



HAL
open science

Zero-error source coding when side information may be present

Nicolas Charpenay, Maël Le Treust, Aline Roumy

► **To cite this version:**

Nicolas Charpenay, Maël Le Treust, Aline Roumy. Zero-error source coding when side information may be present. 2021. hal-03290860v1

HAL Id: hal-03290860

<https://hal.science/hal-03290860v1>

Preprint submitted on 19 Jul 2021 (v1), last revised 16 Dec 2021 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Zero-error source coding when side information may be present

Nicolas Charpenay^{*}, Maël Le Treust[†], Aline Roumy[‡]

^{*}IRISA, 263 avenue du Général Leclerc, 35 042 Rennes cedex, France, nicolas.charpenay@irisa.fr

[†]ETIS UMR 8051, CY Université, ENSEA, CNRS, 6 avenue du Ponceau, 95014 Cergy cedex, France, mael.le-treust@ensea.fr

[‡]INRIA Rennes, Campus de Beaulieu, 35042 Rennes cedex, France, aline.roumy@inria.fr

Abstract—Zero-error source coding when side-information (SI) may be present is a fundamental building block of interactive real-world compression systems. In this paper, we show how to use the SI at the encoder in order to lower the compression rate. Indeed, in general, without SI at the encoder, the minimum zero-error compression rate is the entropy of the source to be compressed. In this paper, we show, that the availability of the SI at the encoder allows to achieve lower rate. The code construction relies on the typicality bipartite graph in which the vertices corresponds to the sequences of source and side information, where an edge links the pair of exactly typical sequences. By considering binary source alphabets, we show that the Hamming distance characterize the graph coloring. For short block length, we show that our type grouping code outperforms any time-sharing strategy between the Huffman code and the conditional Huffman code. Then we extend these results to arbitrary source alphabets.

I. INTRODUCTION

Consider a scenario where a source X is compressed with the help of a side information (SI) Y and sent to two decoders, as shown in Fig. 1. The SI is available at the encoder, and the first decoder, but not at the second decoder. This scheme has been studied in [1] for lossy compression. The lossless scenario is of practical interest, when the data is accessed in an interactive way, i.e. when only a small part of the data is requested by the user. Indeed, in this context the decoding order depends on the request, and is not known a priori by the encoder. This uncertainty can be modeled by the fact that the encoder does not know if a SI is available at the decoder, and if so, which SI, as in [2]. The scheme Fig. 1 considered in this paper corresponds to the one-SI case, where the encoder does not know whether the decoder has access to the SI or not.

For lossless reconstruction with vanishing small error, an incremental random binning strategy as in [3] (which does not use encoder SI) is enough to send at rates $R_1 = \mathbb{H}(X|Y)$, and $R_1 + R_2 = \mathbb{H}(X)$, i.e. at the same rate as if the encoder knew whether the SI is available or not at the decoder. This positive result motivated the use of practical schemes implementing the random binning strategy in interactive compression schemes applied to either omnidirectional images in [2] or data living on graphs in [4]. However, in the latter applications, a zero error code has been used in [5], since, as in many lossy video compression schemes, zero-error source coding is a fundamental building block. More precisely, the no-error property can be achieved by using the knowledge of the SI at the encoder.

If practical zero-error codes exist for the problem in Fig. 1, there is no known single-letter asymptotic characterization in general for the zero-error problem in Fig. 1.

A set of relevant work concerns Slepian-Wolf coding [6] i.e. when the SI is not available at the encoder. However, the result is quite pessimistic as it is shown in [7] that, in general, the zero-error variant of the source coding problem with SI at decoder is only possible if $R_1 \geq \mathbb{H}(X)$. However, as for the zero-error channel coding problem, in certain cases, when certain symbol pairs have zero probability, then zero-error Slepian Wolf coding is possible below the entropy [8], [9].

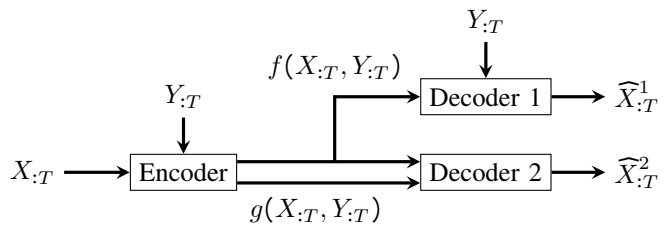


Fig. 1: Source coding when side-information may be present

In this paper we investigate a coding algorithm that is based on a *typicality bipartite graph*, which links the sequences of source and side information that are exactly typical. This object was under study in [10], for strongly typical sequences with a tolerance parameter. We characterize the graph coloring in terms of the Hamming distance between the sequences of binary symbols of source and side information. Then we introduce the type grouping code that groups the types of the sequences, and forms a coloring for all the typicality graphs in this group. The group and the color are transmitted to both decoders, and additional information is conveyed to the second decoder that does not observe the side information. We provide an example for which the rate pair of the type grouping code is below the time-sharing segment between the direct transmission to each decoder, and the conditional coding with respect to the side-information. As an extension, we show that the optimal codes for binary source alphabet can be used to construct optimal codes for arbitrary source alphabets.

