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 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 absence 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 in [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 only a part of the 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 $R_2 = H(Y) \geq I(X;Y)$.

In order to be exploitable by both decoders, part of the information sent through the common channel must be independent from Y. For this reason our setting is closely related to the Slepian and Wolf (SW) problem in [2], seen as lossless source coding with side information at the decoder only. In [3], Csiszar proved in that linear codes achieve the optimal SW rate region. Several works in [4]–[6] investigate the duality between SW setting and channel coding using linear codes, as the side-information Y can be seen as the input of a virtual channel with input X. However these tools cannot be straightforwardly adapted to the zero-error setting, as the linear codes proposed also present a vanishing probability of error.

Our setting can be seen as a zero-error variant with side-informations known at the encoder of the successive refinement problem proposed by Kaspi in [7]; later generalized by Timo et al. in [8] for more than two decoders. Even if the lossy reconstruction of the source makes it fundamentally different from the zero-error setting, there are notable examples that present the same tools as in SW. The side-information scalable source coding (i.e. the decoder 2 has a SI Y' s.t. $X \to Y \to Y'$) in [9] for instance uses nested random binning. This random binning approach was further developed in [10] to give a unified coding scheme that works for both scalable source coding and Wyner-Ziv successive refinement in [11] (i.e. the decoder 2 has a SI Y' s.t. $X \to Y' \to Y$).

In the open problem, the zero-error SW scheme requires to send at rate $H(X)$ to the decoder with side information, see [12]–[15]. In [16], Ma and Cheng use linear codes in a zero-error SW restriction, under symmetry assumptions on the source. However, a zero-error SW coding scheme in our setting does not use at all 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.

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

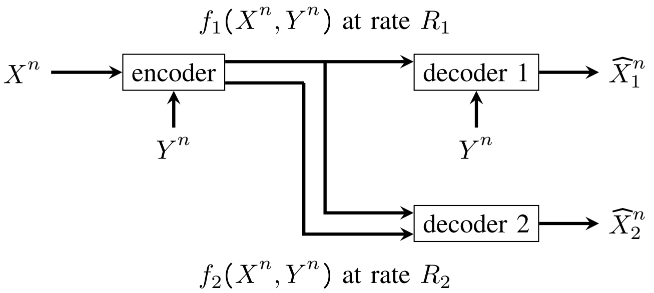

Fig. 1: Source coding when side-information may be present.

use Csiszar and Körner's method of types [17, 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 $x^n = (x_1, ..., x_n)$. 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 $P_X \in \mathcal{P}(\mathcal{X})$. When computing entropies with other distributions than P_X, we specify it in subscript (e.g. $H_Q(X)$ is computed with the distribution Q). 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 $n \in \mathbb{N}^\star = \mathbb{N} \setminus \{0\}$ 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 $n \in \mathbb{N}^\star = \mathbb{N} \setminus \{0\}$, $(R_1^{(n)}, R_2^{(n)}) \in [0, +\infty)^2$, 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:*

$$f_1 : \mathcal{X}^n \times \mathcal{Y}^n \to \{0,1\}^*, \quad f_2 : \mathcal{X}^n \times \mathcal{Y}^n \to \{0,1\}^*, \tag{1}$$

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

that satisfy

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

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

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

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

Theorem II.3

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

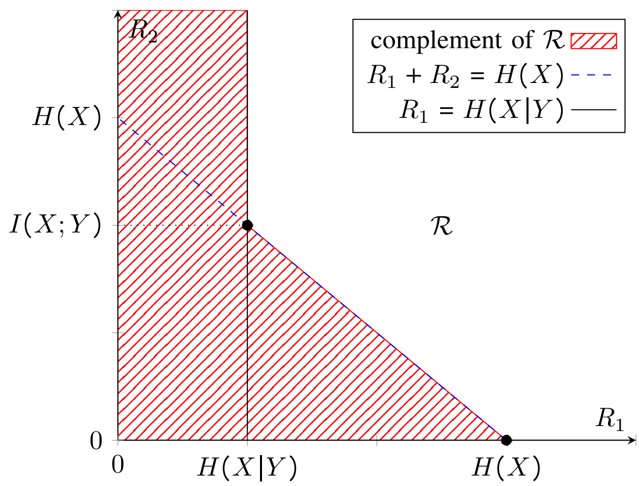

Fig. 2: Zero-error achievable rate region $\mathcal{R}$.

Proof. [Converse of Theorem II.3] In this setting, each decoder must retrieve X with zero-error. Using Shannon lossless source coding result [18, 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. □

III. Achievability Proof of Theorem II.3

In order to prove Theorem II.3, we show that

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

In order to complete the achievability result we use a time sharing with the point $\big(H(X), 0\big)$, 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 Q_{x^n,y^n} that satisfies for all $(x', y') \in \mathcal{X} \times \mathcal{Y}$*

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

We denote the marginal types by Q_{x^n} and Q_{y^n}, respectively. We denote the conditional type of x^n knowing y^n by $Q_{x^n|y^n}$.

