

Side Information Design in Zero-Error Coding for Computing

Nicolas Charpenay ^{1,*}, Maël Le Treust ^{2,*}  and Aline Roumy ³ 

¹ Univ. Rennes, CNRS, IRMAR UMR 6625, F-35000 Rennes, France

² Univ. Rennes, CNRS, Inria, IRISA UMR 6074, F-35000 Rennes, France

³ INRIA Rennes, Campus de Beaulieu, F-35000 Rennes cedex, France; aline.roumy@inria.fr

* Correspondence: nicolas.charpenay@univ-rennes.fr (N.C.); mael.le-treust@cnrs.fr (M.L.T.)

Abstract: We investigate the zero-error coding for computing problems with encoder side information. An encoder provides access to a source X and is furnished with side information $g(Y)$. It communicates with a decoder that possesses side information Y and aims to retrieve $f(X, Y)$ with zero probability of error, where f and g are assumed to be deterministic functions. In previous work, we determined a condition that yields an analytic expression for the optimal rate $R^*(g)$; in particular, it covers the case where $P_{X,Y}$ is full support. In this article, we review this result and study the side information design problem, which consists of finding the best trade-offs between the quality of the encoder’s side information $g(Y)$ and $R^*(g)$. We construct two greedy algorithms that give an achievable set of points in the side information design problem, based on partition refining and coarsening. One of them runs in polynomial time.

Keywords: zero-error information theory; source coding; graph theory

1. Introduction

1.1. Zero-Error Coding for Computing

The problem of Figure 1 is a zero-error setting that relates to Orlitsky and Roche’s coding for computing problems from [1]. This coding problem appears in video compression [2,3], where X^n models a set of images known at the encoder. The decoder does not always want to retrieve each whole image. Instead, the decoder receives, for each image $X_t, t \leq n$, a request Y_t to retrieve information $f(X_t, Y_t)$. This information can, for instance, be a detection: cat, dog, car, bike; or a scene recognition: street/city/mountain, etc. The encoder does not know the decoder’s exact request but has prior information about it (e.g., type of request), which is modeled by $(g(Y_t))_{t \leq n}$. This problem also relates to the zero-error Slepian–Wolf open problem, which corresponds to the special case, where g is constant and $f(X, Y) = X$.

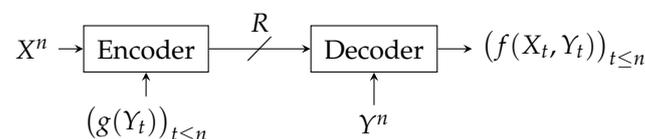


Figure 1. Zero-error coding for computing with side information at the encoder.

Similar schemes to the one depicted in Figure 1 have already been studied, but they differ from the one we are studying in two ways. First, they consider that no side information is available to the encoder. Second, and more importantly, they consider different coding constraints: the lossless case is studied by Orlitsky and Roche in [1], the lossy case by Yamamoto in [4], and the zero-error “unrestricted inputs” case by Shayevitz in [5]. The latter results can be used as bounds for our problem depicted in Figure 1, but do not exactly characterize its optimal rate.



Citation: Charpenay, N.; Le Treust, M.; Roumy, A. Side Information Design in Zero-Error Coding for Computing. *Entropy* **2024**, *26*, 338. <https://doi.org/10.3390/e26040338>

Academic Editor: Igal Sason

Received: 20 December 2023

Revised: 8 April 2024

Accepted: 9 April 2024

Published: 16 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Numerous extensions of the problem depicted in Figure 1 have been studied recently. The distributed context, for instance, has an additional encoder that encodes Y before transmitting it to the decoder. Achievability schemes have been proposed for this setting by Krithivasan and Pradhan in [6] using abelian groups; by Basu et al. in [7] using hypergraphs for the case with maximum distortion criterion; and by Malak and Médard in [8] using hyperplane separations for the continuous lossless case.

Another related context is the network setting, where the function of source random variables from source nodes has to be retrieved at the sink node of a given network. For tree networks, the feasible rate region is characterized by Feizi and Médard in [9] for networks of depth one, and by Sefidgaran and Tchamkerten in [10] under a Markov source distribution hypothesis. In [11], Ravi and Dey consider a bidirectional relay with zero-error “unrestricted inputs” and characterize the rate region for a specific class of functions. In [12], Guang et al. study zero-error function computation on acyclic networks with limited capacities, and give an inner bound based on network cut-sets. For both distributed and network settings, the zero-error coding for computing problems with encoder side information remains open.

In a previous work [13], we determined a condition that we called “pairwise shared side information” such that, if satisfied, the optimal rate $R^*(g)$ has a single-letter expression. This covers many cases of interest, in particular the case where $P_{X,Y}$ is full support for any functions f, g . For the sake of completeness, we review this result. Moreover, we propose an alternative and more interpretable expression for this pairwise shared side information. More precisely, we show that the instances where the “pairwise shared side information” condition is satisfied correspond to the worst possible optimal rates in an auxiliary zero-error Slepian–Wolf problem.

1.2. Encoder’s Side Information Design

In the zero-error coding for computing problems with encoder side information, it can be observed that a “coarse” encoder side information (e.g., if g constant) yields a high optimal rate $R^*(g)$, whereas a “fine” encoder side information (e.g., $g = \text{Id}$) yields a low optimal rate $R^*(g)$. The side information design problem consists of determining the best trade-offs between the optimal rate $R^*(g)$ and the quality of the encoder’s side information, which is measured by its entropy $H(g(Y))$. This expression describes the optimal rate of a zero-error code that transmits the quantized version of Y via the g function. The best trade-offs correspond to the Pareto front of the achievable set, i.e., whose corner-points cannot be obtained by a time sharing between other coding strategies. In short, we aim at determining the Pareto front of the convex hull of the achievable pairs $(H(g(Y)), R^*(g))$.

In this article, we propose a greedy algorithm that gives an achievable set of points in the side information design problem, when $P_{X,Y}$ is full support. Studying our problem with the latter hypothesis is interesting because, unlike the case of the Slepian–Wolf problem, it does not necessarily correspond to a worst-case scenario. Recall indeed, that, when $P_{X,Y}$ is full support, the Slepian–Wolf encoder does not benefit from the side information available at the decoder and needs to send X . In our problem instead, if the retrieval function $f(X, Y) = Y$, since the decoder already has access to Y , no information needs to be sent by the encoder and the optimal rate is 0. Finally, the proposed algorithm relies on our results with “pairwise shared side information”, which gives the optimal rate for all functions g and performs a greedy partition coarsening when choosing the next achievable point. Moreover, it runs in polynomial time.

This paper is organized as follows. In Section 2, we formally present the zero-error coding for computing problems and the encoder’s side information design problem. In Section 3, we give our theoretic results on the zero-error coding for computing problems, including the “pairwise shared side information” condition. In Section 4, we present our greedy algorithms for the encoder’s side information design problem.

2. Formal Presentation of the Problem

We denote sequences by $x^n = (x_1, \dots, x_n)$. The set of probability distributions over \mathcal{X} is denoted by $\Delta(\mathcal{X})$. The distribution of X is denoted by $P_X \in \Delta(\mathcal{X})$ and its support is denoted by $\text{supp } P_X$. Given the sequence length $n \in \mathbb{N}^*$, we denote by $\Delta_n(\mathcal{X}) \subset \Delta(\mathcal{X})$ the set of empirical distributions of sequences from \mathcal{X}^n . We denote by $\{0, 1\}^*$ the set of binary words. The collection of subsets of a set \mathcal{Y} is denoted by $\mathcal{P}(\mathcal{Y})$.

Definition 1. The zero-error source-coding problem of Figure 1 is described by the following:

- Four finite sets $\mathcal{U}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}$ and a source distribution $P_{X,Y} \in \Delta(\mathcal{X} \times \mathcal{Y})$.
- For all $n \in \mathbb{N}^*$, (X^n, Y^n) is the random sequence of n copies of (X, Y) , drawn in an i.i.d. fashion using $P_{X,Y}$.
- Two deterministic functions

$$f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{U}, \tag{1}$$

$$g : \mathcal{Y} \rightarrow \mathcal{Z}. \tag{2}$$

- An encoder that knows X^n and $(g(Y_t))_{t \leq n}$ sends binary strings over a noiseless channel to a decoder that knows Y^n and that wants to retrieve $(f(X_t, Y_t))_{t \leq n}$ without error.

