Distributed Computing and the Graph Entropy Region

Ofer Shayevitz, *Member, IEEE*

***Abstract*— Two remote senders observe X and Y, respectively, and can noiselessly send information via a common relay node to a receiver that observes Z. The receiver wants to compute a function $f(X, Y, Z)$ of these possibly related observations, without error. We study the average number of bits that need to be conveyed to that end by each sender to the relay and by the relay to the receiver, in the limit of multiple instances. We relate these quantities to the entropy region of a probabilistic graph with respect to a Cartesian representation of its vertex set, which we define as a natural extension of graph entropy. General properties and bounds for the graph entropy region are derived, and mapped back to special cases of the distributed computing setup.**

***Index Terms*— Distributed source coding, zero-error information theory, graph entropy.**

I. Introduction

THE ENTROPY of a probabilistic graph was introduced by Körner [1] as a natural generalization of the Shannon entropy, by associating an information source V with a graph G over its alphabet, where two symbols are adjacent in the graph if and only if they can be distinguished. One is then interested in compressing the source such that the source sequence and its associated reconstruction sequence are indistinguishable; the optimal compression rate in the limit of multiple instances such that the probability of indistinguishability approaches one, is called the *graph entropy* of the pair (G, V), and is denoted by $H(G, V)$. For a complete graph, the graph entropy trivially coincides with the Shannon entropy $H(V)$ of the source. More generally, $H(G, V)$ admits a single letter expression as the minimum mutual information over all channels whose input is V and whose output is an independent set of G containing V [1]. Since its introduction, graph entropy has been applied in diverse problems such as perfect hashing [2], Boolean circuit size [3], counting of 'very different' sequences [4], and complexity of sorting from a partial order [5]. For an extensive review of graph entropy and its applications, see [6].

A different information-theoretic interpretation of graph entropy was put forward in [7], where the authors considered a point-to-point source coding problem in which a sender would like to describe X to a receiver that knows a dependent Z, without error. In that work, the minimal one-shot rate from sender to receiver was characterized as the *chromatic entropy* of $(G_{X|Z}, X)$, where $G_{X|Z}$ is the associated *confusability graph*, also known as Witsenhausen's characteristic graph [8]. The minimal asymptotical (per-instance) rate was then shown to be the limit of the (normalized) chromatic entropy of $(G_{X|Z}^{(n)}, X^n)$, where $(\cdot)^{(n)}$ is the n-fold graph AND-product. This quantity was further studied in [9], where it was shown to coincide with the *complementary graph entropy* defined in [10]. A closed form expression for the complementary graph entropy is unknown; in fact, such an expression would yield in particular the zero-error capacity of a graph [11], a notorious open problem. In [7], the authors also considered a smaller family of protocols for *unrestricted inputs*, where the side information sequence z^n is allowed to be arbitrary, and exact reconstruction is guaranteed on each instance k where (X_k, z_k) is in the support set of p_{XZ}. It was shown that the associated minimal asymptotical rate is the limit of the (normalized) chromatic entropy of $(G_{X|Z}^n, X^n)$, where $(\cdot)^n$ is the n-fold graph OR-product, and that this limit is exactly the graph entropy $H(G_{X|Z}, X)$. This serves as an upper bound for the corresponding complementary graph entropy, whereas the conditional Shannon entropy $H(X|Z)$ serves as a trivial lower bound. Both bounds can be arbitrarily loose [7].

A more general problem of zero-error distributed source coding was studied in [12, Section III]. In that setup, two separated senders observe two dependent sources X and Y respectively, and would like to describe their observations to a common receiver, without error. The achievable rate region under unrestricted inputs[1] was given a single letter formula, by considering a natural bipartite graph coloring problem. Specifically, it was demonstrated that in contrast to the standard vanishing error setup of Slepian and Wolf [13], the entire zero-error rate region cannot generally be achieved by time-sharing two point-to-point side information protocols.

In this paper we discuss the problem of zero-error distributed source coding/computing over a simple unidirectional noiseless network consisting of two senders, a relay, and a receiver. In this setting the senders know some dependent X and Y respectively, while the receiver knows a dependent Z, and would like to compute some function $f(X, Y, Z)$, without error. The senders can communicate with the receiver only via the relay. The setting is depicted in Fig. 1. We are interested in the asymptotical rates, i.e., the per-instance expected number of bits, that need to be sent to and from the relay to that end, in the limit of multiple i.i.d. instances,

Manuscript received December 26, 2011; revised October 10, 2013; accepted November 12, 2013. Date of publication January 30, 2014; date of current version May 15, 2014. This work was supported by the Information Theory and Applications Center, University of California, San Diego. This paper was presented at the 2011 Data Compression Conference and the 2011 Information Theory and Applications Workshop.

The author is with the Department of Electrical Engineering-Systems, Tel Aviv University, Tel Aviv 6997801, Israel (e-mail: ofersha@eng.tau.ac.il).

Communicated by J. Körner, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2014.2303802

[1] The paper [12] also discusses the restricted input setting.

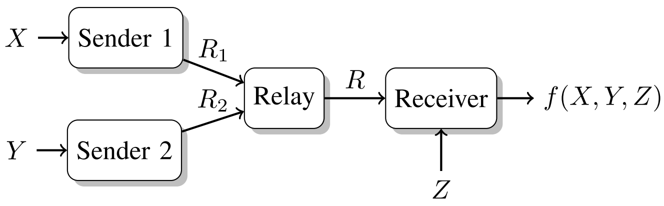

Fig. 1. Distributed computing with side information over a simple relay network.

and under an unrestricted inputs assumption. To that end, we introduce the concept of an entropy region of a probabilistic graph w.r.t a Cartesian representation of is vertex set, a natural generalization of the scalar graph entropy, and show it pertains to the optimal rate region for our distributed computing problem. We derive several inner and outer bounds for the graph entropy region, and discuss some of its properties.

Restricted vs. Unrestricted Inputs. Let us discuss the distinction between restricted and unrestricted inputs in our distributed computing setup. In the restricted input setting, the senders and receiver observe n i.i.d. triplets drawn from a given distribution p, and the receiver wants to compute the function f for each triplet, without error. We are then interested in finding the set of expected rates (as $n \to \infty$) that can be achieved by protocols facilitating this. This interesting setting is notoriously difficult to analyze even in the simplest of cases, e.g., in the aforementioned special case of source coding with side information [7]. Moreover, from a practical standpoint a restricted input protocol is very sensitive to model support errors, since the appearance of even a single input triplet that lies outside the support of p can mess up the computation for other (and possibly all) input triplets. In a more realistic setting, zeros in p may in fact represent outliers or corrupt data events that are unlikely to occur in any given block, hence their absence can be safely assumed for compression purposes; yet, in case they do appear one is still interested in correctly recovering the computation results for all uncorrupted triplets. This naturally leads us to consider unrestricted inputs protocols which further guarantee that the computation result is always correct for each triplet in the support of p, regardless of whether other triplets follow suit. Formally, in the unrestricted input setting the senders and receiver may observe n *arbitrary* triplets, and the receiver is required to compute the function f only for triplets that lie in the support of p, without error. We are then interested in finding the set of expected rates (as $n \to \infty$) achieved by protocols that facilitate this, assuming the triplets were drawn in an i.i.d. fashion from p.[2] Note that in this setup the receiver is not necessarily able to detect corrupt data events; one may either assume that it somehow learns their locations in hindsight and discards them, or that it simply does not care about the value of the function in these locations.

Related Work. Distributed source coding and function computation problems over network setups have been extensively studied in the past under the (markedly different) asymptotically vanishing error probability criterion. In particular, works that probably bear the most resemblance to the settings considered herein are the cascade source coding paper [14], and the distributed function computation papers [15]–[18].

Organization. In Section II some notations are introduced and the necessary mathematical background is provided. In Section III, the graph entropy region is defined, the zero-error computing setup is introduced, and the relation between them is established. In the few Sections that follow, some subregions of the graph entropy region pertaining to special cases of source coding/computing problems are discussed, and several bounds as well as various properties are derived: Section IV characterizes the subregions pertaining to the point-to-point case; Section V characterizes the subregion associated with distributed computing of dependent sources with side information, and establishes some graph entropy region properties; Section VI provides an outer bound for the entire graph entropy region and gives conditions for tightness; Section VII provides inner bounds for the subregion corresponding to the problem of cascade computing with side information; and Section VIII provides inner bounds for the entire graph entropy region. A brief discussion of some open questions appears in Section IX.

II. Preliminaries

A. Notations

A function $f : \mathcal{X} \mapsto \mathcal{Y}$ naturally extends to a function $f : 2^{\mathcal{X}} \mapsto 2^{\mathcal{Y}}$ between the associated power sets via per-element evaluation, i.e., $f(S) \stackrel{\text{def}}{=} \{y : y = f(x), x \in S\}$. We denote the associated inverse image function by $f^{-1} : 2^{\mathcal{Y}} \mapsto 2^{\mathcal{X}}$. Note that we allow the domain of f^{-1} to be the entire power set $2^{\mathcal{Y}}$ and not just $2^{f(\mathcal{X})}$, which means it can return the empty set. We write f^n for the n-fold Cartesian product of f. We denote the set of all finite length binary strings by $\{0,1\}^*$. The length of a string $s \in \{0,1\}^*$ is denoted by $|s|$. The cardinality of a finite set A is denoted by $|A|$. The $+$ operation between two regions in $\mathbb{R}^m$ is understood to be the Minkowski addition, while multiplication by a constant is interpreted as a coordinate-wise operation.

A random variable (r.v.) X taking values in a finite alphabet $\mathcal{X}$ is associated with a probability mass function (p.m.f.) $p_X(x)$ over $\mathcal{X}$, and we write $X \sim p_X(x)$. We omit the subscript when there is no confusion. We write $S_X \stackrel{\text{def}}{=} \{x \in \mathcal{X} : p(x) > 0\}$ for the associated support set. Let $(X,Y) \sim p(x,y)$ be a pair of r.v.'s over a finite product alphabet $\mathcal{X} \times \mathcal{Y}$. For any $y \in \mathcal{Y}$, we write S_{XY} for the joint support, and $S_{X|Y}(y) \stackrel{\text{def}}{=} \{x \in \mathcal{X} : p(x,y) > 0\}$ for the conditional support given $Y = y$. A random sequence $X^n \stackrel{\text{def}}{=} (X_1, \ldots, X_n)$ is said to be *p_X-independent-identically distributed (p_X-i.i.d.)* if $p_{X^n}(x^n) = \prod_{t=1}^n p_X(x_t)$ for all x^n. Let (X^n, Y^n) be two jointly distributed random sequences, and let $p_{Y|X}$ be some conditional distribution. We say that Y^n is *$p_{Y|X}$-independent given X^n* if $p_{Y^n|X^n}(y^n|x^n) = \prod_{t=1}^n p_{Y|X}(y_t|x_t)$ for all y^n and x^n with $p_{X^n}(x^n) > 0$. When we say that two or more random variables/sequences/sets are (possibly conditionally) independent, we mean mutually (possibly conditionally) independent, unless otherwise stated. The indicator r.v. associated with an event A is denoted $\mathbb{1}_A$.

[2] In fact, it is sufficient to assume that X_k, Y_k are drawn i.i.d. and z_k is arbitrary, similar to [7].

Let U be a r.v. distributed over $2^{\mathcal{X}}$. We write $X \in U$ to denote that U contains X with probability one, i.e., that for any $x \in \mathcal{X}$ with $p(x) > 0$,

$$\sum_{u \ni x} p(u|x) = 1.$$

B. Information-Theoretic Notions

The (Shannon) entropy of X is denoted $H(X)$. The mutual information between two r.v.'s (X, Y) is denoted by $I(X; Y)$. For any $x^n \in \mathcal{X}^n$, let ν_{x^n} be the p.m.f. over $\mathcal{X}$ that corresponds to the relative frequency of symbols in x^n. For $\varepsilon > 0$, define the (n, ε)-typical set associated with X to be[3]

$$\mathcal{T}_\varepsilon^n(X) \stackrel{\text{def}}{=} \{x^n \in \mathcal{X}^n : \forall x \in \mathcal{X}, |p(x) - \nu_{x^n}(x)| \le \varepsilon p(x)\}.$$

An important property of this definition of typicality is that $p(x) = 0$ implies that $\nu_{x^n}(x) = 0$ for all $x^n \in \mathcal{T}_\varepsilon^n(X)$. The joint typical set $\mathcal{T}_\varepsilon^n(X, Y)$ associated with a pair of r.v. X, Y is defined similarly.

The following well known Lemmas play a central role in the sequel.

