilde{\mathsf{W}}_{\mathcal{A}'^{\mathrm{c}}}$. Given $\mathcal{E}_1^{\mathrm{c}}$ we have in particular $\widetilde{\mathbf{U}}_{\mathcal{A}_a} \in \mathcal{T}^n_{[\mathsf{U}_{\mathcal{A}_a}]\delta}$. Thus, for any $\boldsymbol{u}_{\mathcal{A}'^{\mathrm{c}}} \in \mathcal{T}^n_{[\mathsf{U}_{\mathcal{A}'^{\mathrm{c}}}]\delta}$, we can conclude,

$$\begin{aligned}&\mathrm{P}\Big\{\mathcal{E}_{\mathcal{A}'}\Big|\mathcal{E}_1^{\mathrm{c}}, \widetilde{\mathbf{U}}_{\mathcal{A}'^{\mathrm{c}}} = \boldsymbol{u}_{\mathcal{A}'^{\mathrm{c}}}\Big\} \\ &\leq \sum_{\widetilde{w}_{\mathcal{A}_a} \in \mathcal{D}_{\mathcal{A}'}} \mathrm{P}\Big\{\mathbf{V}^{(\mathcal{A}_a)}_{\widetilde{w}_{\mathcal{A}_a}} \in \mathcal{T}^n_{[\mathsf{U}_{\mathcal{A}_a}]\delta}\Big|\mathcal{E}_1^{\mathrm{c}}, \widetilde{\mathbf{U}}_{\mathcal{A}'^{\mathrm{c}}} = \boldsymbol{u}_{\mathcal{A}'^{\mathrm{c}}}\Big\} \\ &\leq \exp\left[n\Big(\sum_{k \in \mathcal{A}'} \Delta_k\Big)\right] \frac{\left|\mathcal{T}^n_{[\mathsf{U}_{\mathcal{A}'}|\mathsf{U}_{\mathcal{A}'^{\mathrm{c}}}]\delta}(\boldsymbol{u}_{\mathcal{A}'^{\mathrm{c}}})\right|}{\prod_{k \in \mathcal{A}'}\left|\mathcal{T}^n_{[\mathsf{U}_k]\delta}\right|} \\ &\overset{(k)}{\leq} \exp\left[n\Big(\varepsilon(\delta) + \mathrm{H}(\mathsf{U}_{\mathcal{A}'}|\mathsf{U}_{\mathcal{A}'^{\mathrm{c}}}) + \sum_{k \in \mathcal{A}'}\big(\Delta_k - \mathrm{H}(\mathsf{U}_k)\big)\Big)\right]\end{aligned} \tag{19}$$

where $\varepsilon(\delta) = \sum_{k \in \mathcal{A}' \cup 0} \varepsilon_k(\delta)$ goes to zero as $\delta \to 0$ and (k) follows from the properties of types [10, Sections 2.4 and 2.5]. We observe that the definition of $\widetilde{R}_k$ and (1) imply for any $\varnothing \neq \mathcal{A}' \subseteq \mathcal{A}_a$ with $\mathcal{A}' \cap \mathcal{A}_b \neq \varnothing$,

$$\sum_{k \in \mathcal{A}'} \Delta_k \leq -\frac{\varepsilon''}{2} - \mathrm{H}(\mathsf{U}_{\mathcal{A}'}|\mathsf{U}_{\mathcal{A}'^{\mathrm{c}}}) + \sum_{k \in \mathcal{A}'} \mathrm{H}(\mathsf{U}_k) \ . \tag{20}$$

By taking the marginal distribution on $\widetilde{\mathbf{U}}_{\mathcal{A}'^{\mathrm{c}}}$ in (19) and using (20), we obtain:

$$\mathrm{P}\{\mathcal{E}_{\mathcal{A}'}|\mathcal{E}_1^{\mathrm{c}}\} \leq \exp\left[n\left(\varepsilon(\delta) - \frac{\varepsilon''}{2}\right)\right] \leq \varepsilon' \ , \tag{21}$$

for n large enough and δ small enough. Similar reasoning for $\mathrm{P}\{\mathcal{E}_3|\mathcal{E}_1^{\mathrm{c}}\}$ and combination of (17), (18) and (21) yields

$$\mathrm{P}\{\mathcal{E}_0\} \leq \varepsilon' + 2^{|\mathcal{A}_a|}\varepsilon' + 2^{|\mathcal{B}_a|}\varepsilon' \leq 2^K\varepsilon' \ . \tag{22}$$