A coding scheme in this setting is described by:

- A time horizon $n \in \mathbb{N}^*$ and an encoding function $\phi_e : \mathcal{X}^n \times \mathcal{Z}^n \rightarrow \{0, 1\}^*$ such that $\text{Im } \phi_e$ is prefix-free.
- A decoding function $\phi_d : \mathcal{Y}^n \times \{0, 1\}^* \rightarrow \mathcal{U}^n$.
- The rate is the average length of the codeword per source symbol, i.e., $R \doteq \frac{1}{n} \mathbb{E}[\ell \circ \phi_e(X^n, (g(Y_t))_{t \leq n})]$, where ℓ denotes the codeword length function.
- n, ϕ_e, ϕ_d must satisfy the zero-error property:

$$\mathbb{P}\left(\phi_d\left(Y^n, \phi_e(X^n, (g(Y_t))_{t \leq n})\right) \neq (f(X_t, Y_t))_{t \leq n}\right) = 0. \tag{3}$$

The minimal rate under the zero-error constraint is defined by

$$R^*(g) \doteq \inf_{\substack{n, \phi_e, \phi_d \\ \text{zero-error}}} \frac{1}{n} \mathbb{E}[\ell \circ \phi_e(X^n, (g(Y_t))_{t \leq n})]. \tag{4}$$

The definition of the Pareto front that we give below is adapted to the encoder’s side information design problem and allows us to describe the best trade-off between the quality of the encoder side information and the rate to compute the function $f(X, Y)$ at the decoder. In other works, the definition of a Pareto front may differ depending on the minimization/maximization problem considered and on the number of variables to be optimized.

Definition 2 (Pareto front). Let $\mathcal{S} \subset \mathbb{R}_+^2$ be a set, the Pareto front of \mathcal{S} is defined by

$$\text{Par}(\mathcal{S}) \doteq \left\{ x \in \mathcal{S} \mid \forall x' \in \mathcal{S} \setminus \{x\}, x'_1 > x_1 \text{ or } x'_2 > x_2 \right\}. \tag{5}$$

Definition 3. The side information design problem in Figure 1 consists of determining the Pareto front of the achievable pairs $(H(g(Y)), R^*(g))$:

$$\mathcal{F} \doteq \text{Par} \left(\text{Conv} \left\{ \left(H(g(Y)), R^*(g) \right) \mid g : \mathcal{Y} \rightarrow \mathcal{Z} \right\} \right), \tag{6}$$

where Conv denotes the convex hull.

In our zero-error setup, all alphabets are finite. Therefore, the Pareto front of the convex hull in (6) is computed on a finite set of points, which correspond to the best trade-offs for the encoder’s side information.

3. Theoretic Results

Determining the optimal rate in the zero-error coding for computing problems, with or without encoder side information, is an open problem. In a previous contribution [13], we determined a condition that, when satisfied, yields an analytic expression for the optimal rate. Interestingly, this condition is general as it does not depend on the function f to be retrieved at the decoder.

3.1. General Case

We first build the characteristic graph $G_{[n]}$, which is a probabilistic graph that captures the zero-error encoding constraints on a given number n of source uses. It differs from the graphs used in [5], as we do not need a cartesian representation of these graphs to study the optimal rates. Furthermore, it has a vertex for each possible realization of $(X^n, (g(Y_t))_{t \leq n})$ known at the encoder, instead of \mathcal{X}^n as in the zero-error Slepian–Wolf problem [14].

Definition 4 (Characteristic graph $G_{[n]}$). *The characteristic graph $G_{[n]}$ is defined by the following:*

- $\mathcal{X}^n \times \mathcal{Z}^n$ as a set of vertices with distribution $P_{X,g(Y)}^n$.
- $(x^n, z^n)(x'^n, z'^n)$ are adjacent if $z^n = z'^n$ and there exists $y^n \in g^{-1}(z^n)$ such that

$$\forall t \leq n, P_{X,Y}(x_t, y_t)P_{X,Y}(x'_t, y_t) > 0, \tag{7}$$

$$\text{and } \exists t \leq n, f(x_t, y_t) \neq f(x'_t, y_t); \tag{8}$$

where $g^{-1}(z^n) = \{y^n \in \mathcal{Y}^n \mid (g(y_t))_{t \leq n} = z^n\}$.

The characteristic graph $G_{[n]}$ is designed with the same core idea as in [15]: (x^n, z^n) and (x'^n, z'^n) are adjacent if there exists a side information symbol y^n compatible with the observation of the encoder (i.e., $z^n = z'^n$ and $y^n \in g^{-1}(z^n)$), such that $f(x^n, y^n) \neq f(x'^n, y^n)$. In order to prevent erroneous decodings, the encoder must map adjacent pairs of sequences to different codewords; hence the use of graph colorings, defined below.

Definition 5 (Coloring, independent subset). *Let $G = (\mathcal{V}, \mathcal{E}, P_V)$ be a probabilistic graph. A subset $\mathcal{S} \subseteq \mathcal{V}$ is independent if $xx' \notin \mathcal{E}$ for all $x, x' \in \mathcal{S}$. Let \mathcal{C} be a finite set (the set of colors), a mapping $c : \mathcal{V} \rightarrow \mathcal{C}$ is a coloring if $c^{-1}(i)$ is an independent subset for all $i \in \mathcal{C}$.*

The chromatic entropy of $G_{[n]}$ gives the best rate of n -shot zero-error encoding functions, as in [14].

Definition 6 (Chromatic entropy H_χ). *The chromatic entropy of a probabilistic graph $G = (\mathcal{V}, \mathcal{E}, P_V)$ is defined by*

$$H_\chi(G) = \inf \{H(c(V)) \mid c \text{ is a coloring of } G\}. \tag{9}$$

Theorem 1 (Optimal rate). *The optimal rate is written as follows:*

$$R^*(g) = \lim_{n \rightarrow \infty} \frac{1}{n} H_\chi(G_{[n]}). \tag{10}$$

Proof. By construction, the following holds: for all encoding functions ϕ_e , ϕ_e is a coloring of $G_{[n]}$ if and only if there exists a decoding function ϕ_d such that (n, ϕ_e, ϕ_d) satisfies the zero-error property. Thus, the best achievable rate is written as follows:

$$R^*(g) = \inf_n \inf_{\phi_e \text{ coloring of } G_{[n]}} H\left(\phi_e\left(X^n, (g(Y_t))_{t \leq n}\right)\right) \tag{11}$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} H_\chi(G_{[n]}). \tag{12}$$

where (12) comes from Fekete’s Lemma and from the definition of H_χ . \square

A general single-letter expression for $R^*(g)$ is missing due to the lack of the intrinsic structure of $G_{[n]}$. In Section 3.2, we introduce a hypothesis that gives structure to $G_{[n]}$ and allows us to derive a single-letter expression for $R^*(g)$.

3.2. Pairwise Shared Side Information

Definition 7. The distribution $P_{X,Y}$ and the function g satisfy the “pairwise shared side information” condition if

$$\forall z \in \mathcal{Z}, \forall x, x' \in \mathcal{X}, \exists y \in g^{-1}(z), P_{XY}(x, y)P_{XY}(x', y) > 0, \tag{13}$$

where $Im(g)$ is the image of the function g . This means that for all z output of g , every pair (x, x') “shares” at least one side information symbol $y \in g^{-1}(z)$.

Note that any full-support distribution $P_{X,Y}$ satisfies the “pairwise shared side information” hypothesis. In Theorem 2, we give an interpretation of the “pairwise shared side information” condition in terms of the optimal rate in an auxiliary zero-error Slepian–Wolf problem.

Theorem 2. The tuple $(P_{X,Y}, g)$ satisfies the condition “pairwise shared side information” (13) $\iff R^*(g) = H(X|g(Y))$ in the case $f(X, Y) = X$, and for all $z \in \mathcal{Z}$, $P_{X|g(Y)=z}$ is full support.

The proof of Theorem 2 is given in Appendix A.1.

Definition 8 (AND, OR product). Let $G_1 = (\mathcal{V}_1, \mathcal{E}_1, P_{V_1})$, $G_2 = (\mathcal{V}_2, \mathcal{E}_2, P_{V_2})$ be two probabilistic graphs; their AND (resp. OR) product denoted by $G_1 \wedge G_2$ (resp. $G_1 \vee G_2$) is defined by the following: $\mathcal{V}_1 \times \mathcal{V}_2$ as a set of vertices, $P_{V_1}P_{V_2}$ as probability distribution on the vertices, and $(v_1v_2), (v'_1v'_2)$ are adjacent if

$$v_1v'_1 \in \mathcal{E}_1 \text{ AND } v_2v'_2 \in \mathcal{E}_2, \tag{14}$$

resp. $(v_1v'_1 \in \mathcal{E}_1 \text{ and } v_1 \neq v'_1) \text{ OR } (v_2v'_2 \in \mathcal{E}_2 \text{ and } v_2 \neq v'_2)$;

with the convention that all vertices are self-adjacent. We denote by $G_1^{\wedge n}$ (resp. $G_1^{\vee n}$) the n -th AND (resp. OR) power.

AND and OR powers significantly differ in terms of existing single-letter expression for the associated asymptotic chromatic entropy. Indeed, in the zero-error Slepian–Wolf problem in [14], the optimal rate $\lim_{n \rightarrow \infty} \frac{1}{n} H_\chi(G^{\wedge n})$, which relies on an AND power, does not have a single-letter expression. Instead, closed-form expressions for OR powers of graphs exist. More precisely, as recalled in Proposition 1, $\lim_{n \rightarrow \infty} \frac{1}{n} H_\chi(G^{\vee n})$ admits a single-letter expression called the Körner graph entropy, introduced in [16], and defined below. This observation is key for us to derive a single-letter expression for our problem. More precisely, by using a convex combination of Körner graph entropies, we provide a single-letter expression in Theorem 3 for the optimal rate $R^*(g)$.