Lemma 1 (Conditional Typicality Lemma [20])**.** Let p_{XY} be some joint distribution. Suppose $x^n \in \mathcal{T}_{\varepsilon'}^n(p_X)$ for some $\epsilon' > 0$, and Y^n is $p_{Y|X}$-independent given $X^n = x^n$. Then for every $\varepsilon > \varepsilon'$:

$$\lim_{n\to\infty} \Pr\left((x^n, Y^n) \notin \mathcal{T}_\varepsilon^n(p_{XY})\right) = 0.$$

Lemma 2 (Multivariate Covering Lemma, [20])**.** *Let* $(U_0, U_1, \dots U_k) \sim p(u_0, u_1, \dots, u_k)$ *and* $0 < \varepsilon' < \varepsilon$. *Let* U_0^n *be a random sequence satisfying* $\mathbb{P}(U_0^n \in \mathcal{T}_{\varepsilon'}^n(U_0)) \to 1$ *as* $n \to \infty$. *For each* $j \in \{1 \dots, k\}$, *let* $\{U_j^n(m_j)\}_{m_j=1}^{2^{nr_j}}$ *be a set of pairwise conditionally independent random sequences given* U_0^n, *where each sequence is* $p_{U_j|U_0}$*-independent given* U_0^n. *Assume that the sets* $\{U_1^n(m_1)\}_{m_1=1}^{2^{nr_1}}, \dots \{U_k^n(m_k)\}_{m_k=1}^{2^{nr_k}}$ *are mutually conditionally independent given* U_0^n. *Then there exists* $\delta(\varepsilon) \to 0$ *as* $\varepsilon \to 0$ *such that*

$$\lim_{n\to\infty} \mathbb{P}\left((U_0^n, U_1^n(m_1), \dots, U_k^n(m_k)) \notin \mathcal{T}_\varepsilon^n(U_0, U_1, \dots, U_k)\right.$$
$$\left.\text{for all } (m_1, \dots m_k)\right) = 0,$$

if for any $J \subseteq \{1, \dots, k\}$ *with* $|J| \ge 2$

$$\sum_{j\in J} r_j > \sum_{j\in J} H\left(U_j\right) - H\left(\{U_j\}_{j\in J}\right) + \delta(\varepsilon).$$

C. Graph-Theoretic Notions

Let G be a graph with a vertex set $\mathcal{V}$. A set $A \subseteq \mathcal{V}$ is called an *independent set* of G if no two vertices in A are adjacent in G, and a *maximal independent set* if no other independent set strictly contains it. We denote by $\Gamma(G)$ (resp. $\overline{\Gamma}(G)$) the set of all independent (resp. maximal independent) sets of G. A coloring of G is any function c over $\mathcal{V}$ set such that $c^{-1}(\cdot)$ induces a partition of $\mathcal{V}$ into independent sets of G. For two graphs G, F over a common vertex set, $G \subseteq F$ refers to the inclusion of edge sets. The complementary graph G^c is a graph on the same vertex set, with the complementary edge set. The n-fold OR-product of G, denoted G^n, is a graph with a vertex set $\mathcal{V}^n$ where v^n and v'^n are adjacent if and only if v_k and v'_k are adjacent in G for some $k \in \{1, \dots, n\}$.

Let $(X, Y) \sim p(x, y)$ over a finite product alphabet $\mathcal{X} \times \mathcal{Y}$. The *confusability graph* $G_{X|Y}$ has a vertex set $\mathcal{X}$, where (x, x') is an edge if and only if both $x, x' \in S_{X|Y}(y)$ for some $y \in \mathcal{Y}$. More generally, for any function $f(x, y)$, the f-confusability graph $G_{X|Y}^f$ has (x, x') as an edge if and only if both x, $x' \in S_{X|Y}(y)$ and $f(x, y) \neq f(x', y)$, for some $y \in \mathcal{Y}$.

A *probabilistic graph* is a pair (G, V) where G is graph and V is a r.v. distributed over the vertex set of G. One example is $(G_{X|Y}, X)$. The *graph entropy* of (G, V) is defined to be

$$H(G, V) \stackrel{\text{def}}{=} \min_{V \in U \in \overline{\Gamma}(G)} I(V; U). \tag{1}$$

Namely, the minimum is taken over all conditional distributions $p_{U|V}$ such that U, a random maximal independent set of G, contains V with probability one (recall the definition of the relation $V \in U$).[4] The original definition of graph entropy was in terms of the limiting behavior of the chromatic number of a high probability subgraph of G^n [1], which was then shown to reduce to (1). Here we mention an essentially similar asymptotical characterization, following [7]. Define the chromatic entropy of (G, V) to be

$$H_\chi(G, V) \stackrel{\text{def}}{=} \min\{H(c(V)) : c \text{ is a coloring of } G\}.$$

Lemma 3 (Chromatic entropy characterization [7])**.**

$$H(G, V) = \lim_{n\to\infty} \frac{1}{n} H_\chi(G^n, V^n)$$

where V^n *is* p_V*-i.i.d.*

Graph entropy admits an additional characterization via the notion of *vertex packing* [21]. The *characteristic vector* of a set of vertices $A \subseteq \mathcal{V}$ is a column vector $\mathbf{a} \in \mathbb{R}^{|\mathcal{V}|}$ where $a_i = 1$ if the ith vertex is in A, and $a_i = 0$ otherwise. The *vertex packing polytope* $\mathbb{VP}(G)$ is the convex hull of the characteristic vectors associated with $\Gamma(G)$. Write $\mathbf{p}_V \in \mathbb{R}^{|\mathcal{V}|}$ for the probability column vector associated with V.

Lemma 4 (Vertex packing characterization [21])**.**

$$H(G, V) \stackrel{\text{def}}{=} \min_{\mathbf{a} \in \mathbb{VP}(G), \mathbf{a} > 0} -\mathbf{p}_V^T \cdot \log(\mathbf{a}).$$

In the next two Lemmas we mention some useful properties of graph entropy.

Lemma 5.

(i) If G is empty, then $H(G, V) = 0$.
(ii) If G is complete, then $H(G, V) = H(V)$.
(iii) (*Monotonicity*) If $G \subseteq F$ then $H(G, V) \le H(F, V)$.
(iv) (*Subadditivity*) $H(F \cup G, V) \le H(F, V) + H(G, V)$.

Two probabilistic graphs (G, V) and (F, Q) are said to be *independent*, if 1) their respective vertex sets are disjoint, and 2) V, Q are independent r.v.'s. Let v be some vertex in G. Define a new probabilistic graph $(G_{v\leftarrow F}, V_{v\leftarrow Q})$ by deleting v and connecting every vertex in F to those vertices in G that

[3]This definition of typically, also known as *robust typicality*, was originally introduced in [19].

[4]It is easily verified that minimizing over $V \in U \in \Gamma(G)$, namely without the maximality restriction, yields the same minimum.

were adjacent to v, and letting $V_{v\leftarrow Q} \overset{\text{def}}{=} V\mathbb{1}_{\{V\neq v\}} + Q\mathbb{1}_{\{V=v\}}$. This operation is known as *substitution*.

Lemma 6 (Substitution Lemma [6], [22])**.** Let (G, V) and (F, Q) be a pair of independent probabilistic graphs. Then

$$H(G_{v\leftarrow F}, V_{v\leftarrow Q}) = H(G, V) + P_V(v) \cdot H(F, Q).$$

III. Formulation

The problem of distributed computing via a relay with side information at the receiver is formally described in Subsection A. In Subsection B, the graph entropy region for a probabilistic graph w.r.t. a Cartesian representation of its vertex set is defined. The relation between the two problems is established in Subsection C. The remainder of the paper is then dedicated to studying the properties of the graph entropy region and their distributed computing implications.

A. The Distributed Computing Setup

Let $(X, Y, Z) \sim p(x, y, z)$ over a finite product alphabet $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. A sender that knows X and another sender that knows Y communicate with a receiver that knows Z, via a common relay node. The receiver would like to compute a function $f(X, Y, Z)$, *without error*. We are interested in the asymptotical rates (i.e., the per-instance expected number of bits in the limit of multiple i.i.d. instances) that each sender must transmit to the relay, and the relay in turn to the receiver, to that end. The setting is depicted in Fig. 1. We specifically consider communication protocols for *unrestricted inputs*, in a sense to be described shortly. We also assume that both the relay and the receiver are able to tell when the message they receive ends, although this is not essential to our discussion.

A *(deterministic, zero-error) one-shot protocol* for the computing setup (X, Y, Z, f) consists of two *sender mappings* $\phi_1 : \mathcal{X} \mapsto \{0, 1\}^*$ and $\phi_2 : \mathcal{Y} \mapsto \{0, 1\}^*$, and a *relay mapping* $\phi : \phi_1(\mathcal{X}) \times \phi_2(\mathcal{Y}) \mapsto \{0, 1\}^*$. The mappings satisfy the following properties:

(i) The ranges of ϕ_1, ϕ_2 and ϕ are prefix free sets.[5]

(ii) The pair $(\phi(\phi_1(x), \phi_2(y)), z)$ uniquely determines $f(x, y, z)$ over S_{XYZ}.

A *n-shot protocol* for the computing setup (X, Y, Z, f) is a one-shot protocol for the setup (X^n, Y^n, Z^n, f^n). An *unrestricted inputs n-shot protocol* is an n-shot protocol with the following additional property:

(iii) The pair $(\phi(\phi_1(x^n), \phi_2(y^n)), z^n)$ uniquely determines $f(x_k, y_k, z_k)$ for all k for which $(x_k, y_k, z_k) \in S_{XYZ}$.[6]

Namely, unrestricted inputs protocols are robust in the sense of providing a guarantee that arbitrary errors could affect the computation result only at the instance where they appear.[7]

The rate triplet (R_1, R_2, R) achieved by a n-shot protocol (ϕ_1, ϕ_2, ϕ) is defined to be

$$R_1 \overset{\text{def}}{=} \frac{1}{n}\,\mathbb{E}|\phi_1(X^n)|$$
$$R_2 \overset{\text{def}}{=} \frac{1}{n}\,\mathbb{E}|\phi_2(Y^n)|$$
$$R \overset{\text{def}}{=} \frac{1}{n}\,\mathbb{E}|\phi(\phi_1(X^n), \phi_2(Y^n))|.$$

We define the rate-region $\mathcal{R}(X, Y, Z, f)$ associated with (X, Y, Z, f) to be the closure of the set of all rate triplets (R_1, R_2, R) achievable by some n-shot protocol. Similarly, we define rate-region $\mathscr{R}(X, Y, Z, f)$ to be the closure of the set of all rate triplets (R_1, R_2, R) achievable by some unrestricted inputs n-shot protocol. For the special case of *distributed source coding with side information*, i.e., where $f(x, y, z) = (x, y)$, we omit the function f and write $\mathcal{R}(X, Y, Z)$ and $\mathscr{R}(X, Y, Z)$ for the associated rate regions. In the sequel, we limit our discussion to $\mathscr{R}(X, Y, Z, f)$. Our inner bounds will clearly also hold for $\mathcal{R}(X, Y, Z, f)$, but may be arbitrarily loose.

B. The Graph Entropy Region

A (two-dimensional) *Cartesian representation* of a finite set $\mathcal{V}$ is a one-to-one (but not necessarily onto) mapping $\pi : \mathcal{V} \mapsto \mathcal{X} \times \mathcal{Y}$, where $\mathcal{X}, \mathcal{Y}$ are finite sets. Without loss of generality, we assume throughout that the associated coordinate mappings ($\mathcal{V} \mapsto \mathcal{X}$ and $\mathcal{V} \mapsto \mathcal{Y}$) are both onto. Let G ba a graph and let π be a Cartesian representation of its vertex set $\mathcal{V}$. A triplet of functions (c_1, c_2, c) over $(\mathcal{X}, \mathcal{Y}, \mathcal{V})$ respectively is called a *color cover* for (G, π) if

(i) both $(c_1 \times c_2) \circ \pi$ and c are colorings of G.

(ii) $(c_1 \times c_2) \circ \pi$ refines c, i.e., each color class of the latter is a union of color classes of the former.

Let (G, V, π) be a probabilistic graph with an associated Cartesian representation, and write $(X, Y) \overset{\text{def}}{=} \pi(V)$. These conventions will be used throughout. We define the *chromatic entropy region* of (G, V, π) to be

$$\begin{aligned} H_\chi(G, V, \pi) \overset{\text{def}}{=} \bigcup_{(c_1,c_2,c)} \{(b_1, b_2, b) : b_1 &\geq H(c_1(X)),\\ b_2 &\geq H(c_2(Y)),\\ b &\geq H(c(V))\} \end{aligned}$$

where the union is taken over all color covers for (G, π). We define the corresponding *graph entropy region* to be

$$H(G, V, \pi) \overset{\text{def}}{=} \bigcup_n \frac{1}{n} H_\chi(G^n, V^n, \pi^n).$$

The next lemma provides some basic properties of the graph entropy region.

Lemma 7.

(i) *If G is empty then $H(G, V, \pi) =$ {all nonnegative triplets}*

(ii) *If G is complete and π is onto, then*[8]

$$H(G, V, \pi) = \{(R_1, R_2, R) : R_1 \geq H(X), R_2 \geq H(Y), R \geq H(V)\}.$$

[5]Note that ϕ_1, ϕ_2 are generally not one-to-one mappings. The prefix condition can be relaxed as discussed in [7], but this can save no more than $O(\log n/n)$ in rates for a n-shot protocol (to be immediately defined), hence does not affect our asymptotic discussion.

[6]Note that this additional property in fact implies property (ii) of an n-shot protocol, but not vice-versa, since for a n-shot protocol the pair above uniquely determines the entire vectorized function $f^n(x^n, y^n, z^n)$ over S^n_{XYZ}.

[7]See also a discussion in Section I on the distinction between restricted and unrestricted inputs settings.

[8]The case where π is not onto will generally yield a larger region.

