



**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. IZS 2022 - International Zurich Seminar on Information and Communication, Mar 2022, Zurich, Switzerland. pp.1-5. hal-03290860v2

**HAL Id: hal-03290860**

**<https://hal.science/hal-03290860v2>**

Submitted on 16 Dec 2021

**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<sup>\*</sup>, Maël Le Treust<sup>†</sup>, Aline Roumy<sup>‡</sup>

<sup>\*</sup>IRISA, 263 avenue du Général Leclerc, 35 042 Rennes cedex, France, nicolas.charpenay@irisa.fr

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

<sup>‡</sup>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 such a scenario, the side information may represent an image that could have been requested previously by the user. We aim at designing a two layer zero-error coding scheme that adapts to the presence or not of the side information at the decoder. The scenario we consider involves two decoders and two noiseless channels, the first channel to both decoder and the second channel of additional information to decoder 2 only. The side information is available at the encoder and decoder 1, but not at decoder 2. By using a random coding argument we characterize the zero-error achievable rate region. The code construction relies on coset partitioning obtained from a linear code. The encoder sends the coset of the source sequence on the first channel to all decoders, and sends the index of the source sequence in its coset on the second channel to decoder 2.

## I. INTRODUCTION

We consider the scenario described in Fig. 1 in which the information source  $X$  is correlated to the side information (SI)  $Y$  observed by the encoder and decoder 1 only. The information is sent through a noiseless channel at rate  $R_1$  to both decoders and an additional noiseless channel at rate  $R_2$  to decoder 2, which does not observe the SI. All decoders must recover the source  $X$  with zero-error, i.e. with a probability of error equal to zero, which is a more restrictive assumption than a vanishing probability of error.

This scenario arises in interactive compression, where the user can randomly access part of the data directly in the compressed domain. A source sequence  $X^n$  models the smallest entity that can be requested, for instance a file of a database, a frame of a video, or a block of an omnidirectional image [1]. Upon request of  $X^n$ , and if no request has been previously made (case of decoder 2 in Fig. 1), the encoder sends the complete representation of the data  $(f_1(X^n, Y^n), f_2(X^n, Y^n))$  at rate  $R_1 + R_2$ . If, instead, the block  $Y^n$  has already been requested (case of decoder 1), the encoder sends part of the complete compressed representation namely  $f_1(X^n, Y^n)$  to complete  $Y^n$ . Moreover, we consider the zero-error version of this problem, as zero-error source coding is a fundamental building block of practical video coding schemes. We therefore seek for the set of rates  $(R_1, R_2)$ , which can be achieved in this scenario.

A way to achieve zero-error coding is to use conditional coding, and send the source  $X$  to decoder 1 at rate  $R_1 =$

$H(X|Y)$ , since both encoder and decoder 1 observe the SI  $Y$ . Then, to recover the source  $X$ , decoder 2 needs to obtain the SI  $Y$ , which requires a rate of  $R_2 = H(Y)$ . To reduce  $R_2$ , one can construct a scheme based on Slepian and Wolf (SW) random binning argument [2] for lossless source coding with side information at the decoder only. First, the encoder sends to both decoders the bin index, to which the  $X$  sequence belongs at rate  $R_1 = H(X|Y)$ , which corresponds to the data sent by the SW encoder. Then, a complement is sent to decoder 2, which is the index of the  $X$  sequence in the bin. One can deduce from the random binning argument, that this requires a rate  $R_2 = H(X) - H(X|Y) = I(X; Y)$ , see [3, Sec. 10.3].

These rates however hold for a vanishing error probability only. Indeed, for some probability distributions, the zero-error SW scheme requires to send at rate  $H(X)$  to the decoder with side information, see [4]–[6]. However, the SW scheme does not use the side information knowledge at the encoder. Therefore, we study the role of the side information at the encoder with a zero-error constraint when side information may be present at the decoder.

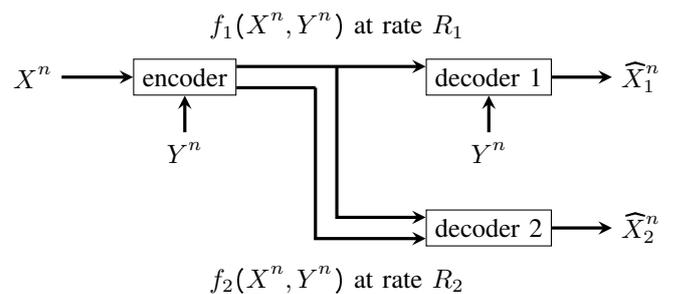


Fig. 1: Source coding when side-information may be present.