Definition 9 (Körner graph entropy H_κ). For all $G = (\mathcal{V}, \mathcal{E}, P_V)$, let $\Gamma(G)$ be the collection of independent sets of vertices in G . The Körner graph entropy of G is defined by

$$H_\kappa(G) = \min_{W \in \Gamma(G)} I(W; V), \tag{15}$$

where the minimum is taken over all distributions $P_{W|V} \in \Delta(\mathcal{W})^\mathcal{V}$, with $\mathcal{W} = \Gamma(G)$ and the constraint that the random vertex V belongs to the random set W with probability one.

Below, we recall that the limit of the normalized chromatic entropy of the OR product of graphs admits a closed-form expression given by the Körner graph entropy H_κ . Moreover, the Körner graph entropy of OR products of graphs is simply the sum of the individual Körner graph entropies.

Proposition 1 (Properties of H_κ). Theorem 5 in [14] for all probabilistic graphs G and G' ,

$$H_\kappa(G) = \lim_{n \rightarrow \infty} \frac{1}{n} H_\kappa(G^{\vee n}), \tag{16}$$

$$H_\kappa(G \vee G') = H_\kappa(G) + H_\kappa(G'). \tag{17}$$

Definition 10 (Auxiliary graph G_z^f). For all $z \in \mathcal{Z}$, we define the auxiliary graph G_z^f by

- \mathcal{X} as set of vertices with distribution $P_{X|g(Y)=z}$;
- xx' are adjacent if $f(x, y) \neq f(x', y)$ for some $y \in g^{-1}(z) \cap \text{supp } P_{Y|X=x} \cap \text{supp } P_{Y|X=x'}$.

Theorem 3 (Pairwise shared side information). If $P_{X,Y}$ and g satisfy (13), the optimal rate is written as follows:

$$R^*(g) = \sum_{z \in \mathcal{Z}} P_{g(Y)}(z) H_\kappa(G_z^f). \tag{18}$$

The proof is in Appendix A.2, the keypoint is the particular structure of $G_{[n]}$: a disjointed union of OR products.

Remark 1. The “pairwise shared side information” assumption (13) implies that the adjacency condition (7) is satisfied, which makes $G_{[n]}$ a disjoint union of OR products. Moreover, Körner graph entropies appear in the final expression for $R^*(g)$, even if $G_{[n]}$ is not an n -th OR power.

Now, consider the case where $P_{X,Y}$ is full support. This is a sufficient condition to have (13). The optimal rate in this setting is derived from Theorem 3, which leads to the analytic expression in Theorem 4.

Theorem 4 (Optimal rate when $P_{X,Y}$ is full support). When $P_{X,Y}$ is full support, the optimal rate is written as follows:

$$R^*(g) = H(j(X, g(Y)) | g(Y)), \tag{19}$$

where the function j returns a word in \mathcal{U}^* , defined by

$$j : \mathcal{X} \times \mathcal{Z} \rightarrow \mathcal{U}^* \tag{20}$$

$$(x, z) \mapsto (f(x, y'))_{y' \in g^{-1}(z)}.$$

Proof. By Theorem 3, $R^*(g) = \sum_{z \in \mathcal{Z}} P_{g(Y)}(z) H_\kappa(G_z^f)$. It can be shown that G_z^f is complete multipartite for all z as $P_{X,Y}$ is full support; and it satisfies $H_\kappa(G_z^f) = H(j(X, g(Y)) | g(Y) = z)$. \square

3.3. Example

In this example, the “pairwise shared side information” assumption is satisfied and $R^*(g)$ is strictly less than a conditional Huffman coding of X knowing $g(Y)$; and also strictly less than the optimal rate without exploiting $g(Y)$ at the encoder.

Consider the probability distribution and function outcomes depicted in Figure 2, with $\mathcal{U} = \{a, b, c\}$, $\mathcal{X} = \{0, \dots, 3\}$, $\mathcal{Y} = \{0, \dots, 7\}$, and $\mathcal{Z} = \{0, 1\}$. Let us show that the “pairwise shared side information” assumption is satisfied. The source symbols $0, 1, 2 \in \mathcal{X}$ share the side information symbol 0 (resp. 5) when $g(Y) = 0$ (resp. $g(Y) = 1$). The source symbol $3 \in \mathcal{X}$ shares the side information symbols 1, 2, 3 with the source symbols 0, 1, 2, respectively, when $g(Y) = 0$, and the source symbol 3 shares the side information symbol 5 with all other source symbols when $g(Y) = 1$.

		Y							
		0	1	2	3	4	5	6	7
		$g(Y) = 0$				$g(Y) = 1$			
X	0	0.1	0.05	*	*	0.05	0.05	*	*
	1	0.1	*	0.05	*	0.05	0.05	0.05	*
	2	0.1	*	*	0.05	*	0.05	*	*
	3	*	0.05	0.05	0.05	*	0.05	*	0.05

		Y							
		0	1	2	3	4	5	6	7
		$g(Y) = 0$				$g(Y) = 1$			
X	0	a	b	*	*	b	a	*	*
	1	a	*	b	*	a	a	b	*
	2	b	*	*	c	*	b	*	*
	3	*	c	c	c	*	c	*	c

Figure 2. An example of $P_{X,Y}$ and g that satisfies (13), along with the outcomes $f(X, Y)$. The elements outside $\text{supp } P_{X,Y}$ are denoted by $*$.

Since the “pairwise shared side information” assumption is satisfied, we can use Theorem 3; the optimal rate is written as follows:

$$R^*(g) = P_{g(Y)=0} H_\kappa(G_0^f) + P_{g(Y)=1} H_\kappa(G_1^f). \tag{21}$$

First, we need to determine the probabilistic graphs G_0^f and G_1^f . In G_0^f , the vertex 0 is adjacent to 2 and 3, as $f(0,0) \neq f(2,0)$ and $f(0,1) \neq f(3,1)$. The vertex 1 is also adjacent to 2 and 3 as $f(1,0) \neq f(2,0)$ and $f(1,2) \neq f(3,2)$. Furthermore $P_{X|g(Y)=0}$ is uniform, hence $G_0^f = (C_4, \text{Unif}(\mathcal{X}))$ where C_4 is the cycle graph with 4 vertices.

In G_1^f , the vertices 1, 2, 3 are pairwise adjacent as $f(1,5), f(2,5)$ and $f(3,5)$ are pairwise different; and 0 is adjacent to 1, 2, and 3 because of the different function outputs generated by $Y = 4$ and $Y = 5$. Thus, $G_1^f = (K_4, P_{X|g(Y)=1})$ with $P_{X|g(Y)=1} = (\frac{1}{4}, \frac{3}{8}, \frac{1}{8}, \frac{1}{4})$, and K_4 is the complete graph with 4 vertices.

Now, let us determine $H_\kappa(G_0^f)$ and $H_\kappa(G_1^f)$. On the one hand,

$$H_\kappa(G_0^f) = H(V_0) - \max_{V_0 \in W \in \Gamma(G_0^f)} H(V_0|W) \tag{22}$$

$$= 2 - 1 = 1, \tag{23}$$

with $V_0 \sim P_{X|g(Y)=0} = \text{Unif}(\mathcal{X})$; and where $H(V_0|W)$ in (22) is maximized by taking $W = \{0, 1\}$ when $V \in \{0, 1\}$, and $W = \{2, 3\}$ otherwise.

On the other hand,

$$H_\kappa(G_1^f) = \min_{V_1 \in W \in \Gamma(G_1^f)} I(W; V_1) \tag{24}$$

$$= H(V_1) \approx 1.906, \tag{25}$$

with $V_1 \sim P_{X|g(Y)=1}$; where (25) follows from $\Gamma(G_1^f) = \{\{0\}, \dots, \{3\}\}$, as G_1^f is complete. Hence, $R^*(g) \approx 1.362$.

The rate that we would obtain by transmitting X knowing $g(Y)$ at both encoder and decoder with a conditional Huffman algorithm is written as $R_{\text{Huff}} = H(X|g(Y)) \approx 1.962$.

The rate that we would obtain without exploiting $g(Y)$ at the encoder is $R_{\text{No } g} = H(X) \approx 1.985$ because of the different function outputs generated by $Y = 4$ and $Y = 5$.

Finally, $H(f(X, Y)|Y) \approx 0.875$.

In this example, we have

$$H(X) = R_{\text{No } g} > R_{\text{Huff}} > R^*(g) > H(f(X, Y)|Y). \tag{26}$$

This illustrates the impact of the side information at the encoder in this setting, as we can observe a large gap between the optimal rate $R^*(g)$ and $R_{\text{No } g}$.

4. Optimization of the Encoder Side Information

4.1. Preliminary Results on Partitions

In order to optimize the function g in the encoder side information, we propose a new equivalent characterization of the function g in the form of a partition of the set \mathcal{Y} . The equivalence is shown in Proposition 2 below.

Proposition 2. For all $g : \mathcal{Y} \rightarrow \mathcal{Z}$, the collection of subsets $(g^{-1}(z))_{z \in \mathcal{Z}}$ is a partition of \mathcal{Y} .