(iii) *(Invariance to row/column permutations) If* $\pi'(v) = (\sigma_1 \times \sigma_2)(\pi(v))$ *where* σ_1 *and* σ_2 *are permutations of* $\mathcal{X}$ *and* $\mathcal{Y}$ *respectively, then* $H(G, V, \pi) = H(G, V, \pi')$.
(iv) *(Monotonicity) If* $G \subseteq F$ *then* $H(G, V, \pi) \supseteq H(F, V, \pi)$.
(v) *(Subadditivity)* $H(F \cup G, V, \pi) \supseteq H(F, V, \pi) + H(G, V, \pi)$.

Proof. See the Appendix. □

A partial generalization of the substitution Lemma will be presented in Subsection V-C.

Projections. In the next subsection, we give an operational interpretation for the graph entropy region in the realm of distributed computing, which provides impetus to study and characterize this region. In particular, we discuss various special cases of the distributed computing setup, which correspond to the following projections of the graph entropy region onto its coordinates:

$H^{(j)}(G, V, \pi) \subseteq \mathbb{R}$: *The projection onto the* jth *coordinate, i.e., the set of all values this coordinate can attain in* $H(G, V, \pi)$*, where* $j \in \{1, 2, 3\}$.

$H^{(i,j)}(G, V, \pi) \subseteq \mathbb{R}^2$ for $i \neq j$: *The projection onto the* (i, j) *coordinates, i.e., the set of all values this coordinate pair can attain in* $H(G, V, \pi)$*, where* $i, j \in \{1, 2, 3\}$ *and* $i \neq j$.

Marginal Graphs. Let us define some natural marginal graphs associated with (G, π), which will prove elemental in the sequel for the purpose of bounding the graph entropy region and its various projections.

(i) The *row-union graph* $\pi^{(1)}(G)$ has vertex set $\mathcal{X}$, and x, x' are adjacent if and only if $\pi^{-1}(x, y)$ and $\pi^{-1}(x', y)$ are adjacent in G for some y.
(ii) The *row-projection graph* $\pi_\perp^{(1)}(G)$ has vertex set $\mathcal{X}$, and x, x' are adjacent if and only if $\pi^{-1}(x, y)$ and $\pi^{-1}(x', y')$ are adjacent in G for some y, y'.
(iii) The *row-support graph* $\underline{\pi}^{(1)}(G)$ has vertex set $\mathcal{X}$, and x, x' are adjacent if and only if $\pi(v) = (x, y)$ and $\pi(v') = (x', y)$ for some v, v' and y.

Note that the row-support graph depends only on π and the vertex set. The *column-union graph* $\pi^{(2)}(G)$, *column-projection graph* $\pi_\perp^{(2)}(G)$, and *column-support graph* $\underline{\pi}^{(2)}(G)$ are defined similarly. The next Lemma summarizes some basic relations between the different marginal graphs.

Lemma 8. *The following relations hold*:[9]

(i) $\pi_\perp^{(i)}(G) \supseteq \pi^{(i)}(G) \subseteq \underline{\pi}^{(i)}(G)$
(ii) $\left(\pi^{(i)}(G^c)\right)^c \subseteq \pi^{(i)}(G)$ *and* $\left(\pi_\perp^{(i)}(G^c)\right)^c \subseteq \pi_\perp^{(i)}(G)$.

Proof. See the Appendix. □

C. Relations

In this subsection we show that the rate region for the distributed computing setup with unrestricted inputs is given by an associated graph entropy region, a generalization of the scalar statement in [7].

Theorem 1. *Let* (G, V, π) *be a probabilistic graph with a Cartesian representation, and set* $(X, Y) \stackrel{\text{def}}{=} \pi(V)$*. Then for any r.v.* Z *and function* f *such that* $G = G^f_{V|Z}$,

$$\mathscr{R}(X, Y, Z, f) = H(G, V, \pi).$$

Furthermore, the following relations hold:

(i) $\pi^{(1)}(G^f_{V|Z}) = G^f_{X|YZ}$.
(ii) $\pi_\perp^{(1)}(G^f_{V|Z}) = G^{f'}_{X|Z}$*, where the function* $f' : \mathcal{X} \times \mathcal{Z} \mapsto \mathcal{X} \cup 2^{f(\mathcal{X}\times\mathcal{Y}\times\mathcal{Z})}$ *is given by*

$$f'(x, z) \stackrel{\text{def}}{=} \begin{cases} \tilde{f}(x, z) & |\tilde{f}(x, z)| \leq 1 \\ x & o.w. \end{cases}$$

and where $\tilde{f} : \mathcal{X} \times \mathcal{Z} \mapsto 2^{f(\mathcal{X}\times\mathcal{Y}\times\mathcal{Z})}$ *is given by*

$$\tilde{f}(x, z) \stackrel{\text{def}}{=} \{f(x, y, z) : y \in \mathcal{Y}, p(x, y, z) > 0\}.$$

(iii) $\underline{\pi}^{(1)}(G^f_{V|Z}) = G_{X|Y}$.

Proof. The idea here is very similar to [7] and [12, Section III]; the equivalence between color covers and protocols follows essentially from definition. Let (ϕ_1, ϕ_2, ϕ) be a one-shot protocol. Clearly, $\phi(\phi_1(x), \phi_2(y))$ is a coloring of $G^f_{XY|Z}$, as otherwise there exist (x, y) and (x', y') in $S_{XY}(z)$ for some z, such that $f(x, y) \neq f(x', y')$, contradicting zero-error. Therefore, $(\phi_1(x), \phi_2(y))$ is also a coloring of $G^f_{XY|Z}$ refining the former coloring. Conversely, the every color cover (c_1, c_2, c) yields a one-shot protocol, by mapping the ranges of the mappings be some prefix free sets. $c(c_1(x), c_2(y))$ and z uniquely determine $f(x, y, z)$ over S_{XYZ} as otherwise there must exist two feasible triplets with the same color but a different value of the function, contradicting the definition of $G^f_{XY|Z}$. Since the ranges are assumed prefix-free, a standard variable-length source coding result [23] implies that the minimal rates achievable by any one-shot protocol is precisely given by what we defined as the associated chromatic entropy region, up to one bit per coordinate.

For unrestricted inputs n-shot protocols, $G^f_{XY|Z}$ should be replaced by its n-fold OR-product.[10] This follows simply since $\overline{\Gamma}(G^n)$ is exactly the n-fold set product of $\overline{\Gamma}(G)$. If the receiver learns some independent set of $(G^f_{XY|Z})^n$ containing (x^n, y^n), it also knows a maximal one containing this pair, which in turn is a product of maximal independent sets of $G_{XY|Z}$. Following the one-shot discussion, the receiver can therefore compute $f(x_k, y_k, z_k)$ whenever $(x_k, y_k, z_k) \in S_{XYZ}$. The converse argument follows similarly. Therefore, the achievable rate region for unrestricted inputs n-shot protocols is the chromatic entropy of $G^n_{XY|Z}$, up to a factor of $O(\frac{1}{n})$ per coordinate. Taking the limit as $n \to \infty$, we obtain the graph entropy region. The relations between the marginal graphs and the confusability graphs are easy to verify. □

Lemma 9. *Let* (G, V) *be a probabilistic graph. Then there exists a r.v.* Z' *such that* $G_{V|Z'} = G$.

Proof. The claim follows by letting Z' be a random edge in G that is connected to V. □

Corollary 1. *For any distributed computing setup* (X, Y, Z, f) *there exists some distributed source coding setup* (X, Y, Z') *such that* $\mathscr{R}(X, Y, Z, f) = \mathscr{R}(X, Y, Z')$*. This immediately follows from Lemma 9 and Theorem 1.*

[9]Note that there is generally no inclusion relation between $\pi_\perp^{(i)}(G)$ and $\underline{\pi}^{(i)}(G)$.

[10]Note that for the restricted inputs setting, the OR-product needs to be replaced by the AND-product, which is significantly more difficult to analyze.

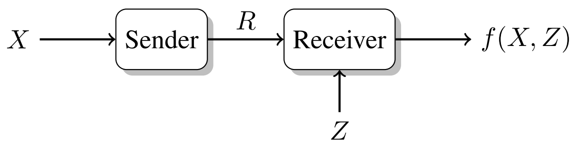

Fig. 2. Point-to-point computing with side information.

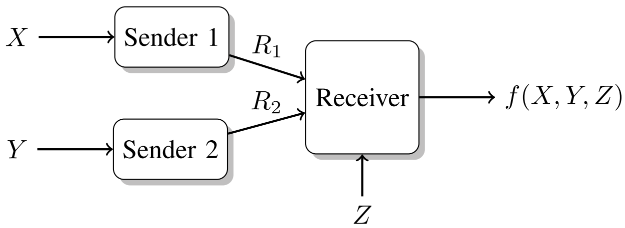

Fig. 3. Distributed computing with side information (no relay).

IV. $H^{(j)}(G, V, \pi)$: The Point to Point Case

The projection of the graph entropy region over each coordinate yields the point-to-point setting with receiver side information, similar to that of Fig. 2.

Theorem 2. *The following hold:*

(i) $H^{(1)}(G, V, \pi) = H(\pi^{(1)}(G), X)$
(ii) $H^{(2)}(G, V, \pi) = H(\pi^{(2)}(G), Y)$
(iii) $H^{(3)}(G, V, \pi) = H(G, V)$.

Proof. Case (i) corresponds to a single sender that knows X, and communicates directly with a receiver that knows (Y, Z). This setting was studied in [7], who proved that the optimal rate, for unrestricted inputs protocols, is given by $H(G_{X|YZ}, X)$. The claim now follows from Theorem 1, property (i). The results for the other two cases follow similarly. □

V. $H^{(1,2)}(G, V, \pi)$: Distributed Computing of Dependent Sources

The projection of the graph entropy region over the first two coordinates eliminates the relay and reduces the problem to that of distributed computing with receiver side information, as depicted in Fig. 3. This problem (sans the side information) was studied in [12, Section III], using a different formulation. In this section, we cast this result within the graph entropy region framework, and then apply it to derive a vertex packing characterization and a generalized version of the substitution Lemma.

A. The Region

Theorem 3. *Let $\widetilde{H}^{(1,2)}(G, V, \pi)$ be the closed convex hull of all pairs (R_1, R_2) satisfying*

$$R_1 \geq I(X; U), \quad R_2 \geq I(Y; W)$$

for some r.v's (U, W) satisfying

(i) $X \in U \in \Gamma(\pi^{(1)}(G))$.
(ii) $Y \in W \in \Gamma(\pi^{(2)}(G))$.
(iii) $\pi^{-1}(u \times w) \in \Gamma(G)$ *whenever* $p_U(u)p_W(w) > 0$.

Then $H^{(1,2)}(G, V, \pi) = \widetilde{H}^{(1,2)}(G, V, \pi)$.

Proof of $\supseteq$ Inclusion: We shall show the existence of a color cover (c_1, c_2, c) for (G^n, π^n) inducing an entropy region that approaches $\widetilde{H}^{(1,2)}(G, V, \pi)$ as $n \to \infty$. Let $\{U^n(m_1)\}_{m_1=1}^{2^{nR_1}}$ be a set of independently drawn p_U-i.i.d. random sequences. Define:

$$c_1(x^n) = \begin{cases} m_1 & (x^n, U^n(m_1)) \in \mathcal{T}_\varepsilon^n(X, U) \\ x^n & o.w. \end{cases} \quad (2)$$

where m_1 is the smallest index (if any) such that the condition above is satisfied. This in particular means that x_k is contained in the independent set $U_k(m_1)$ for all k. Similarly, let $\{W^n(m_2)\}_{m_2=1}^{2^{nR_2}}$ be a set of independently drawn p_W-i.i.d. random sequences, and define $c_2(y^n)$ accordingly. Since the function c is irrelevant here, trivially set $c = (c_1 \times c_2) \circ \pi^n$ to satisfy the refinement property. Now, we only need to show that $(c_1 \times c_2) \circ \pi$ is a coloring of G^n. To that end, we show that each color class is an independent set. We have four cases, depending on whether or not the condition in the definition (2) of c_1 and its counterpart for c_2 hold. For the first case, we have

$$\begin{aligned} &\{v^n \in G^n : (c_1 \times c_2) \circ \pi^n(v^n) = (m_1, m_2)\} \\ &\quad = \{v^n \in G^n : \forall k,\ v_k \in \pi^{-1}(U_k(m_1) \times W_k(m_2))\} \in \Gamma(G^n) \end{aligned}$$

where the equality follows from typicality and conditions (i) and (ii), and the inclusion follows from condition (iii) and the definition of the OR product. For the second case, we have

$$\begin{aligned} &\{v^n \in G^n : (c_1 \times c_2) \circ \pi^n(v^n) = (x^n, m_2)\} \\ &\quad = \{v^n \in G^n : \forall k,\ v_k \in \pi^{-1}(\{x^n\} \times W_k(m_2))\} \in \Gamma(G^n) \end{aligned}$$

where the equality follows from typicality and condition (ii), and the inclusion holds by virtue of condition (ii) and the definition of the OR product. The two other cases follow similarly, thereby confirming that (c_1, c_2, c) is a color cover for (G^n, π^n).