II. PRESENTATION OF THE MODEL

We denote by $x_{:T} = (x_1, \dots, x_T)$ the slicing of a sequence $(x_t)_{t \in \mathbb{N}}$. Integer intervals are denoted by double brackets: $\forall n, n' \in \mathbb{Z}$ such that $n \leq n'$, $\llbracket n, n' \rrbracket \doteq \{n, n+1, \dots, n'-1, n'\}$. Let \mathcal{S} be a finite set, we denote by $\#\mathcal{S}$ its cardinality. Let $\{0, 1\}^*$ be the set of binary words. For all $w \in \{0, 1\}^*$, the length of w is denoted by $|w|$. The random variables will be denoted by capital letters, and their set of possible values with the corresponding calligraphic letters. For example $X : \Omega \rightarrow \mathcal{X}$, where Ω denotes the set of aleas. The set of probability distributions over a finite set \mathcal{X} is denoted by $\Delta(\mathcal{X})$ ($\in \mathbb{R}^{\#\mathcal{X}}$). The distribution of a random variable X is denoted by a vector $p^X \in \Delta(\mathcal{X})$, as its set of possible values will be finite. This will give us geometric interpretations of probabilistic phenomena. The conditional distribution of a random variable Y knowing X is denoted by a matrix $p^{X|Y} \in \mathbb{R}^{\#\mathcal{X} \times \mathcal{Y}}$. This allows for extracting the marginal distributions with a dot product, for example. The same applies for the joint distribution $p^{X,Y}$.

We consider the setting of Fig. 1 described by:

- ◆ An information source over a finite set $\mathcal{X} \times \mathcal{Y}$ with publicly observed distribution $p^{X,Y} \in \Delta(\mathcal{X} \times \mathcal{Y})$, and a pair of random variables $(X, Y) \sim p^{X,Y}$.
- ◆ An encoder that observes the realizations of (X, Y) .
- ◆ Two decoders, where only the decoder 1 observes the realizations of the side-information Y .
- ◆ The encoder transmits over a common channel to both decoders, and also over a secondary channel to the decoder 2 only.
- ◆ For $T \in \mathbb{N}^* = \mathbb{N} \setminus \{0\}$ iterated source uses, we consider a sequence $(X_{:T}, Y_{:T})$ of independent copies of (X, Y) . The codewords sent by the encoder are concatenated, so the decoder has to choose the right decomposition into codewords before decoding.

Definition II.1 (*T*-Source code, rate) A *T*-source code is a pair of mappings $f : \mathcal{X}^T \times \mathcal{Y}^T \rightarrow \{0, 1\}^*$ and $g : \mathcal{X}^T \times \mathcal{Y}^T \rightarrow \{0, 1\}^*$. We define their rates:

$$\begin{aligned} R_1 &\doteq \frac{1}{T} \mathbb{E}_{X,Y} \left[|f(X_{:T}, Y_{:T})| \right], \\ R_2 &\doteq \frac{1}{T} \mathbb{E}_{X,Y} \left[|g(X_{:T}, Y_{:T})| \right]. \end{aligned} \quad (1)$$

The rate R_1 (resp. R_2) corresponds to the quantity of information sent over the common (resp. secondary) channel.

Definition II.2 (Zero-error) In this setting, a *T*-source code is zero-error if both decoders can retrieve X with zero-error. That is, there exists two functions $\widehat{X}^1 : \{0, 1\}^* \times \mathcal{Y}^T \rightarrow \mathcal{X}^T$ and $\widehat{X}^2 : (\{0, 1\}^*)^2 \rightarrow \mathcal{X}^T$ such that

$$\begin{aligned} \widehat{X}_{:T}^1(f(X_{:T}, Y_{:T}), Y_{:T}) &= X_{:T}, \\ \widehat{X}_{:T}^2(f(X_{:T}, Y_{:T}), g(X_{:T}, Y_{:T})) &= X_{:T}, \end{aligned} \quad (2)$$

with probability 1.

Theorem II.3 In this setting, for all $T \in \mathbb{N}$ and for all zero-error *T*-source code (f, g) we have:

$$R_1 \geq \mathbb{H}(X|Y), \quad R_1 + R_2 \geq \mathbb{H}(X).$$

Proof. In this setting, each decoder must retrieve X without loss. Using Shannon lossless source coding result and Slepian-Wolf theorem [6, Theorem 2] on each decoder, we have $R_1 \geq \mathbb{H}(X|Y)$ and $R_1 + R_2 \geq \mathbb{H}(X)$. \square

Definition II.4 (Asymptotically achievable rates) Let $(r_1, r_2) \in \mathbb{R}^2$, the pair (r_1, r_2) is asymptotically achievable if for all $\epsilon > 0$ there exists an integer T and a zero-error *T*-source code (f, g) such that

$$R_1 \leq r_1 + \epsilon, \quad R_2 \leq r_2 + \epsilon.$$

Theorem II.5 The segment with extremities $(r_1, r_2) = (\mathbb{H}(X), 0)$ and $(\mathbb{H}(X|Y), \mathbb{H}(Y))$ is asymptotically achievable.

Proof. On one hand, the pair of rates $(\mathbb{H}(X), 0)$ is asymptotically achievable by using a Huffman code to transmit X on the common channel. On the other hand, the pair of rates $(\mathbb{H}(X|Y), \mathbb{H}(Y))$ is asymptotically achievable by using a conditional Huffman code on the common channel, and by transmitting Y on the secondary channel. With time-sharing, we have that each point of the segment with extremities $(\mathbb{H}(X), 0)$ and $(\mathbb{H}(X|Y), \mathbb{H}(Y))$ is asymptotically achievable. \square

We illustrated a specific example in Fig. 4, with the feasible segment from Theorem II.5 and the non-achievable rates from II.3.