In this paper, we characterize the set of rate pairs that are achievable with zero-error source codes, as depicted in Fig. 1. More precisely, we show that the pair of rates  $(R_1, R_2) = (H(X|Y), I(X; Y))$  is achievable and moreover, it is the corner-point of the set of achievable pair of rates. Our achievability result relies on a random coding argument. We use Csiszar and Körner’s method of types [7, Chapter 2] in order to calibrate a linear code which is used to partition the set of source sequences. The encoder sends the coset of the source sequence to all decoders and the index of the source

sequence in its coset to decoder 2. We show that the zero-error property is satisfied and the corresponding rates converge to the pair of target rates  $(H(X|Y), I(X; Y))$ .

#### A. Notations

Random variables and their realizations are represented by uppercase letters (e.g.,  $X$ ) and lowercase letters (e.g.,  $x$ ), respectively; and their set of possible values with the corresponding calligraphic letters (e.g.,  $\mathcal{X}$ ). We denote by  $|\cdot|$  the cardinality of a set. We denote a sequence of symbols by  $x^n = (x_1, \dots, x_n)$ . The set of probability distributions over a finite set  $\mathcal{X}$  is denoted by  $\mathcal{P}(\mathcal{X})$ . The distribution of a random variable  $X$  is denoted by  $P_X \in \mathcal{P}(\mathcal{X})$ . The conditional distribution of a random variable  $X$  knowing  $Y$  is denoted by  $P_{X|Y}$ , and the joint distribution is denoted by  $P_{X,Y}$ . We denote by  $\{0, 1\}^*$  the set of binary words. Throughout the paper the logarithms are in base two.

### II. PROBLEM STATEMENT AND MAIN RESULT

The setting of Fig. 1 is described by:

- Two finite sets  $\mathcal{X}, \mathcal{Y}$  and a pair of random variables  $(X, Y) \in \mathcal{X} \times \mathcal{Y}$  drawn with the distribution  $P_{X,Y}$ .
- An encoder that observes the realizations of  $(X, Y)$ .
- Two decoders, where only decoder 1 observes the realizations of the side-information  $Y$ .
- The encoder transmits over a first channel to both decoders, and a second channel to decoder 2 only.
- We denote by  $n \in \mathbb{N}^* = \mathbb{N} \setminus \{0\}$  the block size of the coding scheme. For  $n$  iterated source uses, we denote by  $(X^n, Y^n)$  the sequences of independent copies of  $(X, Y)$ .

**Definition II.1** Given  $n \in \mathbb{N}^* = \mathbb{N} \setminus \{0\}$ ,  $(R_1^{(n)}, R_2^{(n)}) \in [0, +\infty)^2$ , a  $(n, R_1^{(n)}, R_2^{(n)})$ -zero-error source code consists of encoding functions  $(f_1, f_2)$  that assigns variable-length binary sequences and decoding functions  $(g_1, g_2)$  defined by:

$$f_1 : \mathcal{X}^n \times \mathcal{Y}^n \rightarrow \{0, 1\}^*, \quad f_2 : \mathcal{X}^n \times \mathcal{Y}^n \rightarrow \{0, 1\}^*, \quad (1)$$

$$g_1 : \{0, 1\}^* \times \mathcal{Y}^n \rightarrow \mathcal{X}^n, \quad g_2 : (\{0, 1\}^*)^2 \rightarrow \mathcal{X}^n, \quad (2)$$

that satisfy

$$R_1^{(n)} = \frac{1}{n} \mathbb{E}[l(f_1(X^n, Y^n))], \quad R_2^{(n)} = \frac{1}{n} \mathbb{E}[l(f_2(X^n, Y^n))],$$

where  $l(\cdot)$  denotes the length of a binary word, and that satisfy the zero-error property, i.e.  $X^n = g_1(f_1(X^n, Y^n), Y^n) = g_2(f_1(X^n, Y^n), f_2(X^n, Y^n))$  with probability 1.

**Definition II.2** A rate pair  $(R_1, R_2) \in [0, +\infty)^2$  is achievable if there exists a sequence of  $(n, R_1^{(n)}, R_2^{(n)})$ -zero-error source codes such that

$$\lim_n R_1^{(n)} = R_1, \quad \lim_n R_2^{(n)} = R_2. \quad (3)$$

We denote by  $\mathcal{R}$  the zero-error achievable rate region.