We now turn to analyze the achieved region. By Lemma 2, if $R_1 > I(X; U) + \delta(\varepsilon)$ then

$$\lim_{n\to\infty} \Pr\left(\exists m_1, (X^n, U^n(m_1)) \in \mathcal{T}_\varepsilon^n(X, U)\right) = 1.$$

Clearly then

$$H(c_1(X^n)) \leq nR_1 + \log|\mathcal{X}| \cdot o(n).$$

Similarly, if $R_2 > I(X; W) + \delta(\varepsilon)$ then

$$H(c_2(Y^n)) \leq nR_2 + \log|\mathcal{Y}| \cdot o(n).$$

The existence of a deterministic protocol achieving the same region follows from a standard argument. □

Proof of $\subseteq$ Inclusion in Theorem 3: Let us establish a simple additivity Lemma, generalizing a similar result for the graph entropy [7].

Lemma 10. $\widetilde{H}^{(1,2)}(G, V, \pi)$ is additive under the OR-product.

Proof. See the Appendix. □

Applying the lemma, we have that

$$\begin{aligned} H^{(1,2)}(G, V, \pi) &= \bigcup_n n^{-1} H_\chi^{(1,2)}(G^n, V^n, \pi^n) \\ &\subseteq \bigcup_n n^{-1} \widetilde{H}^{(1,2)}(G^n, V^n, \pi^n) \\ &= \widetilde{H}^{(1,2)}(G, V, \pi). \end{aligned}$$

□

Corollary 2. $H^{(1,2)}(G, V, \pi)$ is additive under the OR-product.

Remark 1. *Note that the rate region in Theorem 3 depends only on the marginals p_X, p_Y and the graph G. i.e., in the simple distributed computing setting we can assume without loss of generality that $p_{XYZ}(x, y, z) = p_X(x)p_Y(y)p_{Z|XY}(z|x, y)$. A variation of this argument (sans the receiver side information Z) was used in [12, Section III]. It will cease to hold in the sequel, when we allow for relay processing.*

B. A Vertex Packing Characterization

In this subsection, we provide a vertex packing type characterization for the region under discussion. Unfortunately, this characterization is quite cumbersome and not as elegant as its scalar counterpart of Lemma 4.

Define the *vertex packing collection* $\mathbb{VPC}(G)$ to be the set of all polytopes generated by some sub-collection of $\Gamma(G)$ (i.e., as a convex hull of the associated characteristic vectors). The *characteristic matrix* $\mathbf{A}$ of a set $A \subseteq \mathcal{V}$ w.r.t. π is the $|\mathcal{X}| \times |\mathcal{Y}|$-dimensional matrix generated by taking the characteristic vector of A and naturally mapping it to a matrix as dictated by π, where matrix elements not in the associated range are set to one. The set A is said to be *in product form w.r.t.* π if $A = \pi^{-1}(A_1 \times A_2)$ for some $A_1 \subseteq \mathcal{X}$ and $A_2 \subseteq \mathcal{Y}$. Let $\Gamma(G, \pi)$ be the collection of all independent sets of G that are in product form wr.t. π. We define $\mathbb{VP}(G, \pi)$, the *vertex packing polytope of G w.r.t. π*, to be the convex hull of all characteristic matrices pertaining to $\Gamma(G, \pi)$. We write $\mathbf{p}_X$ and $\mathbf{p_Y}$ for the probability column vectors associated with X and Y respectively.

For two sets of column vectors $P_1 \subseteq \mathbb{R}^m$ and $P_2 \subseteq \mathbb{R}^n$, define

$$P_1 * P_2 \stackrel{\text{def}}{=} \{\mathbf{A} \in \mathbb{R}^{m\times n} : \mathbf{A} = \mathbf{a}\mathbf{b}^T, \mathbf{a} \in P_1, \mathbf{b} \in P_2\}.$$

Using Theorem 3, we obtain the following result. The proof is a rather simple extension of the scalar case (see [6]), and is omitted.

Theorem 4. *$H^{(1,2)}(G, V, \pi)$ is given by the set of all rate pairs satisfying*

$$R_1 \geq \min_{\mathbf{a}\in P_1, \mathbf{a}>0} -\mathbf{p}_X^T \cdot \log \mathbf{a}$$
$$R_2 \geq \min_{\mathbf{b}\in P_2, \mathbf{b}>0} -\mathbf{p}_Y^T \cdot \log \mathbf{b}$$

for some $P_1 \in \mathbb{VPC}(\pi^{(1)}(G))$ and $P_2 \in \mathbb{VPC}(\pi^{(2)}(G))$, such that $P_1 * P2 \subseteq \mathbb{VP}(G, \pi)$.

C. Generalized Substitution Lemma

Let (G, V, π) and (F, Q, σ) be a pair of independent probabilistic graphs with associated Cartesian representations. Consider the new triplet $(G_{v\leftarrow F}, V_{v\leftarrow Q}, \pi_{v\leftarrow\sigma})$ obtained via the substitution operation defined in Subsection II-C, where the associated Cartesian representation is defined as $\pi_{v\leftarrow\sigma}(u) = \pi(u)$ for $u \in \mathcal{V} \setminus \{v\}$, and $\pi_{v\leftarrow\sigma}(u) = \sigma(u)$ otherwise. We have the following generalization of the Substitution Lemma.

Lemma 11 (Generalized Substitution Lemma)**.**

$$H^{(1,2)}(G_{v\leftarrow F}, V_{v\leftarrow Q}, \pi_{v\leftarrow\sigma}) = H^{(1,2)}(G, V, \pi) + P_V(v) \cdot H^{(1,2)}(F, Q, \sigma).$$

Proof. See the Appendix. □

VI. An Outer Bound

Recall the definition of $\widetilde{H}^{(1,2)}(G, V, \pi)$ in Theorem 3. The following is an immediate consequence of Theorem 2 and Theorem 3.

Theorem 5. *The following inclusion holds:*

$$H(G, V, \pi) \subseteq \widetilde{H}^{(1,2)}(G, V, \pi) \times \{R : R \geq H(G, V)\}. \quad (3)$$

Specifically,

$$H^{(1,3)}(G, V, \pi) \subseteq \{(R_1, R) : R_1 \geq H(\pi^{(1)}(G), X), R \geq H(G, V)\}. \quad (4)$$

The bound (3) is not tight in general. Let us derive a condition for tightness. Recall that $A \subseteq \mathcal{V}$ is in product form w.r.t. π if $A = \pi^{-1}(A_1 \times A_2)$ for some $A_1 \subseteq \mathcal{X}$ and $A_2 \subseteq \mathcal{Y}$.

Theorem 6. *Suppose that for any non-singleton $A \in \Gamma(G)$ that is in product form w.r.t. π, and for any $a \in A$, the set of all vertices in G that are not adjacent to a is in $\Gamma(G)$. Then the outer bound of Theorem 5 is tight.*

Proof. Loosely speaking, in this case product coloring does not limit the way we can color the whole graph G via coarsening, regardless of the per-coordinate colorings. Precisely, let (U_1, W_1) be some pair satisfying the constraints in Theorem 3, and write $K \stackrel{\text{def}}{=} \pi^{-1}(U_1 \times W_1)$. Let U achieve the maximum in (1), where without loss of generality we can assume that $U - V - K$ forms a Markov chain. The condition in the Theorem implies that any non-singleton $A \in \Gamma(G)$ of product form w.r.t. π has the property that each $a \in A$ has a unique set in $\overline{\Gamma}(G)$ containing it. It is easy to check that this must be the same set for all $a \in A$, hence we can denote it by $m(A)$. Clearly, K is in product form w.r.t. π. Therefore, if $|K| = 1$ then $K = V$, and if $|K| > 1$ then $m(K) = U$, and hence $U - K - V$ forms a Markov chain as well, implying that $I(K; U) = I(V; U) = H(G, V)$. Since $K \in U \in \overline{\Gamma}(G)$, once can clearly achieve $R = H(G, V)$ by random coloring w.r.t. K. □

Example 1. *The outer bound is tight for arbitrary π if G is either empty, complete, or more generally, a complete multipartite graph. The bound is also tight for (G, π) such that G obtained from a complete graph by removing any number of edges that are diagonal w.r.t. π, i.e., edges (v, v') where both $\pi(v)$ and $\pi(v')$ differ in both coordinates.*

VII. $H^{(1,3)}(G, V, \pi)$: Cascade Computing

The projection of the graph entropy region over the first and third (or second and third) coordinates corresponds to the assumption that Y (or X) is known at the relay. This results in the cascade computing setting depicted in Fig. 4. We start by discussing an inner bound based on point-to-point protocols, and then proceed to consider more general protocols. We then study a certain covering problem and utilize it to obtain a better inner bound in a special case.

A. Point-to-Point Protocols

Theorem 7. *The closed convex hull of the union of the following three regions is contained in $H^{(1,3)}(G, V, \pi)$:*

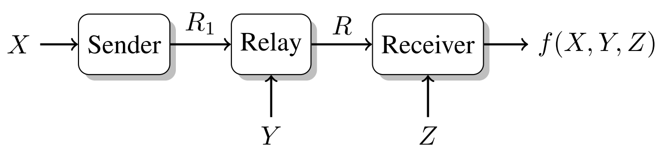

Fig. 4. Cascade computing with side information.

(i) *Decoding relay region:*

$$\{R_1 \geq H(\underline{\pi}^{(1)}(G), X), R \geq H(G, V)\}.$$

(ii) *Forwarding relay region I:*

$$\{R_1 \geq H(\pi_{\perp}^{(1)}(G), X),\\ R \geq H(\pi_{\perp}^{(1)}(G), X) + H(\pi^{(2)}(G), Y)\}.$$

(iii) *Forwarding relay region II:*

$$\{R_1 \geq H(\pi^{(1)}(G, V)),\\ R \geq H(\pi^{(1)}(G), X) + H(\pi_{\perp}^{(2)}(G), Y)\}.$$

Proof. For the purpose of the proof, and by virtue of Corollary 1, we can adopt a distributed source coding with side information setting, i.e., where $f(x, y, z) = (x, y)$, without any loss of generality. Theorem 1 then provides the relation to the marginal graphs. Note that the unrestricted inputs property of the protocol is guaranteed by the unrestricted inputs property of the point-to-point protocols used.

Decoding relay: The sender describes X^n to the relay using $H(G_{X|Y}, X)$ bits per instance. The relay then describes (X^n, Y^n) to the receiver using $H(G_{XY|Z}, (X, Y))$ bits per instance.

Forwarding Relay I: The sender describes X^n to the receiver using $H(G_{X|Z}, X)$ bits per instance. The relay forwards this description, and further describes Y^n to the receiver (that now knows X^n) using $H(G_{Y|XZ}, Y)$ bits per instance. Note that since $f(x, y, z) = (x, y)$, then $G^{f'}_{X|Z} = G_{X|Z}$.

Forwarding Relay II: The sender describes X^n to the receiver using $H(G_{X|YZ}, X)$ bits per instance, assuming the receiver knows Y^n. The relay forwards this description, and further describes Y^n to the receiver using $H(G_{Y|Z}, Y)$ bits per instance. □

Example 2. *Consider the relay-assisted cascade source coding problem, i.e., the cascade computing problem with $f(x, y, z) = x$, and suppose $X = g_1(Y)$ and $Z = g_2(Y)$. Let us compute the decoding relay region, which by Theorem 1 amounts to computing $H(G_{X|Y}, X)$ and $(G^f_{XY|Z}, (X, Y))$. Clearly $G_{X|Y}$ is empty, hence $H(G_{X|Y}, X) = 0$. It is readily verified that $G^f_{XY|Z}$ can be written as a disjoint union of complete multipartite graphs $\{M_z\}_{z\in\mathcal{Z}}$, where the vertices of M_z are essentially $g_2^{-1}(z)$, and the partite sets in M_z correspond to $g_1(g_2^{-1}(z))$. Let ψ map (x, y) to the partite set it belongs to. Now let*

$$U = \bigcup_{z\neq Z} \psi(X_z, Y_z) \cup \psi(X, Y)$$

where $Y_z \sim p_{Y|Z}(\cdot|z)$, $X_z = g_1(Y_z)$ z_i, and all the pairs are independent. This yields $(X, Y) \in U \in \overline{\Gamma}(G^f_{XY|Z})$ and also results in U, Z being independent. Therefore

$$\begin{aligned} H(G^f_{XY|Z}, (X, Y)) &\leq I(X, Y; U)\\ &\overset{(a)}{=} H(U|Z) - H(U|X, Y, Z)\\ &\overset{(b)}{=} H(U|Z) - H(U|X, Z) + H(X|U, Z) + H(Z) - H(Z)\\ &= H(X, U, Z) - H(Z) - H(U|X, Z)\\ &= H(X|Z). \end{aligned}$$

In (a) we used the facts that U and Z are independent and that $Z = g_2(Y)$. In (b) we used the facts that U is a function of X, Z and X is a function of U, Z. The decoding relay region is therefore tight, yielding $\{R_1 \geq 0\,, R \geq H(X|Z)\}$ which is optimal also for general protocols (not only for unrestricted inputs), and even when allowing a vanishing error probability. Now, note that if $G_{X|Z}$ is a full graph, then communicating X directly to the receiver requires a rate of $H(X)$, while when communicating X through the relay that knows Y a possibly much lower sum-rate of $H(X|Z)$ is sufficient. This should be contrasted with the vanishing error probability case where the so-called cutset bound holds, i.e., the relay-receiver rate cannot be smaller than the optimal point-to-point sender-receiver rate [24]. Note also that since the relay knows Z, it seems that a simpler way for zero-error communication would be for the relay to use a conditional codebook (over blocks) which would yield R approaching $H(X|Z)$ as well. However, this latter protocol does not satisfy the unrestricted inputs property.