III. TYPICALITY GRAPH AND PROPERTIES

We consider the case where $\mathcal{X} = \mathcal{Y} = \{0, 1\}$.

Definition III.1 (Type) For all pair of sequences $(x_{:T}, y_{:T}) \in \mathcal{X}^T \times \mathcal{Y}^T$, the joint type is defined by

$$\tau(x_{:T}, y_{:T}) \doteq \left(\frac{1}{T} \#\{t \leq T \mid (x_t, y_t) = (\bar{x}, \bar{y})\} \right)_{(\bar{x}, \bar{y}) \in \mathcal{X} \times \mathcal{Y}}.$$

We denote the marginal types by $\tau(x_{:T})$ and $\tau(y_{:T})$, respectively. The *T*-discretized simplex $\Delta_T(\mathcal{X} \times \mathcal{Y})$ is the set of types from $\Delta(\mathcal{X} \times \mathcal{Y})$ that are achievable using sequences of length *T*, i.e.

$$\Delta_T(\mathcal{X} \times \mathcal{Y}) \doteq \left\{ \tau^* \in \Delta(\mathcal{X} \times \mathcal{Y}) \mid \forall (x, y), T\tau_{x,y}^* \in \mathbb{N} \right\}. \quad (3)$$

In this section, we seek for zero-error decoding functions. Showing the existence of such functions rely on the notion of deterministic degradation defined below.

Definition III.2 (Deterministic degradation \leq) The random variable A is a deterministic degradation of the random variable B if there exists a deterministic function $h : \mathcal{B} \rightarrow \mathcal{A}$ such that $h(B) = A$ almost surely, which is denoted by $A \leq B$. Moreover, $A \leq B$ conditionally given an event e , if $h(B) = A$ whenever e is realized.

Proposition III.3 Let C be a random variable such that $A \preceq B$ conditionally given $C = c$ for all $c \in \mathcal{C}$, then $A \preceq (B, C)$.

Proof. For all $c \in \mathcal{C}$, let $h_c : \mathcal{B} \rightarrow \mathcal{A}$ be the function such that $h_c(B) = A$ whenever $C = c$. Then the mapping $\bar{h} : (b, c) \mapsto h_c(b)$ is deterministic and satisfies $\bar{h}(B, C) = A$ almost surely. Thus, $A \preceq (B, C)$. \square

We now introduce the typicality graph. Unlike in [10], this graph is defined for a given type and not a set of types. This will allow us to characterize explicitly the distance properties between the sequences having the same type in Prop. III.8.

Definition III.4 (Typicality graph) Let $\tau^* \in \Delta_T(\mathcal{X} \times \mathcal{Y}) = \Delta_T(\{0, 1\}^2)$, we denote by $G_{\tau^*} = ((\mathcal{V}_X, \mathcal{V}_Y), \mathcal{E})$ the typicality bipartite graph for the type τ^* . The sets \mathcal{V}_X and \mathcal{V}_Y are the sets of sequences that are typical w.r.t. the corresponding marginal distribution of τ^* . There is an edge between pairs of sequences that are jointly typical w.r.t. τ^* . More formally:

$$\begin{aligned} \mathcal{V}_X &= \{x_{:T} \in \mathcal{X}^T \mid \tau(x_{:T}) = (\tau_{0,0}^* + \tau_{0,1}^*, \tau_{1,0}^* + \tau_{1,1}^*)\}, \quad (4) \\ \mathcal{V}_Y &= \{y_{:T} \in \mathcal{Y}^T \mid \tau(y_{:T}) = (\tau_{0,0}^* + \tau_{1,0}^*, \tau_{0,1}^* + \tau_{1,1}^*)\}, \\ \forall (x_{:T}, y_{:T}) \in \mathcal{V}_X \times \mathcal{V}_Y, (x_{:T}, y_{:T}) \in \mathcal{E} &\text{ if } \tau(x_{:T}, y_{:T}) = \tau^*. \end{aligned}$$

Remark III.5 Let $\tau^* \in \Delta_T(\mathcal{X} \times \mathcal{Y}) = \Delta_T(\{0, 1\}^2)$, we have that all the sequences from \mathcal{V}_X in G_{τ^*} are equiprobable.

For all $i \in \mathbb{N}^*$, we have $\Delta_T(\{0, 1\}^2) \subset \Delta_{iT}(\{0, 1\}^2)$.

Definition III.6 (Neighborhood) Let $\tau^* \in \Delta_T(\mathcal{X} \times \mathcal{Y}) = \Delta_T(\{0, 1\}^2)$, and let $x_{:T} \in \mathcal{V}_X$ in G_{τ^*} . The neighborhood of $x_{:T}$ in G_{τ^*} is defined by

$$\mathcal{N}(x_{:T}) \doteq \{y_{:T} \in \mathcal{V}_Y \mid (x_{:T}, y_{:T}) \in \mathcal{E}\}. \quad (5)$$

Definition III.7 (\oplus , Hamming distance) Let $x_{:T}, x'_{:T} \in \{0, 1\}^T$ be two sequences, we define their addition \oplus by:

$$x_{:T} \oplus x'_{:T} = (x_t + x'_t \pmod{2})_{t \leq T} \quad (6)$$

The Hamming distance between two sequences is defined by $d(x_{:T}, x'_{:T}) = \sum_{t \leq T} \mathbb{1}_{x_t \neq x'_t}$.