### Theorem II.3

$$\mathcal{R} = \left\{ (R_1, R_2), R_1 \geq H(X|Y), R_1 + R_2 \geq H(X) \right\}. \quad (4)$$

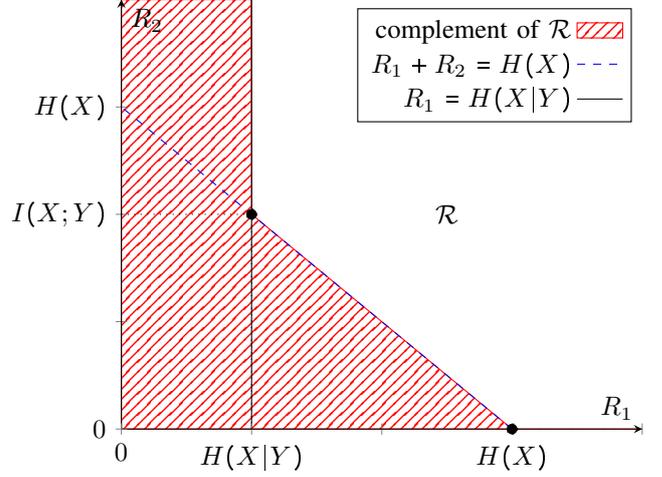


Fig. 2: Zero-error achievable rate region  $\mathcal{R}$ .

*Proof.* [Converse of Theorem II.3] In this setting, each decoder must retrieve  $X$  with zero-error. Using Shannon lossless source coding result [8, Theorem 5.3.1] and Slepian-Wolf Theorem [2, Theorem 2] on each decoder, we have  $R_1 \geq H(X|Y)$  and  $R_1 + R_2 \geq H(X)$ , as the zero-error source codes are a subclass of lossless codes considered for these converses.  $\square$

### III. ACHIEVABILITY PROOF OF THEOREM II.3

In order to complete the proof of Th. II.3, we show that

$$(H(X|Y), I(X; Y)) \in \mathcal{R}. \quad (5)$$

In order to complete the achievability result we use a time sharing with the point  $(H(X), 0)$ , which is known to be achievable by compressing  $X$  using a Huffman code and sending the resulting binary sequence via  $f_1$ .

#### A. Preliminaries

**Definition III.1 (Type)** For all pair of sequences  $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ , the joint type is the distribution from  $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$  denoted  $T_{x^n, y^n}$  that satisfies for all  $(x', y') \in \mathcal{X} \times \mathcal{Y}$

$$T_{x^n, y^n}(x', y') = \frac{1}{n} \left| \left\{ i \leq n \mid (x_i, y_i) = (x', y') \right\} \right|. \quad (6)$$

We denote the marginal types by  $T_{x^n}$  and  $T_{y^n}$ , respectively. The conditional type of  $x^n$  knowing  $y^n$  is denoted  $T_{x^n|y^n}$  and is defined for all  $(x', y') \in \mathcal{X} \times \mathcal{Y}$  by

$$T_{x^n|y^n}(x'|y') = \frac{T_{x^n, y^n}(x', y')}{T_{y^n}(y')} = \frac{T_{x^n, y^n}(x', y')}{\sum_{x''} T_{x^n, y^n}(x'', y')}. \quad (7)$$

We denote by  $T_{X^n, Y^n}$  the random variable of the joint type of the random sequences  $(X^n, Y^n)$ . We denote the random variables of their conditional and marginal types by  $T_{X^n|Y^n}$ ,  $T_{X^n}$  and  $T_{Y^n}$ , respectively.

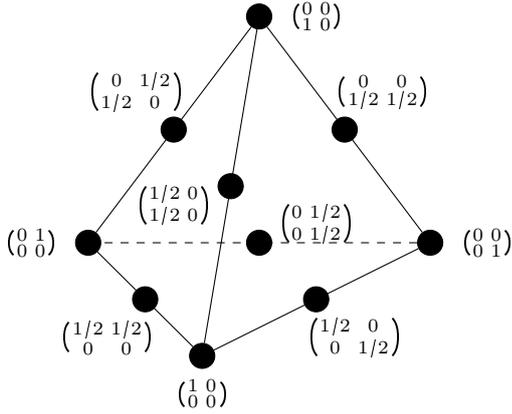


Fig. 3: The 2-discretized probability simplex  $\mathcal{P}_2(\mathcal{X} \times \mathcal{Y})$  for  $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ .

**Definition III.2 (Entropy of a type)** For all  $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ , we define the entropy of the type  $T_{x^n, y^n}$  by:

$$H_{T_{x^n, y^n}}(X, Y) = - \sum_{(x', y')} T_{x^n, y^n}(x', y') \log T_{x^n, y^n}(x', y'),$$

$$H_{T_{x^n, y^n}}(X|Y) = - \sum_{(x', y')} T_{x^n, y^n}(x', y') \log T_{x^n|y^n}(x'|y').$$

For random sequences  $(X^n, Y^n)$ , we denote by  $H_{T_{X^n, Y^n}}(X, Y)$  (resp.  $H_{T_{X^n, Y^n}}(X|Y)$ ), the random entropy computed with respect to the random type  $T_{X^n, Y^n}$ .