In some special cases, one of the bounds in Theorem 7 coincides with the outer bound (4) and yields the exact region.

Lemma 12. *The inner bound in Theorem 7 is tight in each of the following cases:*

(i) $\pi^{(1)}(G) = \underline{\pi}^{(1)}(G)$.

(ii) $\pi^{(1)}(G) = \pi_{\perp}^{(1)}(G) = \left(\pi_{\perp}^{(1)}(G^c)\right)^c$ *and* $\pi^{(2)}(G)$ *is empty.*

(iii) $\pi_{\perp}^{(1)}(G) = \left(\pi_{\perp}^{(1)}(G^c)\right)^c$, $\pi^{(2)}(G)$ *is empty, and either* $\pi_{\perp}^{(1)}(G) \subseteq \underline{\pi}^{(1)}(G)$ *or vice versa.*

(iv) $\pi^{(1)}(G) = \left(\pi^{(1)}(G^c)\right)^c$, $\pi_{\perp}^{(2)}(G) = \left(\pi_{\perp}^{(2)}(G^c)\right)^c$, *and* X, Y *are independent.*

Proof. See the Appendix. □

We now describe four cases of cascade computing where the conditions in Lemma 12 are met. The first three are a *relay-assisted cascade source coding* problems, i.e., cascade computing problems for $f(x, y, z) = x$.

Example 3 (Degraded Receiver)**.** *Suppose $X - Y - Z$ forms a Markov chain. Then for relay-assisted cascade source coding, the decoding relay region is tight. To see this, note that if x, x' are adjacent in $G_{X|Y}$ then $\{x, x'\} \subseteq S_{X|Y}(y)$ for some y. Now, $p(x, y, z) = p(x)p(y|x)p(z|y) > 0$ for any $z \in S_{Z|Y}(y)$ and hence $p(x', y, z) > 0$ as well. Therefore, x, x' are adjacent in $G_{X|YZ}$, and so $G_{X|Y} \subseteq G_{X|YZ}$. The reverse inclusion always holds, hence the two graphs coincide. Using the relations of Theorem 1 we find that condition (i) in Lemma 12 is satisfied. We note in passing that under the vanishing error criterion the corresponding region is also known exactly [18]. The gap between the regions can be arbitrarily large, e.g., consider*

$p(x, y, z)$ with full support such that $H(X|Y) \ll H(X)$ and $H(X|Z) \ll H(X)$.

Example 4 *(Degraded Relay). Suppose $X - Z - Y$ forms a Markov chain, and $S_{Y|Z}(z) = S_Y$ for all $z \in S_Z$. Then for relay-assisted cascade source coding, the first forwarding relay region is tight. To see this, let x, x' be adjacent in $G_{X|Z}$. i.e. $\{x, x'\} \in S_{X|Z}(z)$ for some z. Then for any $y, y' \in S_{Y|Z}(z) = S_Y$ we have $p(x, y, z) = p(z)p(x|z)p(y|z) > 0$, and similarly $p(x', y', z) > 0$. Therefore, (x, y) and (x, y') are adjacent in $G_{XY|Z}$ for any $y, y' \in S_Y$. This implies that x, x' are adjacent in the graph $F = \left(\pi_{\perp}^{(1)}(G^c_{XY|Z})\right)^c$, hence $G_{X|Z} \subseteq F$. By Lemma 8 and Theorem 1 we have $F \subseteq G_{X|Z}$, hence $F = G_{X|Z}$. Now, note that our discussion above holds also for $y = y'$, hence also $F = G_{X|YZ}$. Finally, since $f(x, y, z) = x$ we have that $G^f_{Y|XZ}$ is the empty graph. Appealing to the relations in Theorem 1, we find that condition (VII-A) in Lemma 12 is satisfied.*

Example 5. *Suppose $Y - X - Z$ forms a Markov chain, and either $G_{X|Y} \subseteq G_{X|Z}$ or vice versa. Then for relay-assisted cascade source coding, the first forwarding relay region is tight. To see this, let x, x' be adjacent in $G_{X|Z}$, i.e. $\{x, x'\} \in S_{X|Z}(z)$ for some z. Then $p(x, y, z) = p(x)p(y|x)p(z|x) > 0$ for all $y \in S_{Y|X}(x)$ and similarly $p(x', y', z) > 0$ for all $y' \in S_{Y|X}(x')$. Therefore as in the previous example, we have $G_{X|Z} = F$. It is now immediate to verify that condition (VII-A) in Lemma 12 is satisfied.*

Example 6. *Let $Z = (Z_1, Z_2)$, and suppose (X, Z_1) is independent of (Y, Z_2). Let $f(x, y, z) = (g(x, z_1), y)$. Then the second forwarding relay region is tight. This easily follows from condition (VII-A) in Lemma 12.*

Finally, let us motivate further study with the following example, adapted from [25].

Example 7. *Let X_1, X_2 be a pair of independent r.v's, each uniformly distributed over $\{0, \ldots, t-1\}$. Set $X = (X_1, X_2)$ and let $Y = X_B$ where $B \sim$ Bernoulli$(\frac{1}{2})$ is independent of X. Let $Z = X$ and $f(x, y, z) = y$. The graphs $(G_{Y|Z}, Y)$ and $(G_{Y|X}, Y)$ are complete over t vertices, hence their entropy is $\log t$. The graphs $(G^{f'}_{X|Z}, X)$ and $(G_{X|YZ}, X)$ are empty, hence have zero entropy. The graph $(G^f_{XY|Z}, (X, Y))$ is a disjoint union of size 2 cliques, hence its entropy is one bit. The graph $G_{X|Y}$ has t^2 vertices, and maximal degree $4t - 5$, hence its entropy lies between $\log(4t - 5)$ and $2\log t$. The associated point-to-point regions are given by (i) $\{R_1 \geq \log t, R \geq 1\}$; (ii) $\{R_1 \geq 0, R \geq \log t\}$; (iii) $\{R_1 \geq 0, R \geq \Omega(\log t)\}$. The outer bound is given by $\{R_1 \geq 0, R \geq 1\}$. The gap can be arbitrarily large.*

Interestingly, it is the outer bound that gets it (almost) right. Consider the following unrestricted inputs protocol, originally described in [25] for the almost identical league problem in the context of interactive communication. The sender binary represents X_1 and X_2 using $\lceil \log t \rceil$ bits each, and finds the location L of the first bit where they differ, where we set $L = \lceil \log t \rceil + 1$ if $X_1 = X_2$. Now, the sender describes L to the relay, and the relay sends the Lth bit of Y (or an arbitrary bit if $L = \lceil \log t \rceil + 1$) to the receiver, which can now reconstruct Y. This requires the asymptotical rates $R_1 = H(L) \leq 2$ and $R_2 = 1$, independent of t. Hence, the savings over point-to-point protocols can be arbitrarily large.

B. General Protocols

In this subsection we provide an inner bound for $H^{(1,3)}(G, V, \pi)$, which contains, sometimes strictly, the point-to-point bounds of Theorem 7.

Theorem 8. *Let $(X, Y) \stackrel{\text{def}}{=} \pi(V)$. Then $H^{(1,3)}(G, V, \pi)$ contains the closed convex hull of all rate pairs satisfying*

$$\begin{aligned} R_1 &\geq I(X; U) \\ R &\geq I(Y; W|U) + \min\{I(U; W), I(X; U)\} \end{aligned} \tag{5}$$

for some choice of (U, W) such that

(i) $X \in U \in \Gamma(\pi^{(1)}(G))$
(ii) $\pi^{-1}(U \times Y) \in W \in \Gamma(G)$
(iii) *$U - X - Y$ and $X - (U, Y) - W$ form Markov chains.*

Proof. Set $c_2(y^n) = y^n$ throughout the proof. The inner bound is obtained via two protocols:

Protocol 1: Randomly draw $\{U^n(m_1)\}_{m_1=1}^{2^{nR_1}}$ and $\{W^n(m)\}_{m=1}^{2^{nR}}$ as in Theorem 3, according to the marginals p_U and p_W respectively. Let $c_1(x^n)$ be defined as in (2). As before, we have that if $R_1 > I(X; U) + \delta(\varepsilon)$ then $H(c_1(X^n)) \leq nR_1 + o(n)$.

Now, define $c(v^n)$ as follows. If $c_1(x^n) = m_1$, then set $c(v^n) = m$ where m is (say) the smallest index such that $(U^n(m_1), y^n, W^n(m)) \in \mathcal{T}^n_{3\varepsilon}(U, Y, W)$, if such an index exists. Otherwise, set $c(v^n) = v^n$. In the former case we have that $(x^n, U^n(m_1)) \in \mathcal{T}^n_{\varepsilon}(X, U)$, and since Y^n is generated $p_{Y|X}$-i.i.d. given X^n, then by Lemma 1 we have that $(U^n(m_1), Y^n) \in \mathcal{T}^n_{2\varepsilon}(U, Y)$ following the Markov chain $U - X - Y$, with probability $\to 1$ as $n \to \infty$. Therefore, if also $R > I(Y, U; W) + \delta(\varepsilon)$ then by Lemma 2 there exists a suitable m with probability $\to 1$ as $n \to \infty$. We conclude that under the above condition on R, this choice yields $H(c(V^n)) \leq nR + o(n)$. The Markov relation $X - (U, Y) - W$ is not necessary, but can clearly be assumed without loss of generality.

Protocol 2: For *each* $U^n(m_1)$, independently draw a set $\{W^n(m_1, m)\}_{m=1}^{2^{nr}}$ of independent sequences, where $W^n(m_1, m)$ is $P_{W|U}$-independent given $U^n(m_1)$. If $c_1(x^n) = m_1$, set $c(v^n) = (m_1, m)$ where m is (say) the smallest index such that $(U^n(m_1), y^n, W^n(m_1, m)) \in \mathcal{T}^n_{3\varepsilon}(U, Y, W)$, if such an index exists. Otherwise, set $c(v^n) = v^n$. The analysis continues as in Protocol 1, only now Lemma 2 implies that a suitable m exists with probability $\to 1$ as $n \to \infty$, if $r > I(Y; W|U) + \delta(\varepsilon)$. We conclude that under this condition, our choice yields $H(c(V^n)) \leq n(r + R_1) + o(n)$.

The existence of a deterministic protocol achieving the same rate region follows from a standard argument. It is easy to check that $(c_1 \times c_2) \circ \pi$ indeed refines c for both protocols. $\square$

Remark 2. *The region in Theorem 8 contains the point-to-point region of Theorem 7. To obtain the decoding relay region, set U to achieve $H(\underline{\pi}^{(1)}(G), X)$. In this case X is a function of (U, Y), hence we can set W to achieve $H(G, V)$ while satisfying the Markov chain $W - (X, Y) - U$, yielding $I(Y, U; W) = I(X, Y; W) = H(G, V)$. To obtain* *(and possibly exceed) the first forwarding region, set U to achieve $H(\pi_{\perp}^{(1)}(G), X)$, set W' to achieve $H(\pi^{(2)}(G), Y)$ with $(X, U) - Y - W'$ a Markov chain, and then set $W = \pi^{-1}(U \times W')$. To obtain (and possibly exceed) the second relay forwarding region, set U to achieve $H(\pi^{(1)}(G), X)$ and W' to achieve $H(\pi_{\perp}^{(1)}(G), X)$ while satisfying the Markov chain $(X, U) - Y - W'$, and then set $W = \pi^{-1}(U \times W')$. The inclusion can generally be strict, as we now demonstrate.*

Example 7 (continued): *It is clear that the suggested protocol falls under Theorem 8. Specifically, it corresponds to U being a deterministic function of X, and W a deterministic function of (U, Y).*

Let us now provide another example demonstrating that the region in Theorem 8 can be strictly larger than the point-to-point region of Theorem 7, but requiring stochastic mappings.

Example 8. *Let G be the 5-cycle over $\mathcal{V} = \{0, 1, 2, 3, 4\}$ in this order, V uniformly distributed over $\mathcal{V}$. Let $\pi(\mathcal{V}) = \{(0, 0), (1, 1), (0, 1), (2, 0), (1, 2)\}$ respectively. This results in X having a distribution with values $(\frac{2}{5}, \frac{2}{5}, \frac{1}{5})$ over $\mathcal{X} = \{0, 1, 2\}$ respectively, and the same for Y.*

It can be shown that The entropy $H(G, V)$ is obtained by drawing U uniformly at random over the two possible maximal independent sets containing V. This yields $H(G, V) = \log \frac{5}{2} \approx 1.322$. The graph $\pi^{(1)}(G)$ has a single edge $(0, 1)$, and the associated entropy is achieved by letting $U = (X, 2)$ if $X \in \{0, 1\}$, and choosing U uniformly over $\{(0, 2), (1, 2)\}$ if $X = 2$, which yields $H(\pi^{(1)}(G), X) = 1 - \frac{1}{5}H(\frac{1}{2}, \frac{1}{2}) = \frac{4}{5}$. The graph $\underline{\pi}^{(1)}(G)$ has an edge set $\{(0, 1), (0, 2)\}$, hence $H(\underline{\pi}^{(1)}(G), X) = H(\frac{2}{5}, \frac{3}{5}) \approx 0.971$. Finally, $\pi_{\perp}^{(1)}(G)$ is a complete graph, hence $H(\pi_{\perp}^{(1)}(G), X) = H(X) = H(\frac{2}{5}, \frac{2}{5}, \frac{1}{5}) \approx 1.522$. The graph $\pi^{(2)}(G)$ is empty, hence $H(\pi^{(2)}(G), Y) = 0$. The graph $\pi_{\perp}^{(2)}(G)$ has an edge set $\{(0, 1), (0, 2)\}$, hence $H(\pi_{\perp}^{(2)}(G), X) = H(\frac{2}{5}, \frac{3}{5}) \approx 0.971$.