Lemma 1 If $\tau(x_{:T}) = \tau(x'_{:T})$ then $d(x_{:T}, x'_{:T})$ is even.

The proof is direct. By using the intrinsic properties of the typicality graph we characterize the intersections of the sets of neighbors in terms of the Hamming distance.

Proposition III.8 (Distance and shared neighborhoods)

Let $\tau^* \in \Delta_T(\mathcal{X} \times \mathcal{Y}) = \Delta_T(\{0, 1\}^2)$, and let $x_{:T} \neq x'_{:T} \in \mathcal{V}_X$ in G_{τ^*} . Then $\mathcal{N}(x_{:T}) \cap \mathcal{N}(x'_{:T}) \neq \emptyset$ in G_{τ^*} if and only if $d(x_{:T}, x'_{:T}) \leq 2T \sum_{y \in \mathcal{Y}} \min(\tau_{0,y}^*, \tau_{1,y}^*)$.

The proof of Prop. III.8 is provided in [11].

In order to perform zero-error coding, we color the vertices of typicality graphs such that two vertices from \mathcal{V}_X can share the same color if they can not be confused while observing a sequence from \mathcal{V}_Y ; i.e. if their neighborhoods are disjoint.

Definition III.9 (X-coloring) Let $\tau^* \in \Delta_T(\{0, 1\}^2)$, and G_{τ^*} the associated typicality graph, an X-coloring of G_{τ^*} is an application k from \mathcal{V}_X to a set of colors \mathcal{C} that satisfies: for all $x_{:T} \neq x'_{:T} \in \mathcal{V}_X$, if $k(x_{:T}) = k(x'_{:T})$ then $\mathcal{N}(x_{:T})$ and $\mathcal{N}(x'_{:T})$ are disjoint in G_{τ^*} .

Given $\tau^* \in \Delta_T(\{0, 1\}^2)$, there always exists an X-coloring k in G_{τ^*} : $k(x_{:T}) = x_{:T}$ for all $x_{:T} \in \mathcal{V}_X$.

Corollary III.10 (Minimal distance for shared colors) We have that an application $k : \mathcal{V}_X \rightarrow \mathcal{C}$ is an X-coloring of G_{τ^*} if and only if for all $x_{:T} \neq x'_{:T} \in \mathcal{V}_X$ with $k(x_{:T}) = k(x'_{:T})$, we have $d(x_{:T}, x'_{:T}) > 2T \sum_{y \in \mathcal{Y}} \min(\tau_{0,y}^*, \tau_{1,y}^*)$.

The proof is a direct consequence of Prop. III.8.

Proposition III.11 Let $\tau^* \in \Delta_T(\{0, 1\}^2)$, and G_{τ^*} the associated typicality graph, we have

$$\begin{aligned} \#\mathcal{V}_X &= \binom{T}{T(\tau_{0,0}^* + \tau_{0,1}^*)}, \quad \#\mathcal{V}_Y = \binom{T}{T(\tau_{0,0}^* + \tau_{1,0}^*)}, \\ \forall x_{:T} \in \mathcal{V}_X, \#\mathcal{N}(x_{:T}) &= \binom{T(\tau_{0,0}^* + \tau_{0,1}^*)}{T\tau_{0,0}^*} \binom{T(\tau_{1,0}^* + \tau_{1,1}^*)}{T\tau_{1,1}^*}, \\ \forall y_{:T} \in \mathcal{V}_Y, \#\mathcal{N}(y_{:T}) &= \binom{T(\tau_{0,0}^* + \tau_{1,0}^*)}{T\tau_{0,0}^*} \binom{T(\tau_{0,1}^* + \tau_{1,1}^*)}{T\tau_{1,1}^*}, \\ \#\mathcal{E} &= \binom{T}{T\tau_{0,0}^*, T\tau_{0,1}^*, T\tau_{1,0}^*, T\tau_{1,1}^*}. \end{aligned}$$

Proof. The number of all the possible arrangements of $T(\tau_{0,0}^* + \tau_{0,1}^*)$ "0" symbols in T slots is $\#\mathcal{V}_X = \binom{T}{T(\tau_{0,0}^* + \tau_{0,1}^*)}$. For all $x_{:T} \in \mathcal{V}_X$, the sequences in $\#\mathcal{N}(x_{:T})$ consists of all the possible arrangements of $T\tau_{0,0}^*$ "0" symbols in the $T(\tau_{0,0}^* + \tau_{0,1}^*)$ slots where $x_{:T}$ presents a "0" symbol, times all the possible arrangements of $T\tau_{1,1}^*$ "1" symbols in the $T(\tau_{1,0}^* + \tau_{1,1}^*)$ slots where $x_{:T}$ presents a "1" symbol. Similarly, we determine the value of $\#\mathcal{V}_Y$ and $\#\mathcal{N}(y_{:T})$ for all $\forall y_{:T} \in \mathcal{V}_Y$.