**Definition III.3 ( $n$ -discretized simplex)** The  $n$ -discretized probability simplex  $\mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$  is the set of types that are achievable using sequences of length  $n$ ,

$$\mathcal{P}_n(\mathcal{X} \times \mathcal{Y}) \doteq \left\{ \pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \mid \forall (x', y'), n\pi(x', y') \in \mathbb{N} \right\}.$$

**Definition III.4 (Type class,  $V$ -shell)** For all type  $\pi \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$ , we denote the type class by  $\mathcal{T}_\pi$

$$\mathcal{T}_\pi = \left\{ (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n \mid T_{x^n, y^n} = \pi \right\}. \quad (8)$$

Given a conditional type  $V \in \mathcal{P}(\mathcal{X})^{|\mathcal{Y}|}$ , the  $V$ -shell of a sequence  $y^n$  is the set

$$\mathcal{T}_V(y^n) = \left\{ x^n \in \mathcal{X} \mid T_{x^n|y^n} = V \right\}. \quad (9)$$

**Definition III.5 (Generator/parity matrix, syndrome, coset)** By extending  $P_{X, Y} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$  with zeros, we assume without loss of generality that  $|\mathcal{X}|$  is a prime number, thus giving  $\mathcal{X} \simeq \mathbb{Z}/|\mathcal{X}|\mathbb{Z}$  a field structure. For all  $n, k \in \mathbb{N}^*$ , we denote by  $\mathcal{M}_{n, k}(\mathcal{X})$  the set of  $n \times k$  matrices over the finite field  $\mathcal{X}$ .

Let  $k \in \mathbb{N}^*$ , a generator matrix is a matrix  $\mathbf{G} \in \mathcal{M}_{n, k}(\mathcal{X})$ . An associated parity matrix is a matrix  $\mathbf{H} \in \mathcal{M}_{n-k, n}(\mathcal{X})$  such that  $\text{Im } \mathbf{G} = \text{Ker } \mathbf{H}$ , where  $\text{Im}$  and  $\text{Ker}$  denote the image and the kernel, respectively.

The syndrome of a sequence  $x^n$  is  $\mathbf{H}x^n$ . The coset associated to the syndrome  $\mathbf{H}x^n$  is the set

$$\text{Im } \mathbf{G} + x^n = \{ \tilde{x}^n \in \mathcal{X}^n \mid \mathbf{H}\tilde{x}^n = \mathbf{H}x^n \}. \quad (10)$$

### B. Coding scheme

For all  $n \in \mathbb{N}^*$ , we show the existence of a sequence of  $(n, R_1^{(n)}, R_2^{(n)})$ -zero-error source codes that achieves the corner-point  $(H(X|Y), I(X; Y))$  of the zero-error rate region  $\mathcal{R}$ . Our proof is based on a linear code adjusted depending on  $T_{X^n, Y^n}$ , and coset partitioning of the Hamming space.

We assume w.l.o.g. that  $P_{X, Y} \neq P_X P_Y$ . We fix the block-length  $n$  and a constant parameter  $\delta \in (0; \log |\mathcal{X}| - H(X|Y))$  that will represent a rate penalty.

- Random code generation: For each pair of sequences  $(x^n, y^n)$ , we define the parameter

$$k \doteq \left\lceil n - n \frac{H_{T_{x^n, y^n}}(X|Y) + \delta}{\log |\mathcal{X}|} \right\rceil^+. \quad (11)$$

where  $\lceil \cdot \rceil$  denotes the ceiling function and  $(\cdot)^+$  denotes  $\max(\cdot, 0)$ . We denote by  $K$  the random variable induced by  $k$ , for the random sequences  $(X^n, Y^n)$ . A generator matrix  $\mathbf{G} \in \mathcal{M}_{n, n}(\mathcal{X})$  is randomly drawn, with i.i.d. entries drawn according to the uniform distribution on  $\mathcal{X}$ . If  $K \neq 0$ , let  $\mathbf{G}_K$  be the matrix obtained by extracting the  $K$  first lines of  $\mathbf{G}$ , and  $\mathbf{H}_K$  a parity matrix associated to  $\mathbf{G}_K$ .

The random code  $\mathcal{C}$  consists of the set of random matrices  $\mathcal{C} = \{(\mathbf{G}_k, \mathbf{H}_k), 1 \leq k \leq n\}$ . Before the transmission starts, a code realization is chosen and revealed to the encoder and both decoders.