Conversely, if $\mathcal{A} \subset \mathcal{P}(\mathcal{Y})$ is a partition of \mathcal{Y} , then there exists a mapping $g_{\mathcal{A}} : \mathcal{Y} \rightarrow \mathcal{Z}$ such that $\forall z \in \text{Im } g_{\mathcal{A}}, \exists A_z \in \mathcal{A}, A_z = g_{\mathcal{A}}^{-1}(z)$.

Proof. The direct part results directly from the fact that g is a function. For the converse part, we take \mathcal{Z} such that $|\mathcal{Z}| = |\mathcal{A}|$ and we define $g_{\mathcal{A}} : \mathcal{Y} \rightarrow \mathcal{Z}$ by $g_{\mathcal{A}}(y) = z$, where $z \in \mathcal{Z}$ is the unique index such that $y \in A_z$. The property $\forall z \in \text{Im } g_{\mathcal{A}}, \exists A_z \in \mathcal{A}, A_z = g_{\mathcal{A}}^{-1}(z)$ is therefore satisfied. \square

Now, let us define coarser and finer partitions, with the corresponding notions of merging and splitting. These operations on partitions are the core idea of our greedy algorithms; as illustrated in Proposition 2, the partitions of \mathcal{Y} correspond to functions $g : \mathcal{Y} \rightarrow \mathcal{Z}$ for the encoder’s side information. Therefore, obtaining a partition from another means finding another function $g : \mathcal{Y} \rightarrow \mathcal{Z}$ for the encoder’s side information.

Definition 11 (Coarser, Finer). Let $\mathcal{A}, \mathcal{B} \subset \mathcal{P}(\mathcal{Y})$ be two partitions of the finite set \mathcal{Y} . We say that \mathcal{A} is coarser than \mathcal{B} if

$$\forall B \in \mathcal{B}, \exists A \in \mathcal{A}, B \subset A. \tag{27}$$

If so, we also say that \mathcal{B} is finer than \mathcal{A} .

Example 1. Let $\mathcal{Y} = \{1, 2, 3, 4\}$, the partition $\mathcal{A} = \{\{1\}, \{2, 3, 4\}\}$ is coarser than $\mathcal{B} = \{\{1\}, \{2\}, \{3, 4\}\}$.

Definition 12 (Merging, Splitting). A merging is an operation that maps a partition $\mathcal{A} = \{A_1, \dots, A_i, \dots, A_j, \dots, A_m\}$ to the partition $\mathcal{A}' = \{A_1, \dots, A_i \cup A_j, \dots, A_m\}$. A splitting is an operation that maps a partition $\mathcal{A} = \{A_1, \dots, A_i, \dots, A_m\}$ to the partition $\mathcal{A}' = \{A_1, \dots, A_i^{(1)}, A_i^{(2)}, \dots, A_m\}$, where $\{A_i^{(1)}, A_i^{(2)}\}$ form a partition of the subset A_i .

We also define the set of partitions $\text{Merge}(\mathcal{A})$ (resp. $\text{Split}(\mathcal{A})$), which correspond to all partitions that can be obtained with a merging (resp. splitting) of \mathcal{A} :

$$\text{Merge}(\mathcal{A}) \doteq \left\{ m(\mathcal{A}) \mid m \text{ is a merging} \right\}; \tag{28}$$

$$\text{Split}(\mathcal{A}) \doteq \left\{ s(\mathcal{A}) \mid s \text{ is a splitting} \right\}. \tag{29}$$

Proposition 3. *If \mathcal{A} is coarser (resp. finer) than \mathcal{B} , then \mathcal{A} can be obtained from \mathcal{B} by performing a finite number of mergings (resp. splittings).*

4.2. Greedy Algorithms Based on Partition Coarsening and Refining

In this Section, we assume $P_{X,Y}$ to be full support.

With Proposition 2, we know that determining the Pareto front by a brute force approach would at least require to enumerate the partitions of \mathcal{Y} . Therefore, the complexity of this approach is exponential in $|\mathcal{Y}|$. In the following we describe the greedy Algorithms 1 and 2 that give an achievable set for the encoder’s side information design problem; one of them has a polynomial complexity. Then we give an example where the Pareto front coincides with the boundary of the convex hull of the achievable rate region obtained by both greedy algorithms.

Algorithm 1 Greedy coarsening algorithm

- 1: $\mathcal{A} \leftarrow \{\{1\}, \dots, \{|\mathcal{Y}|\}\}$ // \mathcal{A} starts by being the finest partition of \mathcal{Y} , i.e., $g_{\mathcal{A}} = \text{Id}$.
 - 2: $\text{Front} \leftarrow [\mathcal{A}, \text{ndef}, \dots, \text{ndef}]$ // Will contain the list of the $|\mathcal{Y}|$ partitions chosen during the execution
 - 3:
 - 4: **for** $i \in \{1, \dots, |\mathcal{Y}| - 1\}$ **do**
 - 5: // Maximize over \mathcal{B} merging of \mathcal{A} the slope between $(H(g_{\mathcal{B}}(Y)), R^*(g_{\mathcal{B}}))$ and $(H(g_{\mathcal{A}}(Y)), R^*(g_{\mathcal{A}}))$.
 - 6: $\mathcal{A} \leftarrow \operatorname{argmax}_{\mathcal{B} \in \text{Merge}(\mathcal{A})} \frac{R^*(g_{\mathcal{B}}) - R^*(g_{\mathcal{A}})}{H(g_{\mathcal{B}}(Y)) - H(g_{\mathcal{A}}(Y))}$
 - 7: $\text{Front}[i] \leftarrow \mathcal{A}$
 - 8:
 - 9: **return** Front // $\mathcal{A} = \{\{1, \dots, |\mathcal{Y}|\}\}$ at this point
-

Algorithm 2 Greedy refining algorithm

- 1: $\mathcal{A} \leftarrow \{\{1, \dots, |\mathcal{Y}|\}\}$ // \mathcal{A} starts by being the coarsest partition of \mathcal{Y} , i.e., $g_{\mathcal{A}} = \text{Id}$.
 - 2: $\text{Front} \leftarrow [\mathcal{A}, \text{ndef}, \dots, \text{ndef}]$ // Will contain the list of the $|\mathcal{Y}|$ partitions chosen during the execution
 - 3:
 - 4: **for** $i \in \{1, \dots, |\mathcal{Y}| - 1\}$ **do**
 - 5: // Minimize over \mathcal{B} splitting of \mathcal{A} the slope between $(H(g_{\mathcal{A}}(Y)), R^*(g_{\mathcal{A}}))$ and $(H(g_{\mathcal{B}}(Y)), R^*(g_{\mathcal{B}}))$.
 - 6: $\mathcal{A} \leftarrow \operatorname{argmin}_{\mathcal{B} \in \text{Split}(\mathcal{A})} \frac{R^*(g_{\mathcal{B}}) - R^*(g_{\mathcal{A}})}{H(g_{\mathcal{B}}(Y)) - H(g_{\mathcal{A}}(Y))}$
 - 7: $\text{Front}[i] \leftarrow \mathcal{A}$
 - 8:
 - 9: **return** Front // $\mathcal{A} = \{\{1\}, \dots, \{|\mathcal{Y}|\}\}$ at this point
-

In these a argmin (resp. argmax) means any minimizer (resp. maximizer) of the specified quantity; and the function $g_{\mathcal{A}} : \mathcal{Y} \rightarrow \mathcal{Z}$ is a function for the encoder’s side information corresponding to the partition \mathcal{A} , whose existence is given by Proposition 2.

The coarsening (resp. refining) algorithm starts by computing its first achievable point $(H(g_{\mathcal{A}}(Y)), R^*(g_{\mathcal{A}}))$ with \mathcal{A} being the finest (resp. coarsest) partition: it evaluates $R^*(g_{\mathcal{A}})$, with $g_{\mathcal{A}} = \text{Id}$ (resp. $g_{\mathcal{A}}$ constant); and $H(g_{\mathcal{A}}(Y)) = H(Y)$ (resp. $H(g_{\mathcal{A}}(Y)) = 0$).

Then, at each iteration, the next achievable point will be computed by using a merging (resp. splitting) of the current partition \mathcal{A} . The next partition will be a coarser (resp. finer) partition chosen from $\text{Merge}(\mathcal{A})$ (resp. $\text{Split}(\mathcal{A})$), following a greedy approach. Since we want to achieve good trade-offs between $H(g_{\mathcal{A}}(Y))$ and $R^*(g_{\mathcal{A}})$, we want to decrease $H(g(Y))$ (resp. $R^*(g_{\mathcal{A}})$) as much as possible while increasing the other quantity as less as possible. We do so by maximizing over $\mathcal{B} \in \text{Merge}(\mathcal{A})$ the negative ratio

$$\frac{R^*(g_{\mathcal{B}}) - R^*(g_{\mathcal{A}})}{H(g_{\mathcal{B}}(Y)) - H(g_{\mathcal{A}}(Y))}, \tag{30}$$

resp. minimizing over $\mathcal{B} \in \text{Split}(\mathcal{A})$ the negative ratio

$$\frac{R^*(g_{\mathcal{B}}) - R^*(g_{\mathcal{A}})}{H(g_{\mathcal{B}}(Y)) - H(g_{\mathcal{A}}(Y))}, \tag{31}$$

hence the use of slope maximization (resp. minimization) in the algorithm. At the end, the set of achievable points computed by the algorithm is returned.