It follows that G_{τ^*} is biregular, i.e. $\#\mathcal{N}(x_{:T})$ does not depend on $x_{:T}$ and $\#\mathcal{N}(y_{:T})$ does not depend on $y_{:T}$. Thus for all $x_{:T} \in \mathcal{V}_X$, we have $\#\mathcal{E} = \#\mathcal{V}_X \#\mathcal{N}(x_{:T})$. \square

Proposition III.12 Let $\tau^* \in \Delta_T(\{0, 1\}^2)$. For all $n \in \mathbb{N}^*$ let $G^n = ((\mathcal{V}_X^n, \mathcal{V}_Y^n), \mathcal{E}^n)$ which is the graph G_{τ^*} when τ^* is seen as an element of $\Delta_{nT}(\{0, 1\}^2)$. We have that, in G^n , $\#\mathcal{N}(x_{:nT})$ does not depend on $x_{:nT}$ and $\#\mathcal{N}(y_{:nT})$ does not depend on $y_{:nT}$. Let $N^n \doteq \#\mathcal{N}(x_{:T})$ for any $x_{:T} \in \mathcal{V}_X^n$ and $N^{nT} \doteq \#\mathcal{N}(y_{:T})$ for any $y_{:T} \in \mathcal{V}_Y^n$, then

$$\begin{aligned} \frac{1}{nT} \log \#\mathcal{V}_X^n &\xrightarrow{n \rightarrow \infty} \mathbb{H}(X), \quad \frac{1}{nT} \log \#\mathcal{V}_Y^n \xrightarrow{n \rightarrow \infty} \mathbb{H}(Y), \quad (7) \\ \frac{1}{nT} \log N^n &\xrightarrow{n \rightarrow \infty} \mathbb{H}(Y|X), \quad \frac{1}{nT} \log N^{nT} \xrightarrow{n \rightarrow \infty} \mathbb{H}(X|Y), \\ \frac{1}{nT} \log \#\mathcal{E}^n &\xrightarrow{n \rightarrow \infty} \mathbb{H}(X, Y), \end{aligned}$$

where all the entropies are computed with $(X, Y) \sim \tau^*$.

The proof follows from [12, Lemma 2.3, pp. 17]. In Fig. 2, we built an example of typicality graph for $T = 4$ and for the

distribution $\tau^* = \begin{pmatrix} 2/4 & 0 \\ 1/4 & 1/4 \end{pmatrix}$. The marginal distributions of τ^* over \mathcal{X} and \mathcal{Y} are $(2/4, 2/4)$ and $(3/4, 1/4)$, respectively, and $\#\mathcal{V}_\mathcal{X} = 6$, $\#\mathcal{V}_\mathcal{Y} = 4$.

IV. ZERO-ERROR CODE CONSTRUCTION

The following proposition is a key ingredient of our code construction. Indeed, it shows how to use the X -coloring to achieve zero-error coding for decoder 1. More precisely, we show that given the X -coloring, the side information $Y_{:T}$, the color of $X_{:T}$, and the joint type are sufficient to retrieve $X_{:T}$.

Proposition IV.1 *Let $\tau^* \in \Delta_T(\{0, 1\}^2)$, and k be an X -coloring of G_{τ^*} . Then conditionally given $\tau(X_{:T}, Y_{:T}) = \tau^*$, we have $X_{:T} \preceq (Y_{:T}, k(X_{:T}))$.*

Proof. Let $y_{:T} \in \mathcal{V}_\mathcal{Y}$ in G_{τ^*} , and $x_{:T} \neq x'_{:T} \in \mathcal{N}(y_{:T})$, then we have $\mathcal{N}(x_{:T}) \cap \mathcal{N}(x'_{:T})$ is nonempty as it contains $y_{:T}$; so by Def. III.9, $k(x_{:T}) \neq k(x'_{:T})$. This is true for all $y_{:T} \in \mathcal{V}_\mathcal{Y}$ in G_{τ^*} , and $x_{:T} \neq x'_{:T} \in \mathcal{N}(y_{:T})$; therefore, $k(X_{:T})$ is a color assigned to exactly one sequence in $\mathcal{N}(Y_{:T})$: this sequence is $X_{:T}$. \square

The main idea behind the construction of the type grouping code is to partition the set of types into groups, and build X -colorings that are valid for all types in a group.

Definition IV.2 (Type grouping code) *A type grouping code is a pair (c, Z) : Z is the index of the group which contains the type $\tau(X_{:T}, Y_{:T})$, i.e. $Z \preceq \tau(X_{:T}, Y_{:T})$; and c is a mapping from $\mathcal{X}^T \times \Delta_T(\{0, 1\}^2)$ to a set of colors \mathcal{C} . The pair (c, Z) must satisfy:*

- ◆ $\forall \tau^* \in \Delta_T(\{0, 1\}^2)$, $c(\cdot, \tau^*)$ is an X -coloring of G_{τ^*} ,
- ◆ for all $z \in \mathcal{Z}$, for all pair of types $\tau^*, \tau' \in \Delta_T(\{0, 1\}^2)$ that have the same marginal distribution on \mathcal{X} such that $\mathbb{P}(\tau(X_{:T}, Y_{:T}) = \tau^* | Z = z)$ and $\mathbb{P}(\tau(X_{:T}, Y_{:T}) = \tau' | Z = z)$ are both positive, $c(\cdot, \tau^*) = c(\cdot, \tau')$.