Appealing to Theorem 7, we see that the decoding relay region contains the first forwarding relay region. Taking the convex hull of the decoding relay region and the second forwarding relay region, we obtain the inner bound

$$\begin{aligned} R_1 &\geq 0.8 \\ R &\geq 1.322 \\ R_1 + 0.652 \cdot R &\geq 1.955. \end{aligned} \tag{6}$$

Consider the bound in Theorem 8. Set U to achieve $H(\pi^{(1)}(G), X)$ as above, and restrict $S_W = \{\{0, 3\}, \{1, 3\}, \{2, 4\}\}$. It is easy to check that W is deterministically determined from $\pi^{-1}(U \times Y)$, except when both $U = (1, 2)$ and $Y = 0$. In this latter case, choose W uniformly at random over $\{\{0, 3\}, \{1, 3\}\}$. This results in W having a distribution with values $\{\frac{7}{20}, \frac{1}{4}, \frac{2}{5}\}$ over S_W, respectively. Hence $H(W) \approx 1.559$, and $H(W|U, Y) = 0.1$. Thus, we obtain the following inner bound:

$$\begin{aligned} R_1 &\geq 0.8 \\ R &\geq 1.459 \end{aligned}$$

which contains points strictly outside the region (6).

C. Digression: A Covering Problem

In this subsection we discuss a generic covering problem, which leads to an improved inner bound for a special case presented in the subsequent subsection. Let X, U, W be a triplet of r.v's. over a product alphabet $\mathcal{X} \times \mathcal{U} \times \mathcal{W}$. A set of distinct pairs $S = \{(u^n(t), w^n(t)) \in \mathcal{U}^n \times \mathcal{W}^n\}_{t=1}^T$ is called an (n, ε)-*cover* of X by (U, W) if

$$\mathbb{P}\left(\exists t, (X^n, u^n(t), w^n(t)) \in \mathcal{T}_\varepsilon^n(X, U, W)\right) \geq 1 - \varepsilon.$$

A cover S is associated with a rate pair[11]

$$\begin{aligned} r_1(S) &\stackrel{\text{def}}{=} n^{-1} \log \left|\{u^n(t)\}_{t=1}^T\right| \\ r_2(S) &\stackrel{\text{def}}{=} n^{-1} \log \left|\{w^n(t)\}_{t=1}^T\right|. \end{aligned}$$

A rate pair is called *covering* if for any $\varepsilon > 0$ there exists a (n, ε)-cover of X by (U, W) associated with it, for some large enough n. The *covering rate region* $\mathscr{C}(X|U, W)$ is defined to be the closure of the set of all covering rate pairs.

Problem 1. *Determine* $\mathscr{C}(X|U, W)$.

While we do not know the solution to the problem above, we can derive bounds.

Theorem 9. *$\mathscr{C}(X|U, W)$ contains the closed convex hull of the union of the following regions:*

$$\{(r_1, r_2) : \min(r_1, r_2) \geq I(X; U, W)\}$$

and

$$\begin{aligned} \{(r_1, r_2) : r_1 \geq I(X; U), r_2 \geq I(X; W), \\ r_1 + r_2 \geq I(X; U) + I(X, U; W)\}. \end{aligned}$$

Proof. We pick a random cover in two different ways: 1) Jointly with $r_1 = r_2$ according to p_{UW}, and 2) independently for U and W according to p_U and p_W respectively. By Lemma 2 it is easy to verify that for any $\varepsilon > 0$ these random covers are (n, ε)-covers with probability $\to 1$ as $n \to \infty$ under each of the constraints above. The convex hull is obtained by time-sharing the two strategies. The existence of a deterministic cover achieving the same covering rate region follows from a standard argument. □

The two regions of Theorem 9 do not contain one another in general, as we now exemplify.

Example 9. *Suppose that $X - U - W$ is a Markov chain, $I(X; U|W) > 0$, and $I(X; U) < I(U; W)$. It is easy to show this implies on the one hand that $I(X; W) < I(X; U, W)$ and hence the first region does not include the second, and on the other hand $2I(X; U, W) < I(X; U) + I(X, U; W)$ hence the the second does not include the first. A simple example where these conditions hold is $X \sim$ Bernoulli$(\frac{1}{2})$ and $U = X + Z_1$, $W = U + Z_2$ (mod-2 addition), where $Z_i \sim$ Bernoulli$(\mathrm{p_i})$, $p_2 < p_1 < \frac{1}{2}$, and X, Z_1, Z_2 are independent.*

Remark 3 *(Multiple Descriptions Variant). As we shall see in the following section, the covering region $\mathscr{C}(X|U, W)$ yields $H^{(1,3)}(G, V, \pi)$ for (G, π) of a special structure. It is interesting to note however that the covering region is also related to a variant of the Multiple Descriptions problem [26], described*

[11] Note that we only count the number of distinct elements, hence $r_i(S) < n^{-1} \log T$ is possible.

as follows. A source sequence X^n is to be encoded into two separate descriptions of cardinality 2^{nr_1} and 2^{nr_2} respectively. A receiver observing description $i \in \{0,1\}$ reconstructs a sequence $\widehat{X}^n_{(i)}$ with per symbol mean distortion D_i w.r.t. some distortion function d_i. A receiver observing both descriptions computes both side descriptions $\widehat{X}^n_{(i)}$, and then generates a possibly improved reconstruction by evaluating a per symbol function $\hat{X}_k \stackrel{\text{def}}{=} f(\widehat{X}_{k,(2)}, \widehat{X}_{k,(2)})$, which yields a per symbol mean distortion D w.r.t. some distortion function d. We are interested in the set of all quintuples (r_1, r_2, D_1, D_2, D) that are achievable for some n.

Fixing (D_1, D_2, D), it can be shown that the associated set of achievable pairs is given by the union of all covering regions $\mathscr{C}(X|U, W)$ over the choice of U, W such that $\mathbb{E}d_1(X, g_1(U)) \le D_1$, $\mathbb{E}d_2(X, g_2(V)) \le D_2$, and $\mathbb{E}d(X, g(U, W)) \le D$ for some functions g_1, g_2, g. In particular, Theorem 9 provides an inner bound for that region. Note that the second region in the theorem is very similar to El Gamal - Cover inner bound [26] for the standard Multiple Description problem.

D. Graphs With Singleton Columns

In this subsection, we derive an inner bound for $H^{(1,3)}(G, V, \pi)$ in the case where the Cartesian representation π has *singleton columns*, by which we mean that $\pi(v) = (x, g(x))$ for some function g. This inner bound is then shown to contain rate pairs outside the general inner bound of Theorem 8.

Theorem 10. *Suppose π has singleton columns. Then $H^{(1,3)}(G, V, \pi)$ is the closed convex hull of the union of all covering rate regions of the form $\mathscr{C}(X|U, W)$ where $X \in U \in \Gamma(\pi^{(1)}(G))$ and $\pi^{-1}\{U \times Y\} \in W \in \Gamma(G)$.*

Proof. Proof of $\supseteq$ Inclusion: The protocol is similar to the first protocol in Theorem 8, with the distinction that here the sender can then simulate the relay hence can find both $U^n(m_1)$ and $W^n(m)$ in advance. Precisely, set $c_2(y^n) = y^n$ and let S be an (ε, n) cover for X by U, W satisfying the conditions in the Theorem. For any $t \in \{1, \dots, T\}$, let $m_1(t) \in \{1, \dots, 2^{r_1(S)}\}$ and $m(t) \in \{1, \dots, 2^{r_2(S)}\}$ be the side indices for $u^n(t)$ and $w^n(t)$, respectively. Define

$$c_1(x^n) = \begin{cases} m_1(t) & (x^n, u^n(t), w^n(t)) \in \mathcal{T}^n_\varepsilon(X, U, W) \\ x^n & o.w. \end{cases}$$

where t be the smallest such index, if any. Such a t exists for X^n with probability at least $1 - \varepsilon$. Define further

$$c(v^n) = \begin{cases} m(t') & (u^n(t), y^n, w^n(m(t'))) \in \mathcal{T}^n_\varepsilon(Y, U, W) \\ v^n & o.w. \end{cases}$$

where y^n pertains to the second coordinates of $\pi(v^n)$ and $m(t')$ is the smallest such index, if any. Note that t and t' are not necessarily equal. However, given that t exists, $t' = t$ is an eligible choice since y^n is a function of x^n. Thus, with probability at least $1 - \varepsilon$, $c_1(X^n)$ and $c(V^n)$ take values in alphabets of size $2^{r_1(S)}$ and $2^{r_2(S)}$, respectively. Therefore, the rate pair achievable by this protocol is $R_1 \le r_1 + \varepsilon \log |\mathcal{X}|$ and $R \le r_2 + \varepsilon \log |\mathcal{V}|$. It is easy to check that $(c_1 \times c_2) \circ \pi$ indeed refines c.

Proof of $\subseteq$ Inclusion: Let (R_1, R) be some rate pair in the interior of $H^{(1,3)}(G, V, \pi)$. Let (c_1, c_2, c) be a color cover for (G^n, π^n) such that $H(c_1(X^n)) \ge nR_1$ and $H(c(V^n)) \ge nR$. Since π has singleton columns, we can assume without loss of generality that $V = X$. For any x^n, the set $c_1^{-1}(c_1(x^n))$ (resp. $c^{-1}(c(x^n))$) is contained in some product of n sets in $\Gamma(\pi^{(1)}(G))$ (resp. $\Gamma(G)$). The intersection of all such (respective) product sets is also a product set, which we denote by $\psi^n(x^n)$ (resp. $\lambda^n(x^n)$) i.e., where $\psi_k(x^n) \in \Gamma(\pi^{(1)}(G))$ (resp. $\lambda_k(x^n) \in \Gamma(G)$). Write $x^n \ni x$ if $x_j = x$ for some j. Let $\nu_x(\psi^n(x^n), \lambda^n(x^n))$ be the empirical distribution of the multiset $\{(\psi_k(x^n), \lambda_k(x^n))\}_{k: x_k = x}$, which remains undefined when $x^n \not\ni x$. Define the r.v. pair (U, W) such that

$$p_{UW|X}(\cdot, \cdot|x) \stackrel{\text{def}}{=} \mathbb{E}\left(\nu_x((\psi^n(X^n), \lambda^n(X^n)) \mid X^n \ni x\right)$$

where (X, X^n) is an i.i.d. sequence. By construction, $X \in U_n \in \Gamma(\pi^{(1)}(G))$ and $\pi^{-1}(U_n \times Y) \in \Gamma(G)$.

To conclude the proof, we construct a cover that corresponds to (X, U, W), with the appropriate rates. To that end, consider the k-fold product (c_1^k, c_2^k, c^k) operating on nk-sequences. Trivially, this is a color cover for (G^{nk}, π^{nk}) with $H(c^k(X^{nk})) \ge nkR$ and $H(c_1^k(X^{nk})) \ge nkR_1$. By the asymptotic equipartition property and the above definitions, for any $\varepsilon > 0$ and large enough k (with n fixed), there exists a set $A \subseteq \mathcal{X}^{nk}$ with $P_{X^{nk}}(A) \ge 1 - \varepsilon$, such that $|\psi^{nk}(A)| \le 2^{nk(R+\varepsilon)}$ and $|\lambda^{nk}(A)| \le 2^{nk(R_1+\varepsilon)}$, and where $\left(x^{nk}, \psi^{nk}(x^{nk}), \lambda^{nk}(x^{nk})\right) \in \mathcal{T}^{nk}_\varepsilon(X, U, W)$ for all $x^{nk} \in A$. Therefore, $S \stackrel{\text{def}}{=} \psi^{nk}(A) \times \lambda^{nk}(A)$ is a (nk, ε)-cover with $r_1(S) \le R + \varepsilon$ and $r_2(S) \le R_1 + \varepsilon$.

The next corollary follows from Theorems 9 and 10.

Corollary 3. *Suppose π has singleton columns. Then $H^{(1,3)}(G, V, \pi)$ contains the closed convex hull of all rate pairs satisfying*

$$\begin{aligned} R_1 &\ge I(X; U) \\ R &\ge I(X; W) \\ R_1 + R &\ge I(X; U, W) + I(U; W) \end{aligned} \tag{7}$$

for some choice of (U, W) such that $X \in U \in \Gamma(\pi^{(1)}(G))$ and $\pi^{-1}\{U \times Y\} \in W \in \Gamma(G)$.

Remark 4. *The first rate region in Theorem 9 was not used in Corollary 3 since in this specific case it is already contained in the region of Theorem 8. To see this, let U, W be some pair satisfying the conditions in Theorem 10. For this pair the first region in Theorem 9 is given by*

$$\min(R_1, R) \ge I(X; U, W). \tag{8}$$

Now, let U' be the set of first coordinates of $\pi(W)$. It is readily verified that since π has singleton columns, (U', W) also satisfy the conditions in Theorem 10. Furthermore, U' and W are one-to-one, hence the first region in Theorem 9 for (U', W) yields

$$\min(R_1, R) \ge I(X; W) = I(X; U') \tag{9}$$

which is at least as large as (8). Now note that the pair (U', W) also satisfies the conditions of Theorem 8; plugging it into (5) and using the fact that $I(U'; W) = H(U') \ge I(X; U')$ and $I(Y; W|U) = 0$, reproduces the region (9).