In Figure 3, we show rate pairs associated with all possible partitions of \mathcal{Y} : a point corresponds to a partition of \mathcal{Y} , its position gives the associated rates $(H(g(Y)), R^*(g))$. Two points are linked if their corresponding partitions \mathcal{A}, \mathcal{B} satisfy $\mathcal{A} \in \text{Merge}(\mathcal{B})$ or $\mathcal{A} \in \text{Split}(\mathcal{B})$. The obtained graph is the Hasse diagram for the partial order “coarser than”. Note that due to symmetries in the chosen example, several points associated with different partitions may overlap. In Figure 4, (resp. Figure 5), we give an illustration of the trajectory of the greedy coarsening (resp. refining) algorithm.

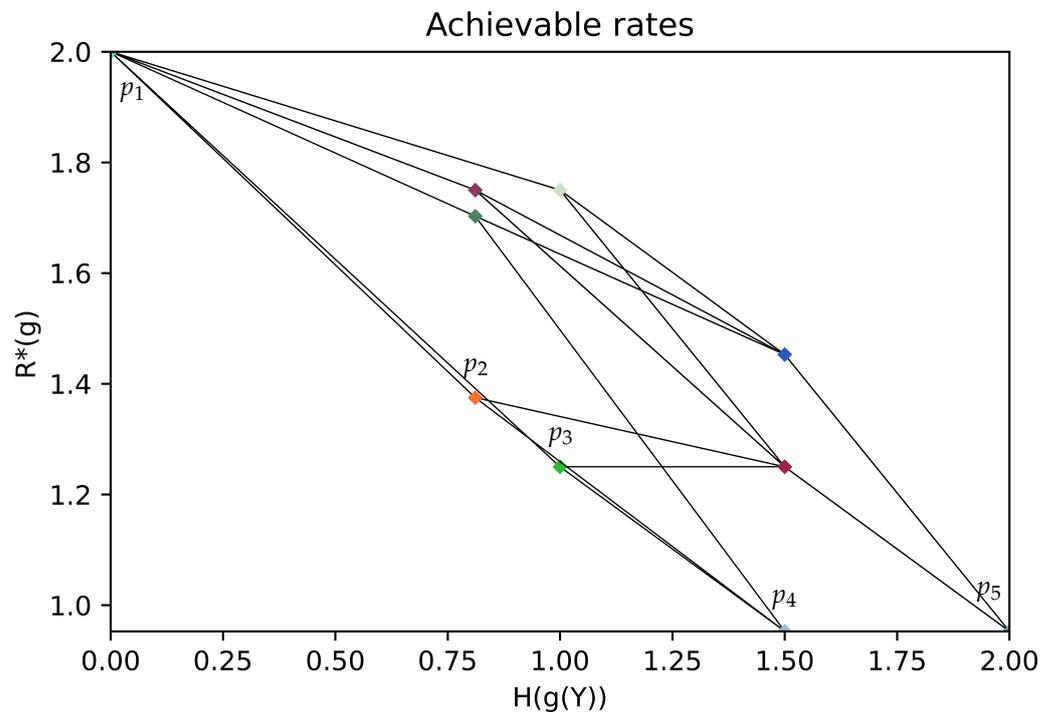


Figure 3. An illustration of the rate pairs associated with all partitions of \mathcal{Y} . The Pareto front is the broken line corresponding to the partitions $p_1 - p_2 - p_3 - p_4 - p_5$; with $p_1 = \{\{1, 2, 3, 4\}\}$, $p_2 = \{\{1, 2, 4\}, \{3\}\}$, $p_3 = \{\{1, 2\}, \{3, 4\}\}$, $p_4 = \{\{1, 2\}, \{3\}, \{4\}\}$, $p_5 = \{\{1\}, \{2\}, \{3\}, \{4\}\}$.

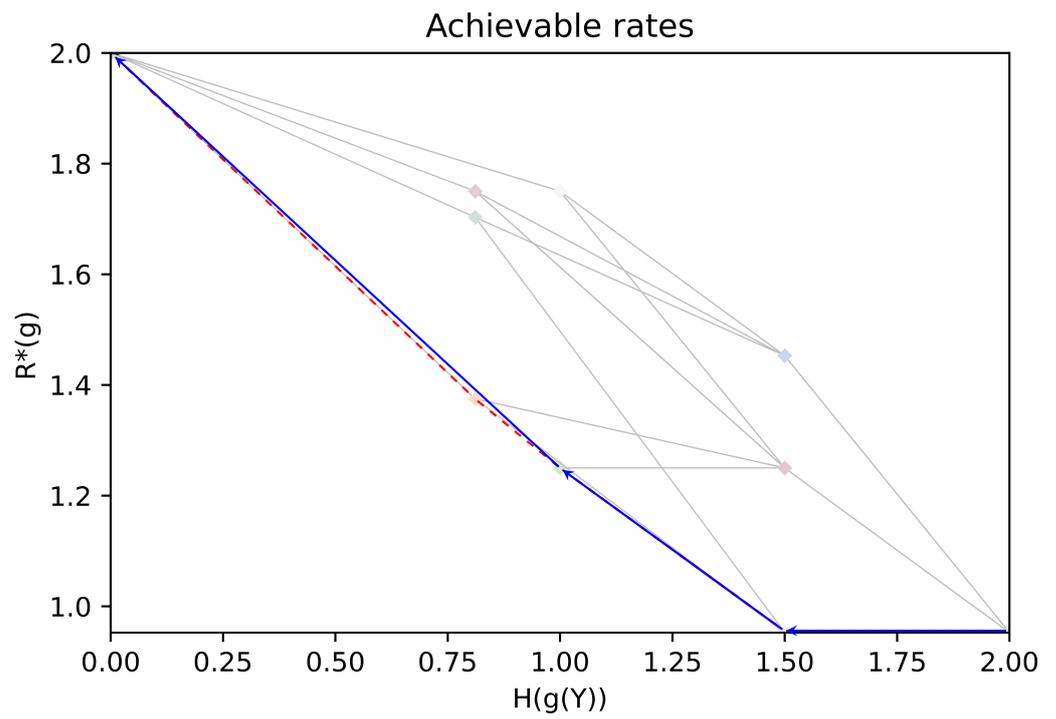


Figure 4. An illustration of the trajectory of the coarsening greedy algorithm (blue), with the Pareto front of the achievable rates (dashed red).

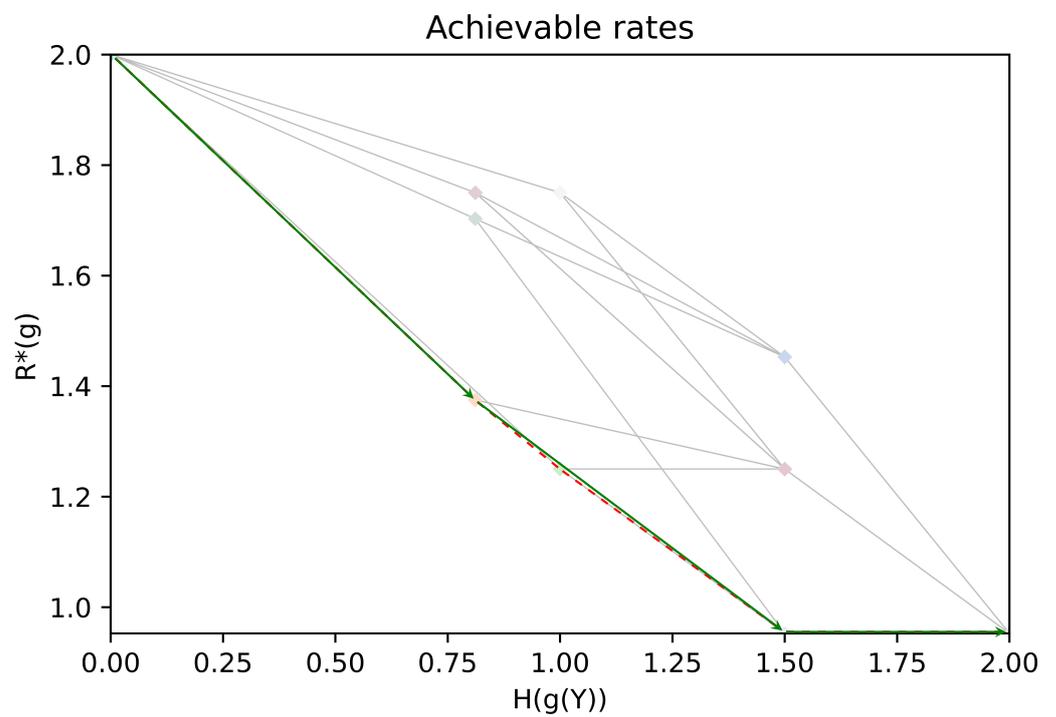


Figure 5. An illustration of the trajectory of the refining greedy algorithm (green), with the Pareto front of the achievable rates (dashed red).