A type grouping code consists in transmitting on the common channel the pair $(c(X_{:T}, \tau(X_{:T}, Y_{:T})), Z)$ with a Huffman code, and transmitting on the secondary channel $X_{:T}$ knowing $(c(X_{:T}, \tau(X_{:T}, Y_{:T})), Z)$ with a conditional Huffman code.

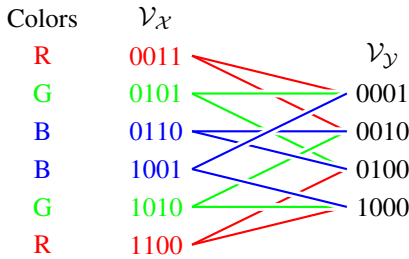


Fig. 2: The graph G_{τ^*} for $T = 4$ and $\tau^* = \begin{pmatrix} 2/4 & 0 \\ 1/4 & 1/4 \end{pmatrix}$. A possible X -coloring is depicted. Note that each sequence of $\mathcal{V}_\mathcal{Y}$ has a unique sequence of each color adjacent to it.

Theorem IV.3 *Let (c, Z) be a type grouping code, then it is zero-error and we have*

$$\begin{aligned} R_1 &= \frac{1}{T} \mathbb{H}(c(X_{:T}, \tau(X_{:T}, Y_{:T})), Z), \\ R_2 &= \frac{1}{T} \mathbb{H}(X_{:T} | c(X_{:T}, \tau(X_{:T}, Y_{:T})), Z). \end{aligned} \quad (8)$$

Proof. By construction, decoder 2 can retrieve $X_{:T}$ with probability 1. We now show that decoder 1 can also retrieve $X_{:T}$. To do so, let $\tau^* \in \Delta_T(\{0, 1\}^2)$, then $c(\cdot, \tau^*)$ is an X -coloring of G_{τ^*} , and by Prop. IV.1, we have $X_{:T} \preceq (Y_{:T}, c(X_{:T}, \tau(X_{:T}, Y_{:T})))$ conditionally given $\tau(X_{:T}, Y_{:T}) = \tau^*$.

Let h_{τ^*} be a mapping such that $X_{:T} = h_{\tau^*}(Y_{:T}, c(X_{:T}, \tau(X_{:T}, Y_{:T})))$ whenever $\tau(X_{:T}, Y_{:T}) = \tau^*$, that is, $X_{:T} = h_{\tau^*}(Y_{:T}, c(X_{:T}, \tau^*))$. Let $z \in \mathcal{Z}$ be the unique element such that $\mathbb{P}(\tau(X_{:T}, Y_{:T}) = \tau^* | Z = z)$ is positive, then we have $X_{:T} = h_{\tau^*}(Y_{:T}, c(X_{:T}, \tau^*))$ for all τ^* such that $\mathbb{P}(\tau(X_{:T}, Y_{:T}) = \tau^* | Z = z)$ is positive.

Thus $X_{:T} \preceq (Y_{:T}, c(X_{:T}, \tau(X_{:T}, Y_{:T})))$ conditionally given $Z = z$, by using the mapping h_{τ^*} . Using Prop. III.3, we have $X_{:T} \preceq (Y_{:T}, c(X_{:T}, \tau(X_{:T}, Y_{:T})), Z)$ and the decoder 1 can retrieve $X_{:T}$ with probability 1.

The coding rates are obtained by using Huffman codes. \square

V. ILLUSTRATIVE EXAMPLE

Definition V.1 *We consider $T = 6$ and we define $\bar{\tau} = \begin{pmatrix} 2/6 & 1/6 \\ 1/6 & 2/6 \end{pmatrix} \in \Delta_6(\mathcal{X} \times \mathcal{Y})$, and for all sequence $x_{:T} \in \mathcal{V}_\mathcal{X}$ in $G_{\bar{\tau}}$, let $k(x_{:T}) = x_{2:T} \oplus (x_1, \dots, x_1)$.*

Proposition V.2 *The mapping k is an X -coloring of $G_{\bar{\tau}}$.*

Proof. We have $2T \sum_{y \in \mathcal{Y}} \min(\bar{\tau}_{0,y}, \bar{\tau}_{1,y}) = 4$. From Corollary III.10, two different x -sequences with the same color must be at distance $d(x_{:T}, x'_{:T}) > 5$. From Lemma 1, their distance $d(x_{:T}, x'_{:T})$ must be even. Therefore, only sequences at distance at least 6 pairwise can share the same color in any X -coloring of $G_{\bar{\tau}}$. Note that 6 is the maximal Hamming distance between any sequences from $\mathcal{V}_\mathcal{X}$, as $T = 6$.