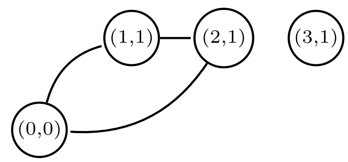

Fig. 5. (G, π) for Example 10.

For π with singleton columns, the bound in Corollary 3 can sometimes improve upon the general bound of Theorem 8. To show this, we need the following Lemma.

Lemma 13. *Suppose π has singleton columns, and let (R_1, R) be contained in the inner bound of Theorem 8. Then $R \geq \underline{H}(G, V)$. Furthermore, suppose that for any $V \in W \in \overline{\Gamma}(G)$ that achieves $H(G, V)$, the mapping $x \mapsto p_{W|X}(\cdot|x)$ is one-to-one. Then $R = H(G, V)$ implies that $R_1 \geq$* $\min(H(G, V), H(\underline{\pi}^{(1)}(G), X))$.

Proof. See the Appendix. □

Let us use the above Lemma to demonstrate that the region in Corollary 3 can indeed contain pairs outside that of Theorem 8.

Example 10. *Let (G, V, π) be described in Fig. 5, where V is uniformly distributed over the vertex set. As a source coding problem this setting corresponds to the case where The first sender is in possession of $X \sim \mathrm{Uniform}\{0, 1, 2, 3\}$, the second sender knows $Y = \min(X, 1)$, the receiver knows $Z = \max(X, 2)$, and would like to learn X.*

It is easy to verify that $H(G, V)$ is achieved by setting $U = \{0, V\}$ if $V \neq 0$, and choosing U uniformly at random over $\overline{\Gamma}(G) = \{\{0, 1\}, \{0, 2\}, \{0, 3\}\}$ otherwise. This yields

$$H(G, V) = H(U) - H(U|V) = \log 3 - \frac{\log 3}{4} = \frac{3 \log 3}{4}.$$

$\underline{\pi}^{(1)}(G)$ is the union of cliques $\{0\} \cup \{1, 2, 3\}$, hence is isomorphic to G. Since X is uniformly distributed, $H(\underline{\pi}^{(1)}(G), X) = H(G, V)$. Using that in Lemma 13, we have that for any point $(R_1, H(G, V))$ within the region of Theorem 8 it must be that $R_1 \geq H(G, V)$. We now proceed to show that the bound in Corollary 3 contains rate pairs $(R_1, H(G, V))$ with $R_1 < H(G, V)$.

Set $U = \{0, X, 3\}$ for $X \in \{1, 2\}$, then distributed over $\{\{0, 1, 3\}, \{0, 2, 3\}, \{0, 3\}\}$ with probabilities $\{\frac{4}{9}, \frac{4}{9}, \frac{1}{9}\}$ respectively for $X = 0$, and chosen uniformly at random over that set for $X = 3$. This choice of parameters nicely yields U that is distributed with probabilities $\{\frac{4}{9}, \frac{4}{9}, \frac{1}{9}\}$ respectively over that same set. Thus:

$$\begin{aligned} I(X; U) &= H\left(\frac{4}{9}, \frac{4}{9}, \frac{1}{9}\right) - \frac{1}{4} H\left(\frac{4}{9}, \frac{4}{9}, \frac{1}{9}\right) - \frac{1}{4} \log 3 \\ &= \frac{5 \log 3}{4} - \frac{4}{3}. \end{aligned}$$

Furthermore, $\pi^{-1}(U \times Y)$ takes values in the set $\{\{0\}, \{3\}, \{1, 3\}, \{2, 3\}\}$ with probabilities $\{\frac{1}{4}, \frac{1}{12}, \frac{1}{3}, \frac{1}{3}\}$. Set $W = \pi^{-1}(U \times Y)$ if $\pi^{-1}(U \times Y) \in \{\{1, 3\}, \{2, 3\}\}$, and $W = \{0, 3\}$ otherwise. This yields $W = \{0, X\}$ if $X \neq 3$, and W uniformly distributed over $\{\{0, 1\}, \{0, 2\}, \{0, 3\}\}$ if $X = 3$. This choice achieves $H(G, V)$, i.e.,

$$I(X; W) = H(G, V) = \frac{3 \log 3}{4}.$$

Furthermore,

$$\begin{aligned} I(X; U, W) + I(U; W) &= I(X; U) + I(X; W) + I(U; W|X) \\ &= I(X; U) + I(X; W) + H(W|X) \\ &= \frac{9 \log 3}{4} - \frac{4}{3}. \end{aligned}$$

Therefore the bound of Corollary 3 contains all rate pairs satisfying

$$\begin{aligned} R_1 &\geq \frac{5 \log 3}{4} - \frac{4}{3} \\ R_2 &\geq \frac{3 \log 3}{4} \\ R_1 + R_2 &\geq \frac{9 \log 3}{4} - \frac{4}{3}. \end{aligned}$$

Specifically, this contains the rate pair $(R_1, H(G, V))$ for $R_1 = \frac{6 \log 3}{4} - \frac{4}{3} < H(G, V)$. which by the previous discussion lies outside the region of Theorem 8.

VIII. $H(G, V, \pi)$: Distributed Computing With Relay Processing

As already discussed, the entire graph entropy region corresponds to the problem of distributed computing with relay processing and side information at the receiver. This setting is depicted in Fig. 1. An outer bound for the region was discussed in Section VI. In this section, we provide two inner bounds in the spirit of Section VII. The derivation is very similar; the main difference is that in order to prove that the sequences $U_1^n(m_1)$ and $U_2^n(m_2)$ that color X^n and Y^n respectively are jointly typical with probability $\to 1$ as $n \to \infty$, we need to use the Markov Lemma [20] in lieu of the conditional typicality Lemma, as in the derivation of the Berger-Tung inner bound in distributed lossy source coding [20], [27], [28]. The details are omitted.

Theorem 11. *$H(G, V, \pi)$ contains the closed convex hull of all rate pairs satisfying*

$$\begin{aligned} R_1 &\geq I(X; U_1) \\ R_2 &\geq I(Y; U_2) \\ R &\geq \min\{I(U_1, U_2; W), I(U_2; W|U_1) + I(X; U_1) \\ &\qquad I(U_1; W|U_2) + I(Y; U_2)\} \end{aligned}$$

for some choice of (U_1, U_2, W) such that

(i) $X \in U_1 \in \Gamma(\pi^{(i)}(G))$.
(ii) $Y \in U_2 \in \Gamma(\pi^{(2)}(G))$.
(iii) $\pi^{-1}(u_1 \times u_2) \in \Gamma(G)$ *whenever* $p_{U_1}(u_1) p_{U_2}(u_2) > 0$.
(iv) $\pi^{-1}(U_1 \times U_2) \in W \in \Gamma(G)$.
(v) $U_1 - X - Y$, $U_2 - Y - (X, U_1)$ *and* $W - (U_1, U_2) - (X, Y)$ *form Markov chains.*

Theorem 12. *Suppose π has singleton columns. Then $H(G, V, \pi)$ contains the closed convex hull of all rate pairs* *satisfying*

$$
\begin{aligned}
R_1 &\geq I(X; U_1)\\
R_2 &\geq I(X; U_2)\\
R_1 + R_2 &\geq I(X; U_1, U_2) + I(U_1; U_2)\\
R_1 + R &\geq I(X; U_1, W) + I(U_1; W)\\
R_2 + R &\geq I(X; U_2, W) + I(U_2; W)\\
R_1 + R_2 + R &\geq I(X; U_1, U_2, W) + I(W; U_1, U_2) + I(U_1; U_2)
\end{aligned}
$$

for some choice of (U_1, U_2, W) *satisfying conditions (i)–(iv) of Theorem 11.*

IX. Further Research

A general characterization of the graph entropy region, and hence the optimal rate region for the associated distributed computing problem, remains an open problem. Furthermore, the various expressions and bounds obtained in this paper are at times cumbersome and not as intuitively appealing as in the one-dimensional case. It remains to be seen if there is a simpler more natural way of approaching these type of problems, or whether the increased complexity is somehow endemic to the setup, as is sometimes the case in other multi-user settings.

The concept of a graph entropy region in itself raises some questions of separate (thought related) interest. For instance, is the entire region additive w.r.t. the OR-product? Does the generalized substitution Lemma apply? We have only established these properties for the projection $H^{(1,2)}$. And, what are the conditions for additivity w.r.t. graph union? This latter question has been well studied in the scalar case, and has revealed fascinating relations to the perfect graph property. Furthermore, it may be interesting to study how the region behaves for a fixed (G, V) as a function of π, and to characterize the induced partial order on the set of Cartesian representations.

The subadditivity of graph entropy w.r.t. to unions has been used as a bounding technique in various problems outside information theory and graph theory, such as counting problems and complexity of algorithms. The underlying idea is to represent the problem in graph covering terms, where the task is to find the minimum number of graphs from a certain class whose union yields a some target graph. The entropy of the target graph and the maximum entropy over the class of graphs then translate into a lower bound on the sought number. It would be interesting to examine whether subadditivity of the graph entropy region w.r.t. unions can be used similarly, possibly in problems where operations are inherently distributed.

Appendix

Proof of Lemma 7: (i) follows since a constant color cover applies. (ii) follows easily since G^n is complete and π^n is onto, hence only one-to-one color covers are possible. (iii) follows since there is a trivial one-to-one mapping between color covers for (G, V, π) and (G, V, π') (via the permutations/their inverses) preserving the chromatic entropy region. (iv) follows since any color cover for (F, V, π) is also a color cover for (G, V, π). (v) follows by noting that the Cartesian product of color covers for (F, V, π) and (G, V, π) yields a color cover for $(F \cup G, V, \pi)$, and then using the subadditivity of entropy. □

Proof of Lemma 8.

(i) Trivial.

(ii) Suppose x, x' are not connected in $\pi^{(i)}(G^c)$. Then for all $y \in \mathcal{Y}$ we have that $\pi^{-1}(x, y)$ and $\pi^{-1}(x', y)$ are not adjacent in G^c, or equivalently, $\pi^{-1}(x, y)$ and $\pi^{-1}(x', y)$ are adjacent in G for all $y \in \mathcal{Y}$. Hence by definition, x and x' are adjacent in $\pi^{(i)}(G)$ and the first relation is established. The second relation follows similarly. □

Proof of Lemma 10. We clearly have the inclusion $\widetilde{H}^{(1,2)}(G, V, \pi) \subseteq n^{-1}\widetilde{H}^{(1,2)}(G^n, V^n, \pi^n)$. To establish the converse inclusion, let (U^n, W^n) be some pair of r.v.'s taking values in $2^{\mathcal{X}^n}$ and $2^{\mathcal{Y}^n}$ respectively (not necessarily i.i.d. component-wise). Then

$$
\begin{aligned}
I(X^n; U^n) &= \sum_{k=1}^{n} H(X_k) - H(X_k|U^n, X_{k-1})\\
&\geq \sum_{k=1}^{n} H(X_k) - H(X_k|U_k) = \sum_{k=1}^{n} I(X_k; U_k).
\end{aligned}
$$

Where we have used the fact that X^n is i.i.d. Similarly, $I(Y^n; W^n) \geq \sum I(Y_k; W_k)$. Therefore the rate pair achieved by U^n, W^n, when normalized by n, is greater (simultaneously on both coordinates) than a convex combination of rate pairs achieved by (U_k, W_k). To conclude we need to show that these rate pairs are all in $\widetilde{H}^{(1,2)}(G, V, \pi)$, i.e., that (U_k, W_k) satisfies the conditions of Theorem 3. The first two conditions are clearly satisfied, since $X^n \in U^n$ and $Y^n \in W^n$. Now, assume that $p_{U_k}(u)p_{W_k}(w) > 0$ for some u, w. Then there exists u^n, w^n with $u_k = u, w_k = w$, such that $p_{U^n}(u^n)p_{W^n}(w^n) > 0$, and hence $(\pi^n)^{-1}(u^n, w^n) \in \Gamma(G^n)$. By the definition of the OR-product, this means that $\pi^{-1}(u_k \times w_k) \in \Gamma(G)$. Hence, (U_k, W_k) satisfies the third condition as well. □

Lemma 11, Proof of $\subseteq$ *Inclusion.* Let $(\overline{X}, \overline{Y}) \stackrel{\text{def}}{=} \pi_{v\leftarrow\sigma}(V_{v\leftarrow Q})$ and $E = \mathbb{1}_{\{V=v\}}$. Furthermore, let $(X, Y) = (\overline{X}, \overline{Y})$ if $E = 0$, and $(X, Y) = \pi(v)$ otherwise. Denote $(x, y) \stackrel{\text{def}}{=} \pi(v)$. Define $(X', Y') = \sigma(Q)$ if $E = 1$, and $(X', Y') = (e, e)$ otherwise, for some unique auxiliary symbol e. Clearly, $\overline{X}$ and $\overline{Y}$ are one-to-one with (X, X', E) and (Y, Y', E) respectively.