For a set $\varnothing \neq \mathcal{A} \subseteq \mathcal{K}$, we next analyze the random quantity $\mathsf{L} \triangleq |\mathcal{C}_{\mathcal{A}} \cap \mathcal{T}^n_{[\mathsf{U}_{\mathcal{A}}]\delta}|$. For large n and $\mathbf{V}_{\mathcal{A}} \in \mathcal{C}_{\mathcal{A}}$, we have

$$\begin{aligned}\mathbb{E}[\mathsf{L}] &\leq \sum_{\mathbf{V}_{\mathcal{A}} \in \mathcal{C}_{\mathcal{A}}} \mathbb{E}\Big[\mathbb{1}_{\mathcal{T}^n_{[\mathsf{U}_{\mathcal{A}}]\delta}}(\mathbf{V}_{\mathcal{A}})\Big] = \left(\prod_{k \in \mathcal{A}} \widetilde{M}_k\right)\mathbb{E}\Big[\mathbb{1}_{\mathcal{T}^n_{[\mathsf{U}_{\mathcal{A}}]\delta}}(\mathbf{V}_{\mathcal{A}})\Big] \\ &= \left(\prod_{k \in \mathcal{A}} \widetilde{M}_k\right)\frac{\left|\mathcal{T}^n_{[\mathsf{U}_{\mathcal{A}}]\delta}\right|}{\prod_{k \in \mathcal{A}}\left|\mathcal{T}^n_{[\mathsf{U}_k]\delta}\right|} \\ &\overset{(l)}{\leq} \left(\prod_{k \in \mathcal{A}} \widetilde{M}_k\right)\frac{\exp[n(\mathrm{H}(\mathsf{U}_{\mathcal{A}}) + \varepsilon_0(\delta))]}{\exp\big[n(\sum_{k \in \mathcal{A}} \mathrm{H}(\mathsf{U}_k) - \varepsilon_k(\delta))\big]} \\ &\leq \exp\left[n\left(\mathrm{I}(\mathsf{U}_{\mathcal{A}}; \mathsf{X}_{\mathcal{A}}) + \hat{\varepsilon}(\delta) + |\mathcal{A}|\frac{\varepsilon''}{2}\right)\right] \ ,\end{aligned}$$

where $\mathbb{1}_{\mathcal{Z}}(\cdot)$ denotes the indicator function of a set $\mathcal{Z}$ and $\hat{\varepsilon}(\delta) = \sum_{k \in \mathcal{A} \cup 0} \varepsilon_k(\delta)$ vanishes to zero as $\delta \to 0$. Here, (l) follows from the properties of types [10, Sections 2.4 and 2.5]. We can choose ε'' enough small s.t. $\hat{\varepsilon}(\delta) + K\varepsilon''/2 < \varepsilon$ for suitably small δ. Let $\mathcal{E}_4 = \{\mathsf{L} \geq \exp[n(\mathrm{I}(\mathsf{U}_{\mathcal{A}}; \mathsf{X}_{\mathcal{A}}) + \varepsilon)]\}$ and from Markov's inequality provided that n is large enough:

$$\mathrm{P}\{\mathcal{E}_4\} \leq \exp\left[n\left(\hat{\varepsilon}(\delta) - \varepsilon + |\mathcal{A}|\frac{\varepsilon''}{2}\right)\right] \leq \varepsilon' \ . \tag{23}$$

As ε' was arbitrary in (22) and (23), we can obtain deterministic encoding and decoding functions, s.t. (4) holds whenever the conditions of part 1 are satisfied. Taking into account that $g_{\mathcal{A}_a,\mathcal{A}_b}(\mathcal{M}_{\mathcal{A}_a}) \times g_{\mathcal{B}_a,\mathcal{B}_b}(\mathcal{M}_{\mathcal{B}_a}) \subseteq \mathcal{C}_{\mathcal{A}_b \cup \mathcal{B}_b}$, we also have (5). Notice that, given a specific code, $\mathrm{P}\{\mathcal{E}_4\} < 1$ already implies $\mathrm{P}\{\mathcal{E}_4\} = 0$ as the event $\mathcal{E}_4$ is fully determined by the code $\mathcal{C}_{\mathcal{K}}$ alone. ■

REFERENCES

[1] G. Pichler, P. Piantanida, and G. Matz, "Distributed information-theoretic biclustering of two memoryless sources," in *53rd Annual Allerton Conference on Communication, Control, and Computing*, 2015.

[2] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, "Cluster analysis and display of genome-wide expression patterns," *Proc. of the Nat. Academy of Sciences*, vol. 95, no. 25, pp. 14 863–14 868, 1998.

[3] K.-P. Wong, D. Feng, S. R. Meikle, and M. J. Fulham, "Segmentation of dynamic PET images using cluster analysis," *IEEE Trans. Nucl. Sci.*, vol. 49, no. 1, pp. 200–207, 2002.

[4] G. Punj and D. W. Stewart, "Cluster analysis in marketing research: review and suggestions for application," *J. of Marketing Research*, vol. 20, no. 2, pp. 134–148, 1983.

[5] T. S. Han, "Hypothesis testing with multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 33, no. 6, pp. 759–772, 1987.

[6] T. S. Han and S. Amari, "Statistical inference under multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2300–2324, 1998.

[7] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources," *IEEE Trans. Inf. Theory*, vol. 25, no. 2, pp. 219–221, March 1979.

[8] T. A. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 740–761, Jan 2014.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 2006.

[10] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[11] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

[12] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

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