Let $x_{:T}, x'_{:T} \in \mathcal{V}_\mathcal{X}$ such that $k(x_{:T}) = k(x'_{:T})$, then we have either $x_1 = x'_1$, in that case $x_{2:T} \oplus (x_1, \dots, x_1) = x'_{2:T} \oplus (x'_1, \dots, x'_1) \Rightarrow x_{:T} = x'_{:T}$; or $x_1 \neq x'_1$, in that case $x_{2:T} \oplus (x_1, \dots, x_1) = x'_{2:T} \oplus (x'_1, \dots, x'_1) \Rightarrow x_{:T} \oplus x'_{:T} = (1, \dots, 1)$. The sequences that share the same color are at Hamming distance 6, thus k is an X -coloring of $G_{\bar{\tau}}$. \square

In Fig. 3, we illustrate the graph $G_{\bar{\tau}}$, with its X -coloring k . We have that $\bar{\tau}$ has the same marginal type on both \mathcal{X} and \mathcal{Y} : $(3/6, 3/6)$. Using Prop. III.11 there are $\binom{6}{3} = 20$ sequences in both $\mathcal{V}_\mathcal{X}$ and $\mathcal{V}_\mathcal{Y}$ in $G_{\bar{\tau}}$.

For all $x_{:T} \in \mathcal{V}_\mathcal{X}$, there exists a unique sequence $x'_{:T} \in \mathcal{V}_\mathcal{X}$ such that $d(x_{:T}, x'_{:T}) = 6$: $x'_{:T} = x_{:T} \oplus (1, \dots, 1)$. Therefore we have two sequences per color, which gives 10 colors as there are 20 vertices in $\mathcal{V}_\mathcal{X}$. Thus, when $k(X_{:T})$ is given, an additional bit (which is X_1) is required to retrieve $X_{:T}$.

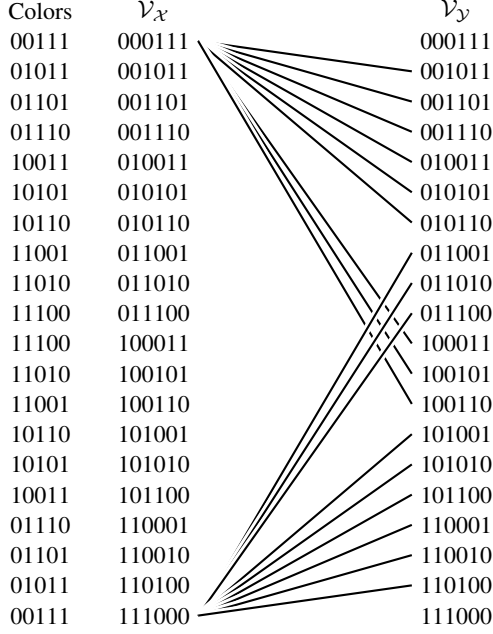


Fig. 3: For the sake of clarity, we represent only a subset of edges of the graph G_{τ} that correspond to two vertices at distance 6. The remaining edges can be drawn similarly by symmetry. We describe the X -coloring k of the sequences from \mathcal{V}_X on the left, the sequences \mathcal{V}_Y are depicted on the right.

Definition V.3 (Coding scheme) Let Z be the auxiliary random variable defined by:

$$Z = \begin{cases} 2 & \text{if } \tau(X_{:T}, Y_{:T}) = \begin{pmatrix} 2/6 & 1/6 \\ 1/6 & 2/6 \end{pmatrix} \\ 1 & \text{if } \tau(X_{:T}, Y_{:T}) = \begin{pmatrix} * & 0 \\ 0 & * \end{pmatrix} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

where $*$ represents any value. Let $c : \mathcal{X}^T \times \Delta_T(\mathcal{X} \times \mathcal{Y}) \rightarrow \mathcal{X}^T$ be the function defined by:

$$c(X_{:T}, \tau(X_{:T}, Y_{:T})) = \begin{cases} k(X_{:T}) & \text{if } \tau(X_{:T}, Y_{:T}) = \begin{pmatrix} 2/6 & 1/6 \\ 1/6 & 2/6 \end{pmatrix} \\ 0, \dots, 0 & \text{if } \tau(X_{:T}, Y_{:T}) = \begin{pmatrix} * & 0 \\ 0 & * \end{pmatrix} \\ X_{:T} & \text{otherwise} \end{cases} \quad (10)$$

where k is the coloring function defined in Def. V.1. We define the following coding scheme:

- ◆ On the common channel, $c(X_{:T}, \tau(X_{:T}, Y_{:T})), Z$ are jointly transmitted using a Huffman code.
- ◆ On the secondary channel, $X_{:T}$ is transmitted using a conditional Huffman code knowing $c(X_{:T}, \tau(X_{:T}, Y_{:T})), Z$.

Theorem V.4 (Achieved rates) The coding scheme (c, Z) defined in Def. V.3 is a type grouping code that achieves the rates:

$$R_1 = 0.5786, \quad R_2 = 0.5345. \quad (11)$$

The numerical values are obtained with the source distribution $p^{X,Y} = \begin{pmatrix} 9/20 & 1/20 \\ 1/20 & 9/20 \end{pmatrix}$ and $T = 6$.

The proof of Theorem V.4 is provided in [11].

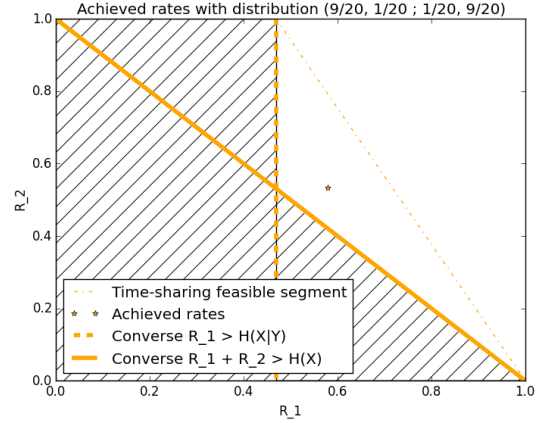


Fig. 4: The performances of the type grouping code of Def. V.3 are below the time-sharing feasible segment.