Figures 3–5 are obtained with the following problem data:

$$P_{X,Y} = \text{Unif}(\mathcal{X} \times \mathcal{Y}) \quad f(\cdot, \cdot) = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}. \tag{32}$$

As stated in Theorem 5, the complexity of the coarsening greedy algorithm is polynomial since $|\text{Merge}(\mathcal{A})|$ is quadratic in $|\mathcal{Y}|$ and the evaluation of $R^*(g)$ can be conducted in polynomial time. This polynomial complexity property is not satisfied by the refining greedy algorithm, as $|\text{Split}(\mathcal{A})|$ is exponential in $|\mathcal{Y}|$.

Theorem 5. *The coarsening greedy algorithm runs in polynomial time in $|\mathcal{Y}|$. The refining greedy algorithm runs in exponential time in $|\mathcal{Y}|$.*

Proof. The number of points evaluated by the coarsening (resp. refining) greedy algorithm is $O(|\mathcal{Y}|^3)$ (resp. $O(2^{|\mathcal{Y}|})$): $O(|\mathcal{Y}|)$ mergings (resp. splittings) are made; and for each of these mergings, all points from $\text{Merge}(\mathcal{A})$ (resp. $\text{Split}(\mathcal{A})$) are evaluated; they are, at most, $\binom{|\mathcal{Y}|}{2} = O(|\mathcal{Y}|^2)$ (resp. $O(2^{|\mathcal{Y}|})$), in the worst case $\mathcal{A} = \{\{1, \dots, |\mathcal{Y}|\}\}$. Since the expression $R^*(g) = H(j(X, g(Y))|g(Y))$ from Theorem 4 allows for an evaluation of $R^*(g)$ in polynomial time in $|\mathcal{Y}|$, the coarsening (resp. refining) greedy algorithm has a polynomial (resp. exponential) time complexity. \square

Author Contributions: Formal analysis, N.C.; validation, N.C., M.L.T. and A.R.; writing—original draft preparation, N.C.; writing—review and editing, M.L.T. and A.R.; supervision, M.L.T. and A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the Cominlabs excellence laboratory with the French National Research Agency’s funding (ANR-10-LABX-07-01).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Appendix A.1. Proof of Theorem 2

Consider the particular case $f(X, Y) = X$ of Figure 1. The optimal rate in this particular case equals the optimal rate $R^*(g)$ in the following auxiliary problem, depicted in Figure A1: $(X, g(Y))$ as source available at the encoder and to be retrieved by the decoder which knows Y (thus, expecting it to retrieve $g(Y)$ in addition to X does not change the optimal rate).

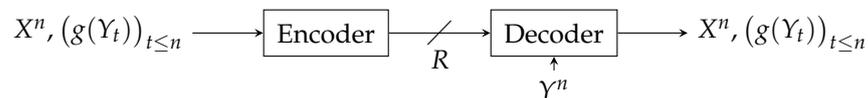


Figure A1. An auxiliary zero-error Slepian–Wolf problem.

Definition A1 (Characteristic graph for the zero-error Slepian–Wolf problem). *Let \mathcal{X}, \mathcal{Y} be two finite sets and $P_{Y|X}$ be a conditional distribution from $\Delta(\mathcal{Y})^{|\mathcal{X}|}$. The characteristic graph associated with $P_{Y|X}$ is defined by the following:*

- \mathcal{X} as set of vertices;
- $x, x' \in \mathcal{X}$ are adjacent if $P_{Y|X}(y|x)P_{Y|X}(y|x') > 0$ for some $y \in \mathcal{Y}$.

This auxiliary problem is a particular instance of the zero-error Slepian–Wolf problem; its optimal rate is written as $\bar{H}(G)$, where $\bar{H}(G)$ is the complementary graph entropy [17]

and G is the characteristic graph in the Slepian–Wolf problem, defined in Definition A1, for the pair $((X, g(Y)), Y)$. The graph G has $\mathcal{X} \times \mathcal{Z}$ as a set of vertices, and (x, z) is adjacent to (x', z') if there exists a side information symbol $y \in \mathcal{Y}$ such that $P_{X,Y,g(Y)}(x, y, z)P_{X,Y,g(Y)}(x', y, z') > 0$. It can be observed that the vertices (x, z) and (x', z') such that $z \neq z'$ are not adjacent in G . The graph G is therefore a disjoint union indexed by \mathcal{Z} :

$$G = \bigsqcup_{z \in \mathcal{Z}}^{P_{g(Y)}} G_z; \tag{A1}$$

$$R^*(g) = \overline{H}(G) = \overline{H}\left(\bigsqcup_{z \in \mathcal{Z}}^{P_{g(Y)}} G_z\right); \tag{A2}$$

where for all $z \in \mathcal{Z}$, G_z is the characteristic graph defined in Definition A1 for the pair $(X'_z, Y'_z) \sim P_{X,Y|g(Y)=z}$.

(\Rightarrow) Assume that g and $P_{X,Y}$ satisfy the “pairwise shared side information” condition. It directly follows that $P_{X|g(Y)=z}$ is full support for all $z \in \mathcal{Z}$. Let $z \in \mathcal{Z}$, and let $(x, z), (x', z)$ be any two vertices of G_z . By construction, there exists $y \in g^{-1}(z)$ such that $P_{X,Y}(x, y)P_{X,Y}(x', y) > 0$; hence, $P_{X,Y,g(Y)}(x, y, z)P_{X,Y,g(Y)}(x', y, z) > 0$, and $(x, z), (x', z)$ are adjacent in G_z . Each graph G_z is therefore complete and perfect; the graph $G = \bigsqcup_{z \in \mathcal{Z}}^{P_{g(Y)}} G_z$ is a disjoint union of perfect graphs and is therefore also perfect. We have the following:

$$R^*(g) = \overline{H}\left(\bigsqcup_{z \in \mathcal{Z}}^{P_{g(Y)}} G_z\right) \tag{A3}$$

$$= H_\kappa\left(\bigsqcup_{z \in \mathcal{Z}}^{P_{g(Y)}} G_z\right) \tag{A4}$$

$$= \sum_{z \in \mathcal{Z}} P_{g(Y)}(z)H_\kappa(G_z) \tag{A5}$$

$$= \sum_{z \in \mathcal{Z}} P_{g(Y)}(z)H(P_{X|g(Y)=z}) \tag{A6}$$

$$= H(X|g(Y)); \tag{A7}$$

where (A3) comes from (A2); (A4) and (A5) follow from Corollary 12 in [18] used on the perfect graph $\bigsqcup_{z \in \mathcal{Z}}^{P_{g(Y)}} G_z$; and (A6) holds as the independent subsets of the complete graph G_z are singletons containing one of its vertices.

(\Leftarrow) Conversely, assume that $P_{X|g(Y)=z}$ is full support for all $z \in \mathcal{Z}$, and $R^*(g) = H(X|g(Y))$.

Assume, ad absurdum, that at least one of the $(G_z)_{z \in \mathcal{Z}}$ is not complete; then, there exists a coloring of that graph that maps two different vertices to the same color. Thus, there exists $z \in \mathcal{Z}$ such that

$$\overline{H}(G_z) < H(P_{X|g(Y)=z}), \tag{A8}$$

as $P_{X|g(Y)=z}$ is full support. We have

$$H(X|g(Y)) = R^*(g) \tag{A9}$$

$$= \overline{H}\left(\bigsqcup_{z \in \mathcal{Z}}^{P_{g(Y)}} G_z\right) \tag{A10}$$

$$\leq \sum_{z \in \mathcal{Z}} P_{g(Y)}(z)\overline{H}(G_z) \tag{A11}$$

$$< H(X|g(Y)); \tag{A12}$$

where (A10) comes from (A2), (A11) results from Theorem 2 in [17], and (A12) follows from (A8). We arrive at a contradiction, and hence all the graphs $(G_z)_{z \in \mathcal{Z}}$ are complete: for all $z \in \mathcal{Z}$ and $x, x' \in \mathcal{X}$, there exists a side information symbol $y \in \mathcal{Y}$ such that $P_{X,Y,g(Y)}(x, y, z)P_{X,Y,g(Y)}(x', y, z) > 0$; hence, $y \in g^{-1}(z)$ and satisfies $P_{X,Y}(x, y)P_{X,Y}(x', y) > 0$. The condition “pairwise shared side information” is satisfied by $P_{X,Y}, g$.

Appendix A.2. Proof of Theorem 3

Let us specify the adjacency condition in $G_{[n]}$ under assumption (13). Two vertices are adjacent if they satisfy (7) and (8); however, (7) is always satisfied under (13). Thus, $(x^n, z^n)(x'^n, z'^n)$ are adjacent if $z^n = z'^n$ and

$$\exists y^n \in g^{-1}(z^n), \exists t \leq n, f(x_t, y_t) \neq f(x'_t, y_t). \tag{A13}$$

It can be observed that condition (A13) is the adjacency condition of an OR product of adequate graphs; more precisely,

$$G_{[n]} = \bigsqcup_{z^n \in \mathcal{Z}^n} \bigvee_{t \leq n} G_{z_t}^f. \tag{A14}$$

Although $G_{[n]}$ cannot be expressed as an n -th OR power, we will show that its chromatic entropy asymptotically coincides with that of an appropriate OR power: we now search for an asymptotic equivalent of $H_\chi(G_{[n]})$.