- Encoding function  $f_1$ : Let  $E \in \{0, 1\}$  be such that  $E = 0$  if  $K \neq 0$  and  $(\text{Im } \mathbf{G}_K + X^n) \cap \mathcal{T}_{T_{X^n|Y^n}}(Y^n) = \{X^n\}$ ;  $E = 1$  otherwise. Then we define

$$f_1(X^n, Y^n) = \begin{cases} b(T_{X^n, Y^n}, E, \mathbf{H}_K X^n) & \text{if } E = 0, \\ b(T_{X^n, Y^n}, E, X^n) & \text{if } E = 1, \end{cases} \quad (12)$$

where  $b(\cdot)$  denotes the binary expansion.

- Encoding function  $f_2$ : If  $E = 0$ , the index of  $X^n$  in its coset  $\text{Im } \mathbf{G}_K + X^n$  is compressed using a Huffman code with the distribution  $P_{X^n}$ . Let  $B(\mathbf{G}_K, X^n, Y^n)$  be the resulting binary sequence, then we set

$$f_2(X^n, Y^n) = B(\mathbf{G}_K, X^n, Y^n). \quad (13)$$

Otherwise,  $f_2(X^n, Y^n) = 0$ .

- Decoding function  $g_1$ : It observes  $f_1(X^n, Y^n)$  and extracts  $E$  and  $T_{X^n, Y^n}$ . If  $E = 1$ ,

$$g_1(f_1(X^n, Y^n), Y^n) = X^n. \quad (14)$$

Otherwise  $E = 0$ , it extracts  $\mathbf{H}_K X^n$  and determines the coset  $\text{Im } \mathbf{G}_K + X^n$ . Moreover, by using  $T_{X^n, Y^n}$  and  $Y^n$  it determines the  $T_{X^n|Y^n}$ -shell  $\mathcal{T}_{T_{X^n|Y^n}}(Y^n)$ , and therefore returns an element

$$g_1(f_1(X^n, Y^n), Y^n) \in (\text{Im } \mathbf{G}_K + X^n) \cap \mathcal{T}_{T_{X^n|Y^n}}(Y^n).$$

- ◆ Decoding function  $g_2$ : It observes  $f_1(X^n, Y^n)$  and extracts  $E$  and  $T_{X^n, Y^n}$ . If  $E = 0$ , it extracts  $\mathbf{H}_K X^n$  and determines the coset  $\text{Im } \mathbf{G}_K + X^n$ , and it returns  $g_2(f_1(X^n, Y^n), f_2(X^n, Y^n))$ , the element of  $\text{Im } \mathbf{G}_K + X^n$  with index  $f_2(X^n, Y^n)$ . If  $E = 1$ , it returns

$$g_2(f_1(X^n, Y^n), f_2(X^n, Y^n)) = X^n.$$

**Remark III.6** The parameter  $K$  is selected so that when  $K > 0$ , the number of parity bits of the linear code asymptotically matches the conditional entropy:

$$\frac{(n - K) \log |\mathcal{X}|}{n} = H_{T_{X^n, Y^n}}(X|Y) + \delta + O\left(\frac{1}{n}\right). \quad (15)$$

### C. Zero-error property

We now prove that the code built in Section III-B satisfies the zero-error property. It is clear that both decoders retrieve  $X^n$  with zero-error when  $E = 1$ .

If  $E = 0$ , then by definition of  $E$  we have  $(\text{Im } \mathbf{G}_K + X^n) \cap \mathcal{T}_{T_{X^n, Y^n}}(Y^n) = \{X^n\}$ , hence  $g_1(f_1(X^n, Y^n), Y^n) = X^n$  with probability 1. On the other hand,  $f_2(X^n, Y^n) = B(\mathbf{G}_K, X^n, Y^n)$ , so the element of  $\text{Im } \mathbf{G}_K + X^n$  with index  $f_2(X^n, Y^n)$  is  $X^n$ . Thus,  $g_2(f_1(X^n, Y^n), f_2(X^n, Y^n)) = X^n$  with probability 1.