VI. EXTENSION

The following theorem states that for $T = 1$, it is possible to design an optimal code adapted for arbitrary finite sets \mathcal{X} and \mathcal{Y} , if we know optimal codes for \mathcal{X} binary and \mathcal{Y} arbitrary. The adaptation for $T > 1$ is under study. In this section we index $\mathcal{X} = \{1, \dots, \#\mathcal{X}\}$.

Lemma 2 $\sum_{i < \#\mathcal{X}} \mathbb{H}(\mathbb{1}_{X=i} | Y, X \geq i) \mathbb{P}(X \geq i) = \mathbb{H}(X|Y)$

The proof of Lemma 2 is provided in [11].

Theorem VI.1 Let $(f_i)_{i < \#\mathcal{X}}, (g_i)_{i < \#\mathcal{X}}$ be two families of zero-error variable-length 1-source codes adapted for the probability distribution $(p^{(\mathbb{1}_{X=i}, Y) | X \geq i})_{1 \leq i < \#\mathcal{X}}$. If we have for all $i < \#\mathcal{X}$

$$\begin{aligned} \mathbb{E}_{X,Y} [|f_i(\mathbb{1}_{X=i}, Y)| | X \geq i] &= \mathbb{H}(\mathbb{1}_{X=i} | Y, X \geq i) \\ \mathbb{E}_{X,Y} [|g_i(\mathbb{1}_{X=i}, Y)| | X \geq i] &= \mathbb{I}(\mathbb{1}_{X=i}; Y | X \geq i) \end{aligned} \quad (12)$$

then the 1-source codes

$$\begin{aligned} \bar{f} : \mathcal{X} \times \mathcal{Y} &\rightarrow \{0, 1\}^* \\ (x, y) &\mapsto f_1(0, y) f_2(0, y) \dots f_{x-1}(0, y) f_x(1, y) \\ \bar{g} : \mathcal{X} \times \mathcal{Y} &\rightarrow \{0, 1\}^* \\ (x, y) &\mapsto g_1(0, y) g_2(0, y) \dots g_{x-1}(0, y) g_x(1, y) \end{aligned} \quad (13)$$

have a rate of $\mathbb{H}(X|Y)$ and $\mathbb{I}(X; Y)$, respectively.

The proof of Theorem VI.1 is provided in [11].

REFERENCES

- [1] A. Kaspi, "Rate-distortion function when side-information may be present at the decoder," *IEEE Transactions on Information Theory*, vol. 40, no. 6, pp. 2031–2034, Nov. 1994. [Online]. Available: <http://ieeexplore.ieee.org/document/340475/>
- [2] N. M. Bidgoli, T. Maugey, and A. Roumy, "Fine granularity access in interactive compression of 360-degree images based on rate-adaptive channel codes," *IEEE Transactions on Multimedia*, 2020. [Online]. Available: <https://hal.inria.fr/hal-02946795>
- [3] S. C. Draper, "Universal incremental Slepian-Wolf coding," in *42nd Annual Allerton Conf*, 2004.
- [4] T. Maugey, A. Roumy, E. Dupraz, and M. Kieffer, "Incremental Coding for Extractable Compression in the Context of Massive Random Access," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 6, pp. 251–260, Dec. 2020. [Online]. Available: <https://hal.inria.fr/hal-02542972>
- [5] F. Ye, E. Dupraz, Z. Mheich, and K. Amis, "Optimized Rate-Adaptive Protograph-Based LDPC Codes for Source Coding With Side Information," *IEEE Transactions on Communications*, vol. 67, no. 6, pp. 3879–3889, Jun. 2019. [Online]. Available: <https://ieeexplore.ieee.org/document/8662618/>
- [6] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973. [Online]. Available: <http://ieeexplore.ieee.org/document/1055037/>
- [7] A. El Gamal and A. Orlitsky, "Interactive data compression," in *25th Annual Symposium on Foundations of Computer Science, 1984*. IEEE, 1984, pp. 100–108.
- [8] P. Koulgi, E. Tuncel, S. Regunathan, and K. Rose, "On zero-error source coding with decoder side information," *IEEE Transactions on Information Theory*, vol. 49, no. 1, pp. 99–111, Jan. 2003. [Online]. Available: <http://ieeexplore.ieee.org/document/1159765/>
- [9] E. Tuncel, J. Nayak, P. Koulgi, and K. Rose, "Zero-error Distributed source coding," in *Distributed source coding: theory, algorithms, and applications*. Elsevier, 2009.
- [10] A. Nazari, D. Krithivasan, S. S. Pradhan, A. Anastasopoulos, and R. Venkataramanan, "Typicality graphs and their properties," in *2010 IEEE International Symposium on Information Theory*, 2010, pp. 520–524.
- [11] A. R. N. Charpenay, M. Le Treust, "Zero-error source coding when side information may be present - technical report," [Online]. Available: https://drive.google.com/drive/folders/1wlHghok8eM8P9pWhI8BG2qHRKGsp4Pz_?usp=sharing, May 2021.
- [12] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.