Definition A2. \mathcal{S}_n is the set of colorings of $G_{[n]}$ that can be written as $(x^n, z^n) \mapsto (T_{z^n}, \tilde{c}(x^n, z^n))$ for some mapping $\tilde{c} : \mathcal{X}^n \times \mathcal{Z}^n \rightarrow \tilde{\mathcal{C}}$; where T_{z^n} denotes the type of z^n .

In the following, we define $Z^n \doteq (g(Y_t))_{t \leq n}$. Now, we need several Lemmas. Lemma A1 states that the optimal coloring $c(x^n, z^n)$ of $G_{[n]}$ has the type of z^n as a prefix at a negligible rate cost. Lemma A3 gives an asymptotic formula for the minimal entropy of the colorings from \mathcal{S}_n .

Lemma A1. The following asymptotic comparison holds as follows:

$$H_\chi(G_{[n]}) = \inf_{\substack{c \text{ coloring of } G_{[n]} \\ \text{s.t. } c \in \mathcal{S}_n}} H(c(X^n, Z^n)) + O(\log n). \tag{A15}$$

Definition A3 (Isomorphic probabilistic graphs). Let $G_1 = (\mathcal{V}_1, \mathcal{E}_1, P_{V_1})$ and $G_2 = (\mathcal{V}_2, \mathcal{E}_2, P_{V_2})$ be two probabilistic graphs. We say that G_1 is isomorphic to G_2 (denoted by $G_1 \simeq G_2$) if there exists an isomorphism between them, i.e., a bijection $\psi : \mathcal{V}_1 \rightarrow \mathcal{V}_2$ such that

- For all $v_1, v'_1 \in \mathcal{V}_1, v_1 v'_1 \in \mathcal{E}_1 \iff \psi(v_1) \psi(v'_1) \in \mathcal{E}_2$;
- For all $v_1 \in \mathcal{V}_1, P_{V_1}(v_1) = P_{V_2}(\psi(v_1))$.

Lemma A2. Let \mathcal{B} be a finite set, let $P_B \in \Delta(\mathcal{B})$ and let $(G_b)_{b \in \mathcal{B}}$ be a family of isomorphic probabilistic graphs, then $H_\chi(\bigsqcup_{b' \in \mathcal{B}}^{P_B} G_{b'}) = H_\chi(G_b)$ for all $b \in \mathcal{B}$.

Lemma A3. The following asymptotic comparison holds as follows:

$$\inf_{\substack{c \text{ coloring of } G_{[n]} \\ \text{s.t. } c \in \mathcal{S}_n}} H(c(X^n, Z^n)) = n \sum_{z \in \mathcal{Z}} P_{g(Y)}(z) H_\kappa(G_z^f) + o(n). \tag{A16}$$

The proof of Lemma A1 is given in Appendix A.3 and its keypoint is the asymptotically negligible entropy of the prefix T_{Z^n} of the colorings of \mathcal{S}_n . The proof of Lemma A2 is given in Appendix A.5. The proof of Lemma A3 is given in Appendix A.4 and relies on the decomposition $G_{[n]} = \bigsqcup_{Q_n \in \Delta_n(\mathcal{Z})} G_{[n]}^{Q_n}$, where $G_{[n]}^{Q_n}$ is the subgraph induced by the vertices (x^n, z^n) such that the type of z^n is Q_n . We show that $G_{[n]}^{Q_n}$ is a disjoint union of isomorphic graphs whose chromatic entropy is given by Lemma A2 and (17): $|H_\chi(G_{[n]}^{Q_n}) - n \sum_{z \in \mathcal{Z}} Q_n(z) H_\kappa(G_z^f)| \leq n \epsilon_n$. Finally, uniform convergence arguments enable us draw a conclusion.

Now, let us combine these results together as follows:

$$R^*(g) = \frac{1}{n} H_\chi(G_{[n]}) + o(1) \tag{A17}$$

$$= \frac{1}{n} \inf_{\substack{c \text{ coloring of } G_{[n]} \\ \text{s.t. } c \in \mathcal{S}_n}} H(c(X^n, Z^n)) + o(1) \tag{A18}$$

$$= \sum_{z \in \mathcal{Z}} P_{g(\gamma)}(z) H_\kappa(G_z^f) + o(1), \tag{A19}$$

where (A17) comes from Theorem 1, (A18) comes from Lemma A1, and (A19) comes from Lemma A3. The proof of Theorem 3 is complete.

Appendix A.3. Proof of Lemma A1

Let c_n^* be the coloring of $G_{[n]}$ with minimal entropy. Then, we have the following:

$$H_\chi(G_{[n]}) = \inf_{c \text{ coloring of } G_{[n]}} H(c(X^n, Z^n)) \tag{A20}$$

$$\leq \inf_{\substack{c \text{ coloring of } G_{[n]} \\ \text{s.t. } c \in \mathcal{S}_n}} H(c(X^n, Z^n)) \tag{A21}$$

$$= \inf_{\substack{c: (x^n, z^n) \\ \mapsto (T_{z^n}, \tilde{c}(x^n, z^n))}} H(T_{Z^n}, \tilde{c}(X^n, Z^n)) \tag{A22}$$

$$\leq H(T_{Z^n}) + H(c_n^*(X^n, Z^n)) \tag{A23}$$

$$= H_\chi(G_{[n]}) + O(\log n), \tag{A24}$$

where (A22) comes from Definition A2; (A23) comes from the subadditivity of the entropy, $(x^n, z^n) \mapsto (T_{z^n}, c_n^*(x^n, z^n))$ is a coloring of $G_{[n]}$ that belongs to \mathcal{S}_n , and (A24) comes from $H(T_{Z^n}) = O(\log n)$, as $\log |\Delta_n(\mathcal{Z})| = O(\log n)$. The desired equality comes from the bounds $H_\chi(G_{[n]})$ and $H_\chi(G_{[n]}) + O(\log n)$ on (A21).

Appendix A.4. Proof of Lemma A3

For all $Q_n \in \Delta_n(\mathcal{Z})$, let

$$G_{[n]}^{Q_n} = \bigsqcup_{\substack{z^n \in \mathcal{Z}^n \\ \tilde{T}_{z^n} = Q_n}} \bigvee_{t \leq n} G_{z_t}^f, \tag{A25}$$

with the probability distribution induced by $P_{X,Z}^n$. This graph is formed of the connected components of $G_{[n]}$ whose corresponding z^n has type Q_n . We need to find an equivalent for $H_\chi(G_{[n]}^{Q_n})$. Since $G_{[n]}^{Q_n}$ is a disjoint union of isomorphic graphs, we can use Lemma A2 as follows:

$$H_\chi(G_{[n]}^{Q_n}) = H_\chi\left(\bigvee_{z \in \mathcal{Z}} (G_z^f)^{\vee n Q_n(z)}\right). \tag{A26}$$

On one hand,

$$H_\chi\left(\bigvee_{z \in \mathcal{Z}} (G_z^f)^{\vee n Q_n(z)}\right) \geq H_\kappa\left(\bigvee_{z \in \mathcal{Z}} (G_z^f)^{\vee n Q_n(z)}\right) \tag{A27}$$

$$= n \sum_{z \in \mathcal{Z}} Q_n(z) H_\kappa(G_z^f), \tag{A28}$$

where (A27) comes from $H_\kappa \leq H_\chi$ Lemma 14 in [14], (A28) comes from (17). On the other hand,

$$H_\chi \left(\bigvee_{z \in \mathcal{Z}} (G_z^f)^{\vee n Q_n(z)} \right) \leq \sum_{z \in \mathcal{Z}} Q_n(z) H_\chi((G_z^f)^{\vee n}) \tag{A29}$$

$$= n \sum_{z \in \mathcal{Z}} Q_n(z) H_\kappa(G_z^f) + n\epsilon_n, \tag{A30}$$

where $\epsilon_n \doteq \max_z \frac{1}{n} H_\chi((G_z^f)^{\vee n}) - H_\kappa(G_z^f)$ is a quantity that does not depend on Q_n and satisfies $\lim_{n \rightarrow \infty} \epsilon_n = 0$; (A29) comes from the subadditivity of H_χ . Combining Equations (A26), (A28), and (A30) yields

$$\left| H_\chi(G_{[n]}^{Q_n}) - n \sum_{z \in \mathcal{Z}} Q_n(z) H_\kappa(G_z^f) \right| \leq n\epsilon_n. \tag{A31}$$

Now, we have an equivalent for $H_\chi(G_{[n]}^{Q_n})$.