### D. Rate analysis

Now we prove that for all parameter  $\delta > 0$ , the sequence of rates of the codes built in Section III-B satisfy

$$R_1^{(n)} \xrightarrow{n \rightarrow \infty} H(X|Y) + \delta, \quad R_2^{(n)} \xrightarrow{n \rightarrow \infty} I(X; Y). \quad (16)$$

**Lemma 1 (Large deviations)** Let  $X^l$  be a random variable such that  $P_{X^l}$  is the uniform distribution over  $\mathcal{X}$ . Then for each pair of sequences  $(x^n, y^n)$ , we have:

$$\Pr(T_{X^n, y^n} = T_{x^n, y^n}) = 2^{nH_{T_{x^n, y^n}}(X|Y) - n \log |\mathcal{X}| + o(n)} \quad (17)$$

*Proof.* Since  $P_{X^l}$  is uniform:

$$\Pr(T_{X^n, y^n} = T_{x^n, y^n}) = |\mathcal{X}|^{-n} \left| \mathcal{T}_{T_{x^n, y^n}}(y^n) \right| \quad (18)$$

$$= 2^{-n \log |\mathcal{X}|} 2^{nH_{T_{x^n, y^n}}(X|Y) + o(n)},$$

as [7, Lemma 2.5] gives the asymptotic size of the  $T_{x^n, y^n}$ -shell  $\mathcal{T}_{T_{x^n, y^n}}(y^n)$ .

□

**Probability of decoding ambiguity.** We need to estimate  $\Pr(E = 1)$ . We have  $E = 1$  iff  $K = 0$  or there exists  $(\alpha_1, \dots, \alpha_K) \in \mathcal{X}^K \setminus \{0, \dots, 0\}$  such that  $T_{(X^n + \sum_{i \leq K} \alpha_i \mathbf{G}_K^{(i)})} Y^n = T_{X^n, Y^n}$ , where  $\mathbf{G}_K^{(i)}$  denotes the  $i$ -th column of  $\mathbf{G}_K$ . Thus

$$\Pr(E = 1) \leq \Pr(K = 0) \quad (19)$$

$$+ \Pr\left(\bigcup_{\substack{\alpha \in \mathcal{X}^K \\ \alpha \neq 0}} \left[ T_{(X^n + \sum_{i \leq K} \alpha_i \mathbf{G}_K^{(i)})} Y^n = T_{X^n, Y^n} \right] \middle| K \neq 0\right).$$

We provide an upper bound on the second term in (19). For all  $(x^n, y^n)$  such that  $k \neq 0$ , we have:

$$\Pr\left(\bigcup_{\substack{\alpha \in \mathcal{X}^k \\ \alpha \neq 0}} \left[ T_{(x^n + \sum_{i \leq k} \alpha_i \mathbf{G}_k^{(i)})} y^n = T_{x^n, y^n} \right]\right)$$

$$\leq \sum_{\substack{\alpha \in \mathcal{X}^k \\ \alpha \neq 0}} \Pr\left(T_{(x^n + \sum_{i \leq k} \alpha_i \mathbf{G}_k^{(i)})} y^n = T_{x^n, y^n}\right) \quad (20)$$

$$\leq |\mathcal{X}|^k 2^{nH_{T_{x^n, y^n}}(X|Y) - n \log |\mathcal{X}| + o(n)} \quad (21)$$

$$\leq 2^{n \log |\mathcal{X}| - nH_{T_{x^n, y^n}}(X|Y) - \delta n + o(n)} \\ \times 2^{nH_{T_{x^n, y^n}}(X|Y) - n \log |\mathcal{X}| + o(n)} \leq 2^{-\delta n + o(n)}, \quad (22)$$

where (21) comes from Lemma 1 and (22) comes from (11). Therefore,

$$\Pr\left(\bigcup_{\substack{\alpha \in \mathcal{X}^K \\ \alpha \neq 0}} \left[ T_{(X^n + \sum_{i \leq K} \alpha_i \mathbf{G}_K^{(i)})} Y^n = T_{X^n, Y^n} \right] \middle| K \neq 0\right)$$

$$= \sum_{x^n, y^n} \Pr\left((X^n, Y^n) = (x^n, y^n) \middle| K \neq 0\right)$$

$$\times \Pr\left(\bigcup_{\substack{\alpha \in \mathcal{X}^K \\ \alpha \neq 0}} \left[ T_{(X^n + \sum_{i \leq K} \alpha_i \mathbf{G}_K^{(i)})} Y^n = T_{X^n, Y^n} \right] \right.$$

$$\left. \middle| K \neq 0, (X^n, Y^n) = (x^n, y^n) \right) \quad (23)$$

$$\leq \sum_{x^n, y^n} \Pr\left((X^n, Y^n) = (x^n, y^n) \middle| K \neq 0\right) 2^{-\delta n + o(n)} \quad (24)$$

$$\leq 2^{-\delta n + o(n)}, \quad (25)$$

where (24) comes from (22) and the fact that  $\mathbf{G}$  is independent of  $(X, Y)$ .