Let $(\overline{U}, \overline{W})$ be a pair satisfying the conditions in Theorem 3 for $(G_{v\leftarrow F}, V_{v\leftarrow Q}, \pi_{v\leftarrow\sigma})$. Denote the alphabets for the first coordinate pertaning to π and σ by $\mathcal{X}_\pi$ and $\mathcal{X}_\sigma$ respectively. Let $U = \overline{U}$ if $\overline{U} \cap \mathcal{X}_\sigma = \emptyset$, and $U = (\overline{U} \cap \mathcal{X}_\pi) \cup \{x\}$ otherwise. Let $U' = \overline{U} \cap \mathcal{X}_\sigma$. Define W, W' similarly. Clearly then, $\overline{U}$ and $\overline{W}$ are one-to-one with (U, U') and (W, W') respectively. It is readily verified that (U, W) satisfies the conditions in Theorem 3 for (G, V, π), and that given $V = v$, (U', W') satisfies these conditions for (F, Q, σ). Thus:

$$
\begin{aligned}
&I(\overline{X}; \overline{U})\\
&\quad= H(X, X', E) - H(X, X', E|U, U')\\
&\quad= H(X) + H(E|X) + H(X'|E, X)\\
&\qquad - H(X|U, U') - H(E|X, U, U') - H(X'|E, X, U, U')
\end{aligned}
$$

$$
\begin{aligned}
&= H(X) + H(E|X) + H(X'|E) \\
&\quad - H(X|U) - H(E|X, U, U') - H(X'|E, U') \\
&= I(X; U) + I(X'; U'|E) + I(E; U, U'|X) \\
&= I(X; U) + P_V(v) I(X'; U'|V = v) + I(E; U, U'|X)
\end{aligned} \tag{10}
$$

where we have used the facts that $X - E - X'$, $X - U - U'$ and $X' - (E, U') - (X, U)$ are Markov chains. Similarly, we obtain

$$
\begin{aligned}
I(\overline{Y}; \overline{W}) = I(Y; W) + P_V(v) I(Y'; W'|V = v) \\
+ I(E; W, W'|Y)
\end{aligned} \tag{11}
$$

Noting that $(X', Y') = \sigma(Q)$ given $V = v$, the inclusion follows by taking the union over all feasible $(\overline{U}, \overline{W})$, and by the nonnegativity of the mutual information. □

Lemma 11, Proof of of $\supseteq$ Inclusion: Let (U_1, W_1) and (U_2, W_2) be two pairs satisfying the conditions in Theorem 3 for the (G, V, π) and (F, Q, σ) respectively. Without loss of generality, assume that the triplets (U_1, W_1, V) and (U_2, W_2, Q) are independent, U_1 is independent of W_1, and U_2 is independent of W_2. We now show that we can generate a corresponding $(\overline{U}, \overline{W})$ for which the extra mutual information term vanishes. Let $\overline{U} = U_1 \cup U_2 \setminus \{x\}$ if $x \in U_1$, and $\overline{U} = U_1$ if $x \notin U_1$. Define $\overline{W}$ similarly. It is easily verified that the pair $(\overline{U}, \overline{W})$ satisfies the conditions in Theorem 3 for the $(G_{v \leftarrow F}, V_{v \leftarrow Q}, \pi_{v \leftarrow \sigma})$. Now note that the former derivation of (U, W) and (U', W') yields $(U, W) = (U_1, W_1)$ and $\Pr((U', V') = (U_2, V_2)|V = v) = 1$. Furthermore, it is easily verified that in this case $E - X - (U, U')$ forms a Markov chain, hence the term $I(E; U, U'|X)$ vanishes in (10). Similarly, the term $I(E; W, W'|Y)$ vanishes in (11). The inclusion follows by taking the union over all feasible (U_1, W_1) and (U_2, W_2). □

Proof of Lemma 12.

(i) Trivial, via the decoding relay region.

(ii) The first equality guarantees the optimality of R_1 in the first forwarding relay region. The second equality implies that (x, y) is adjacent to (x', y') in G for $x \neq x'$, if and only if the two associated columns are fully interconnected. i.e., for all feasible y, y'. Since $\pi^{(2)}(G)$ is empty there are no intra-column edges. Therefore, G can be obtained by starting with $\pi_{\perp}^{(1)}(G)$, and substituting each vertex x with an empty graph over $S_{Y|X}(x)$ with a probability distribution $p_{Y|X}(\cdot|x)$. By the Substitution Lemma (Lemma 6) we have that $H(G, V) = H(\pi_{\perp}^{(1)}(G), X)$. This guarantees the optimality of R in the first forwarding relay region.

(iii) Following the discussion in item (ii) above, we have that different columns are either disconnected or fully interconnected, and that there are no intra-column edges. Therefore, if $\pi_{\perp}^{(1)}(G) \subseteq \underline{\pi}^{(1)}(G)$ then $\pi_{\perp}^{(1)}(G) = \pi^{(1)}(G)$, and if $\underline{\pi}^{(1)}(G) \subseteq \pi_{\perp}^{(1)}(G)$ then $\underline{\pi}^{(1)}(G) = \pi^{(1)}(G)$. The first case implies the tightness of the first forwarding relay region as in (ii), while the second implies the tightness of the decoding relay region.

(iv) The first equality implies that all the rows are identical in terms of intra-row edges (independence is needed here to guarantee that there are no missing vertices). The second equality implies that different rows are either disconnected or fully interconnected. Therefore, using the independence again, (G, V) can be constructed by starting with $(\pi_{\perp}^{(2)}(G), Y)$ and substituting each vertex y with the probabilistic graph $(\pi^{(1)}(G), X)$. This yields $H(G, V) = H(\pi^{(1)}(G), X)) + H(\pi_{\perp}^{(2)}(G).Y)$, implying tightness of the second forwarding relay region. □

Proof of Lemma 13. Since π has singleton columns, we can assume without loss of generality that $X = V$. Let (U, W) be some pair satisfying the conditions in Theorem 8. Assume first that $I(U; W) \leq I(X; U)$. The lower bound on R reads

$$
\begin{aligned}
I(Y; W|U) + I(U; W) = I(Y, U; W) \geq I(X; W) \\
= I(V; W) \geq H(G, V)
\end{aligned}
$$

where we have used the Markov chain $X - (U, Y) - W$ for the inequality transition. This inequality is tight if and only if $X - W - (U, Y)$ forms a Markov chain and W is an optimal choice achieving $H(G, V)$. Since $X - (U, Y) - W$ as well, we have that $p_{W|X}(\cdot|x) = p_{W|UY}(\cdot|u, y)$ for all $(x, y, u) \in S_{XYU}$. Since by our assumption $x \mapsto p_{W|X}(\cdot|x)$ is one-to-one, we conclude that X is a function of (U, Y). Therefore, by virtue of Theorem 1 it must be that $X \in U \in \Gamma(\underline{\pi}^{(1)}(G))$. Thus the lower bound on R_1 yields

$$
I(X; U) \geq H(\underline{\pi}^{(1)}(G), X).
$$

Now assume that $I(U; W) \geq I(X; U)$. The lower bound on R reads:

$$
\begin{aligned}
&I(Y; W|U) + I(X; U) \\
&\quad \overset{(a)}{=} I(Y; W|U) + I(X; U) + I(X; W|U, Y) \\
&\quad = I(X; U) + I(XY; W|U) \\
&\quad \overset{(b)}{=} I(X; U, W) = I(X; W) + I(X; U|W) \\
&\quad = I(V; W) + I(X; U|W) \\
&\quad \geq H(G, V)
\end{aligned}
$$

where in (a) we have used the Markov chain $X - (U, Y) - W$, and in (b) the fact that Y is a function of X. The inequality above is tight if and only if W is the optimal choice and $X - W - U$ forms a Markov chain. In this case the bound of R_1 yields

$$
R_1 \geq I(X; U) \geq I(X; W) = H(G, V). \qquad \square
$$

Acknowledgment

This research has emanated from a discussion of a relay-assisted cascade source coding problem under a vanishing error criterion, introduced to the author by Haim Permuter. Useful discussions with Young-Han Kim, Amir Leshem, Alon Orlitsky and Haim Permuter in various stages of this work are greatly appreciated. The author is also thankful to the reviewers whose helpful comments improved the presentation of the paper.

REFERENCES

[1] J. Körner, "Coding of an information source having ambiguous alphabet and the entropy of graphs," in *Proc. 6th Prague Conf Inf. Theory*, 1973, pp. 411–425.

[2] M. Fredman and J. Komlós, "On the size of separating systems and families of perfect hash functions," *SIAM J. Algebraic Discrete Methods*, vol. 5, no. 1, pp. 61–68, 1984.

[3] I. Newman P. Ragde, and A. Wigderson, "Perfect hashing, graph entropy, and circuit complexity," in *Proc. 5th Annu. Struct. Complex. Theory Conf.*, 1990, pp. 91–99.

[4] E. Fachini and J. Körner, "A note on counting very different sequences," *Combinatorics, Probab. Comput.*, vol. 10, no. 6, pp. 501–504, 2001.

[5] J. Kahn and J. H. Han, "Entropy and sorting," in *Proc. 24th Annu. ACM Symp. Theory Comput.*, 1992, pp. 178–187.

[6] G. Simonyi, "Graph entropy: A survey," in *Proc. DIMACS*, vol. 20. 1995, pp. 399–441.

[7] N. Alon and A. Orlitsky "Source coding and graph entropies," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1329–1339, Sep. 1996.

[8] H. Witsenhausen, "The zero-error side information problem and chromatic numbers," *IEEE Trans. Inf. Theory*, vol. 22, no. 5, pp. 592–593, Sep. 1976.

[9] P. Koulgi, E. Tuncel, S. L. Regunathan, and K. Rose, "On zero-error source coding with decoder side information," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 99–111, Jan. 2003.

[10] J. Körner and G. Longo, "Two-step encoding for finite sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 6, pp. 778–782, Nov. 1973.

[11] K. Marton, "On the Shannon capacity of probabilistic graphs," *J. Combinat. Theory, Ser. B*, vol. 57, no. 2, pp. 183–195, Mar. 1993.

[12] P. Koulgi, E. Tuncel, S. L. Regunathan, and K. Rose, "On zero-error coding of correlated sources," *IEEE Trans. Inf. Theory*, vol. 49, no. 11, pp. 2586–2873, Nov. 2003.

[13] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.

[14] P. Cuff, H.-I. Su, and A. El Gamal, "Cascade multiterminal source coding," in *Proc. ISIT*, Jul. 2009, pp. 1199–1203.

[15] V. Doshi, D. Shah, M. Medard, and S. Jaggi, "Distributed functional compression through graph coloring," in *Proc. Data Compress. Conf.*, 2007, pp. 93–102.

[16] V. Doshi, D. Shah, M. Medard, and M. Effros, "Functional compression through graph coloring," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3901–3917, Aug. 2010.

[17] M. Sefidgaran and A. Tchamkerten, "Computing a function of correlated sources: A rate region," in *Proc. ISIT*, Mar. 2011, pp. 1856–1860.

[18] M. Sefidgaran and A. Tchamkerten, "On function computation over a cascade network," in *Proc. IEEE ITW*, Sep. 2012, pp. 472–476.

[19] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 903–917, Mar. 2001.

[20] A. El Gamal and Y. H. Kim, *Lecture Notes on Network Information Theory*. Stanford, CA, USA: Stanford Univ., 2009.

[21] I. Csiszár, J. Körner, L. Lovász, K. Marton, and G. Simonyi, "Entropy splitting for antiblocking corners and perfect graphs," *Combinatorica*, vol. 10, no. 1, pp. 27–40, 1990.

[22] J. Körner, G. Simonyi, and Z. Tuza, "Perfect couples of graphs," *Combinatorica*, vol. 12, no. 2, pp. 179–192, 1992.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

[24] P. Ishwar and S. S. Pradhan, "A relay-assisted distributed source coding problem," in *Proc. Inf. Theory Appl. Workshop*, 2008, pp. 136–141.

[25] A. Orlitsky, "Worst-case interactive communication I: Two messages are almost optimal," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1111–1126, Sep. 1990.

[26] A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 851–857, Nov. 1982.

[27] S.-Y. Tung, *Multiterminal Source Coding*. Ithaca, NY, USA: Cornell Univ., 1978.

[28] T. Berger, *Multiterminal Source Coding*. New York, NY, USA: Springer-Verlag, 1978.

Ofer Shayevitz received the B.Sc. degree (summa cum laude) from the Technion Institute of Technology, Haifa, Israel, in 1997 and the M.Sc. and Ph.D. degrees from the Tel-Aviv University, Tel Aviv, Israel, in 2004 and 2009, respectively, all in electrical engineering.

He is currently a senior lecturer in the department of EE - Systems at the Tel Aviv University. Before joining the department, he was a postdoctoral fellow in the Information Theory and Applications (ITA) Center at the University of California, San Diego (2008-2011), and worked as a quantitative analyst with the D. E. Shaw group in New York (2011-2013). Prior to his graduate studies, he served as an engineer and team leader in the Israeli Defense Forces (1997-2003), and as an algorithms engineer at CellGuide (2003-2004). Dr. Shayevitz is the recipient of the ITA postdoctoral fellowship (2009-2011), the Adams fellowship (2006-2008) awarded by the Israel Academy of Sciences and Humanities, the Advanced Communication Center (ACC) Feder Family award for an outstanding Ph.D. thesis (2009), and the Weinstein prize (2006-2009) for research and publications in signal processing.