$$\inf_{\substack{c \text{ coloring of } G_{[n]} \\ \text{s.t. } c \in \mathcal{S}_n}} H(c(X^n, Z^n)) \tag{A32}$$

$$= \inf_{\substack{c: (x^n, z^n) \\ \mapsto (T_{Z^n}, \tilde{c}(x^n, z^n))}} H(\tilde{c}(X^n, Z^n) | T_{Z^n}) + H(T_{Z^n}) \tag{A33}$$

$$= \inf_{\substack{c: (x^n, z^n) \\ \mapsto (T_{Z^n}, \tilde{c}(x^n, z^n))}} \sum_{Q_n \in \Delta_n(\mathcal{Z})} P_{T_{Z^n}}(Q_n) H(\tilde{c}(X^n, Z^n) | T_{Z^n} = Q_n) + O(\log n) \tag{A34}$$

$$= \sum_{Q_n \in \Delta_n(\mathcal{Z})} P_{T_{Z^n}}(Q_n) \inf_{c_{Q_n} \text{ coloring of } G_{[n]}^{Q_n}} H(c_{Q_n}(X^n, Z^n) | T_{Z^n} = Q_n) + O(\log n) \tag{A35}$$

$$= \sum_{Q_n \in \Delta_n(\mathcal{Z})} P_{T_{Z^n}}(Q_n) H_\chi(G_{[n]}^{Q_n}) + O(\log n) \tag{A36}$$

$$= \sum_{Q_n \in \Delta_n(\mathcal{Z})} P_{T_{Z^n}}(Q_n) \left(n \sum_{z \in \mathcal{Z}} Q_n(z) H_\kappa(G_z^f) \pm n\epsilon_n \right) + O(\log n) \tag{A37}$$

$$= n \sum_{Q_n \in \Delta_n(\mathcal{Z})} 2^{-nD(Q_n \| P_{g(Y)}) + o(n)} \left(\sum_{z \in \mathcal{Z}} Q_n(z) H_\kappa(G_z^f) \right) \pm n\epsilon_n + O(\log n) \tag{A38}$$

$$= n \sum_{z \in \mathcal{Z}} P_{g(Y)}(z) H_\kappa(G_z^f) + o(n), \tag{A39}$$

where (A34) comes from $H(T_{Z^n}) = O(\log n)$, as $\log |\Delta_n(\mathcal{Z})| = O(\log n)$; (A35) follows from the fact that the entropy of \tilde{c} can be minimized independently on each $G_{[n]}^{Q_n}$; (A36) follows from the definition of $G_{[n]}^{Q_n}$; (A37) comes from (A31); (A38) comes from Lemma 2.6 in [19] and the fact that ϵ_n does not depend on Q_n .

Appendix A.5. Proof of Lemma A2

Let $(\tilde{G}_i)_{i \leq N}$ be isomorphic probabilistic graphs and G such that $G = \bigsqcup_i \tilde{G}_i$. Let $c_1^* : \mathcal{V}_1 \rightarrow \mathcal{C}$ be the coloring of \tilde{G}_1 with minimal entropy, and let c^* be the coloring of G defined by

$$c^* : \mathcal{V} \rightarrow \mathcal{C} \tag{A40}$$

$$\text{vs. } \mapsto c_1^* \circ \psi_{i_v \rightarrow 1}(v), \tag{A41}$$

where i_v is the unique integer such that $v \in \mathcal{V}_{i_v}$ and $\psi_{i_v \rightarrow 1} : \mathcal{V}_{i_v} \rightarrow \mathcal{V}_1$ is an isomorphism between \tilde{G}_{i_v} and \tilde{G}_1 . In other words, c^* applies the same coloring pattern c_1^* on each connected component of G . We have

$$H_\chi(G) \leq H(c^*(V)) \quad (\text{A42})$$

$$= h\left(\sum_{j \leq N} P_{i_v}(j) P_{c^*(V_j)}\right) \quad (\text{A43})$$

$$= h\left(\sum_{j \leq N} P_{i_v}(j) P_{c_1^*(V_1)}\right) \quad (\text{A44})$$

$$= H(c_1^*(V_1)) \quad (\text{A45})$$

$$= H_\chi(\tilde{G}_1), \quad (\text{A46})$$

where h denotes the entropy of a distribution; (A44) comes from the definition of c^* ; and (A46) comes from the definition of c_1^* .

Now, let us prove the upper bound on $H_\chi(\tilde{G}_1)$. Let c be a coloring of G , and let $i^* \doteq \operatorname{argmin}_i H(c(V_i))$ (i.e., i^* is the index of the connected component for which the entropy of the coloring induced by c is minimal). We have

$$H(c(V)) = h\left(\sum_{j \leq N} P_{i_v}(j) P_{c(V_j)}\right) \quad (\text{A47})$$

$$\geq \sum_{j \leq N} P_{i_v}(j) h(P_{c(V_j)}) \quad (\text{A48})$$

$$\geq \sum_{j \leq N} P_{i_v}(j) H(c(V_{i^*})) \quad (\text{A49})$$

$$\geq H_\chi(\tilde{G}_{i^*}), \quad (\text{A50})$$

$$= H_\chi(\tilde{G}_1), \quad (\text{A51})$$

where (A48) follows from the concavity of h ; (A49) follows from the definition of i^* ; (A50) comes from the fact that c induces a coloring of \tilde{G}_{i^*} ; (A51) comes from the fact that \tilde{G}_1 and \tilde{G}_{i^*} are isomorphic. Now, we can combine the bounds (A46) and (A51): for all coloring c of G we have

$$H_\chi(G) \leq H_\chi(\tilde{G}_1) \leq H(c(V)), \quad (\text{A52})$$

which yields the desired equality when taking the infimum over c .

References

- Orlitsky, A.; Roche, J.R. Coding for computing. In Proceedings of the IEEE 36th Annual Foundations of Computer Science, Milwaukee, WI, USA, 23–25 October 1995; IEEE: Piscataway, NJ, USA, 1995; pp. 502–511.
- Duan, L.; Liu, J.; Yang, W.; Huang, T.; Gao, W. Video coding for machines: A paradigm of collaborative compression and intelligent analytics. *IEEE Trans. Image Process.* **2020**, *29*, 8680–8695. [[CrossRef](#)] [[PubMed](#)]
- Gao, W.; Liu, S.; Xu, X.; Rafie, M.; Zhang, Y.; Curcio, I. Recent standard development activities on video coding for machines. *arXiv* **2021**, arXiv:2105.12653.
- Yamamoto, H. Wyner-ziv theory for a general function of the correlated sources (corresp.). *IEEE Trans. Inf. Theory* **1982**, *28*, 803–807. [[CrossRef](#)]
- Shayevitz, O. Distributed computing and the graph entropy region. *IEEE Trans. Inf. Theory* **2014**, *60*, 3435–3449. [[CrossRef](#)]
- Krithivasan, D.; Pradhan, S.S. Distributed source coding using abelian group codes: A new achievable rate-distortion region. *IEEE Trans. Inf. Theory* **2011**, *57*, 1495–1519. [[CrossRef](#)]
- Basu, S.; Seo, D.; Varshney, L.R. Hypergraph-based Coding Schemes for Two Source Coding Problems under Maximal Distortion. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Los Angeles, CA, USA, 21–26 June 2020.
- Malak, D.; Médard, M. Hyper Binning for Distributed Function Coding. In Proceedings of the 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Atlanta, GA, USA, 26–29 May 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–5.
- Feizi, S.; Médard, M. On network functional compression. *IEEE Trans. Inf. Theory* **2014**, *60*, 5387–5401. [[CrossRef](#)]
- Sefidgaran, M.; Tchamkerten, A. Distributed function computation over a rooted directed tree. *IEEE Trans. Inf. Theory* **2016**, *62*, 7135–7152. [[CrossRef](#)]
- Ravi, J.; Dey, B.K. Function Computation Through a Bidirectional Relay. *IEEE Trans. Inf. Theory* **2018**, *65*, 902–916. [[CrossRef](#)]

12. Guang, X.; Yeung, R.W.; Yang, S.; Li, C. Improved upper bound on the network function computing capacity. *IEEE Trans. Inf. Theory* **2019**, *65*, 3790–3811. [[CrossRef](#)]
13. Charpenay, N.; Le Treust, M.; Roumy, A. Optimal Zero-Error Coding for Computing under Pairwise Shared Side Information. In Proceedings of the 2023 IEEE Information Theory Workshop (ITW), Saint-Malo, France, 23–28 April 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 97–101.
14. Alon, N.; Orlitsky, A. Source coding and graph entropies. *IEEE Trans. Inf. Theory* **1996**, *42*, 1329–1339. [[CrossRef](#)]
15. Witsenhausen, H. The zero-error side information problem and chromatic numbers (corresp.). *IEEE Trans. Inf. Theory* **1976**, *22*, 592–593. [[CrossRef](#)]
16. Körner, J. Coding of an information source having ambiguous alphabet and the entropy of graphs. In Proceedings of the 6th Prague Conference on Information Theory, Prague, 19–25 September 1973; pp. 411–425.
17. Tuncel, E.; Nayak, J.; Koulgi, P.; Rose, K. On complementary graph entropy. *IEEE Trans. Inf. Theory* **2009**, *55*, 2537–2546. [[CrossRef](#)]
18. Csiszár, I.; Körner, J.; Lovász, L.; Marton, K.; Simonyi, G. Entropy splitting for antiblocking corners and perfect graphs. *Combinatorica* **1990**, *10*, 27–40. [[CrossRef](#)]
19. Csiszár, I.; Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*; Cambridge University Press: Cambridge, UK, 2011.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.