We now provide an upper bound on the first term in (19).

$$\mathcal{S} \doteq \left\{ \pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}), \quad 1 - \frac{H_\pi(X|Y) + \delta}{\log |\mathcal{X}|} \leq 0 \right\}. \quad (26)$$

Then we have:

$$\Pr(K = 0) \quad (27)$$

$$= \Pr\left(\left[ n - n \frac{H_{T_{X^n, Y^n}}(X|Y) + \delta}{\log |\mathcal{X}|} \right]^+ = 0\right) \quad (28)$$

$$= \Pr\left(n - n \frac{H_{T_{X^n, Y^n}}(X|Y) + \delta}{\log |\mathcal{X}|} \leq 0\right) \quad (29)$$

$$= \Pr(T_{X^n, Y^n} \in \mathcal{S}) \quad (30)$$

$$= \sum_{\pi \in \mathcal{S} \cap \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})} \Pr(T_{X^n, Y^n} = \pi) \quad (31)$$

$$\leq |\mathcal{S} \cap \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})| \sup_{\pi \in \mathcal{S} \cap \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})} \Pr(T_{X^n, Y^n} = \pi) \quad (32)$$

$$\leq |\mathcal{S} \cap \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})| \sup_{\pi \in \mathcal{S} \cap \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})} 2^{-nD(\pi \| P_{X, Y})} \quad (33)$$

$$\leq |\mathcal{S} \cap \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})| \sup_{\pi \in \mathcal{S}} 2^{-nD(\pi \| P_{X, Y})} \quad (34)$$

$$\leq 2^{-n \inf_{\pi \in \mathcal{S}} D(\pi \| P_{X,Y}) + o(n)}, \quad (35)$$

where (33) comes from [7, Lemma 2.6]. Since  $P_{X,Y} \notin \mathcal{S}$  by definition of  $\delta$ , we have  $\inf_{\pi \in \mathcal{S}} D(P_{X,Y} \| \pi) > 0$ . Thus there exists a positive constant  $\beta > 0$  such that

$$\Pr(K = 0) \leq 2^{-\beta n + o(n)}. \quad (36)$$

Thus by combining (19), (25), (36), we have:

$$\Pr(E = 1) \leq 2^{-\delta n + o(n)} + 2^{-\beta n + o(n)}. \quad (37)$$

**Rate on the common channel.** The encoding function  $f_1$  defined in (12) returns  $T_{X^n, Y^n}$  and  $E$ . When  $E = 0$ , it sends the syndrome  $\mathbf{H}_K X^n$  at rate  $\frac{n-K}{n} \log |\mathcal{X}|$ , otherwise, it sends  $X^n$ . Therefore,

$$\begin{aligned} nR_1^{(n)} &= 1 + |\mathcal{X}||\mathcal{Y}| \log_2(n+1) + \Pr(E=1)n \log |\mathcal{X}| \\ &\quad + \Pr(E=0) \sum_{x^n, y^n} \Pr((X^n, Y^n) = (x^n, y^n) | E=0) \\ &\quad \times (n-k) \log |\mathcal{X}| \end{aligned} \quad (38)$$

$$\leq 1 + |\mathcal{X}||\mathcal{Y}| \log_2(n+1) + \Pr(E=1)n \log |\mathcal{X}| + (n - \mathbb{E}[K]) \log |\mathcal{X}| \quad (39)$$

$$\leq 1 + |\mathcal{X}||\mathcal{Y}| \log_2(n+1) + \Pr(E=1)n \log |\mathcal{X}| + n\mathbb{E}[H_{T_{X^n, Y^n}}(X|Y)] + n\delta + 1, \quad (40)$$

where (39) comes from  $n - k \geq 0$  for all  $(x^n, y^n)$ , and (40) comes from (11).

By the law of large numbers [8, Theorem 11.2.1]  $\mathbb{E}[H_{T_{X^n, Y^n}}(X|Y)] \xrightarrow{n \rightarrow \infty} H(X|Y)$ , and by using (37), we obtain

$$\lim_{n \rightarrow \infty} R_1^{(n)} \leq H(X|Y) + \delta. \quad (41)$$

**Rate on the secondary channel.** The encoding function  $f_2$  is defined in (13). If  $E = 0$ , then  $K \neq 0$  and the encoder transmits the index of  $X^n$  in its coset. The Huffman algorithm has an average output length  $R_2^{(n)}$  that satisfies

$$\begin{aligned} R_2^{(n)} &\leq \frac{1}{n} \left( 1 + \sum_{k \neq 0} \Pr(K = k | E = 0) \right) \\ &\quad \times H(X^n | \mathbf{H}_k X^n, K = k, \mathcal{C}, E = 0) \end{aligned} \quad (42)$$

$$\begin{aligned} &= \frac{1}{n} + \frac{1}{n} H(X^n | K, \mathcal{C}, E = 0) \\ &\quad - \frac{1}{n} H(\mathbf{H}_K X^n | K, \mathcal{C}, E = 0), \end{aligned} \quad (43)$$

where (43) follows from the fact that  $\mathbf{H}_K X^n$  is a deterministic function of  $X^n$ , given a random code  $\mathcal{C}$ .

We now provide an upper bound to the last term  $-\frac{1}{n} H(\mathbf{H}_K X^n | K, \mathcal{C}, E = 0)$  in (43). To do so, we introduce a new encoding scheme that first encodes the sequences  $X^n$  and  $Y^n$  with the encoding function  $f_1$ , and then encode the output by using an entropy coder. The rate of this code  $r$  is upperbounded by  $H(f_1(X^n, Y^n) | \mathcal{C}) + 1$ . Moreover, since decoder 1 retrieves  $X^n$  with zero error (see Sec. III-C), and since the entropy coder is also lossless, the rate of this

code must necessarily be greater than the rate achieved by a conditional entropy coder that compresses  $X^n$ , while knowing the side information  $Y^n$  at both encoder and decoder 2, whose rate is lower bounded by  $nH(X|Y)$ . Therefore, we have

$$nH(X|Y) \leq r < H(f_1(X^n, Y^n) | \mathcal{C}) + 1 \quad (44)$$

$$\begin{aligned} &= 1 + H(T_{X^n, Y^n}, E | \mathcal{C}) \\ &\quad + \Pr(E = 0) H(\mathbf{H}_K X^n | T_{X^n, Y^n}, \mathcal{C}, E = 0) \\ &\quad + \Pr(E = 1) H(X^n | T_{X^n, Y^n}, \mathcal{C}, E = 1) \end{aligned} \quad (45)$$

$$\leq H(\mathbf{H}_K X^n | T_{X^n, Y^n}, \mathcal{C}, E = 0) + o(n) \quad (46)$$

$$= H(\mathbf{H}_K X^n | T_{X^n, Y^n}, K, \mathcal{C}, E = 0) + o(n) \quad (47)$$

$$\leq H(\mathbf{H}_K X^n | K, \mathcal{C}, E = 0) + o(n) \quad (48)$$

where  $o(n)$  in (46) corresponds to the term  $1 + H(T_{X^n, Y^n}, E | \mathcal{C}) + \Pr(E = 1) H(X^n | T_{X^n, Y^n}, \mathcal{C}, E = 1)$ , and (47) follows from the fact that  $K$  is a deterministic function of  $T_{X^n, Y^n}$ .

We now provide an upper bound on the second term of (43).

$$\begin{aligned} \frac{1}{n} H(X^n | K, \mathcal{C}, E = 0) &\leq \frac{1}{n \Pr(E = 0)} \left( H(X^n | K, \mathcal{C}, E) \right. \\ &\quad \left. - \Pr(E = 1) H(X^n | K, \mathcal{C}, E = 1) \right) \end{aligned}$$

$$\leq \frac{1}{n} H(X^n | K, \mathcal{C}, E) + o(1) \quad (49)$$

$$\leq H(X) + o(1). \quad (50)$$

By combining (43), (48) and (50), we obtain

$$\lim_{n \rightarrow \infty} R_2^{(n)} \leq I(X; Y). \quad (51)$$

**Conclusion.** The rates in (41) and (51) are evaluated on average over the random code  $\mathcal{C}$  with a parameter  $\delta > 0$  arbitrarily small. This shows that there exists a sequence of  $(n, R_1^{(n)}, R_2^{(n)})$ -zero-error source codes, such that

$$\left( R_1^{(n)}, R_2^{(n)} \right) \xrightarrow{n \rightarrow \infty} (H(X|Y), I(X; Y)). \quad (52)$$

## REFERENCES

- [1] 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.
- [2] 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.
- [3] A. A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge ; New York: Cambridge University Press, 2011.
- [4] A. El Gamal and A. Orlitsky, "Interactive data compression," in *25th Annual Symposium on Foundations of Computer Science, 1984*. IEEE, 1984, pp. 100–108.
- [5] 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.
- [6] E. Tuncel, J. Nayak, P. Koulgi, and K. Rose, "Zero-error Distributed source coding," in *Distributed source coding: theory, algorithms, and applications*. Elsevier, 2009.
- [7] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.
- [8] T. M. Cover, *Elements of information theory*. John Wiley & Sons, 1999.