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.

We denote by Q_{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 $Q_{X^n|Y^n}$, Q_{X^n} and Q_{Y^n}, respectively.

Definition III.2 (Type class, V-shell) *For all type $\pi \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$, we denote the type class by $\mathcal{T}_\pi$*

$$\mathcal{T}_\pi = \left\{ (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n \;\middle|\; Q_{x^n,y^n} = \pi \right\}. \quad (7)$$

Given a conditional type $V \in \mathcal{P}(\mathcal{X})^{|\mathcal{Y}|}$, the V-shell of a sequence y^n is the set $\mathcal{T}_V(y^n) = \left\{ x^n \in \mathcal{X} \;\middle|\; Q_{x^n|y^n} = V \right\}$.

Definition III.3 (Generator/parity matrix, syndrome, coset) *Let $\mathcal{A}$ be a finite set such that $|\mathcal{A}|$ is prime, so we can give $\mathcal{A} \simeq \mathbb{Z}/|\mathcal{A}|\mathbb{Z}$ a field structure. For all $n, k \in \mathbb{N}^\star$, we denote by $\mathcal{M}_{n,k}(\mathcal{A})$ the set of $n \times k$ matrices over the finite field $\mathcal{A}$.*

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

The syndrome of a sequence $a^n \in \mathcal{A}^n$ is $\boldsymbol{H}x^n$. The coset associated to the syndrome $\boldsymbol{H}a^n$ is the set Im $\boldsymbol{G} + a^n = \{\tilde{a}^n \in \mathcal{A}^n \mid \boldsymbol{H}\tilde{a}^n = \boldsymbol{H}a^n\}$.

B. Coding scheme

For all $n \in \mathbb{N}^\star$, we show the existence of a sequence of $(n, R_1^{(n)}, R_2^{(n)})$-zero-error source codes that achieves the corner-point $\big(H(X|Y), I(X;Y)\big)$ of the zero-error rate region $\mathcal{R}$. Our proof is based on a linear code adjusted depending on Q_{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 also assume w.l.o.g. that $|\mathcal{X}|$ is prime number by padding (i.e. extending with zeros) $P_{X,Y}$ if necessary. 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

$$k \doteq \left\lceil n - n \frac{H_{Q_{x^n,y^n}}(X|Y) + \delta}{\log|\mathcal{X}|} \right\rceil^+. \quad (8)$$

  where $\lceil \cdot \rceil$ denotes the ceiling function and $(\cdot)^+$ denotes $\max(\cdot, 0)$. We denote by K the random variable induced by k defined in (8), 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 $\big(\text{Im } \mathbf{G}_K + X^n\big) \cap \mathcal{T}_{Q_{X^n|Y^n}}(Y^n) = \{X^n\}$; $E = 1$ otherwise. Then we define

$$f_1(X^n, Y^n) = \begin{cases} b(Q_{X^n,Y^n}, E, \mathbf{H}_K X^n) & \text{if } E = 0, \\ b(Q_{X^n,Y^n}, E, X^n) & \text{if } E = 1, \end{cases} \quad (9)$$

  where $b(\cdot)$ denotes the binary expansion.
- Encoding function f_2: If $E = 0$, the index of X^n in its coset 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

$$f_2(X^n, Y^n) = B(\mathbf{G}_K, X^n, Y^n). \quad (10)$$

  Otherwise, $f_2(X^n, Y^n) = 0$.
- Decoding function g_1: It observes $f_1(X^n, Y^n)$ and extracts E and Q_{X^n,Y^n}. If $E = 1$,

$$g_1(f_1(X^n, Y^n), Y^n) = X^n. \quad (11)$$

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

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

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

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

Remark III.4 *The parameter K is selected so that when $K > 0$, the number of parity bits of the linear code asymptotically matches the conditional entropy: $\frac{(n-K)\log|\mathcal{X}|}{n} = H_{Q_{X^n,Y^n}}(X|Y) + \delta + O\left(\frac{1}{n}\right)$.*

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 $\big(\text{Im } \mathbf{G}_K + X^n\big) \cap \mathcal{T}_{Q_{X^n|Y^n}}(Y^n) = \{X^n\}$, hence $g_1\big(f_1(X^n, Y^n), Y^n\big) = X^n$ 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 Im $\mathbf{G}_K + X^n$ with index $f_2(X^n, Y^n)$ is X^n. Thus, $g_2\big(f_1(X^n, Y^n), f_2(X^n, Y^n)\big) = X^n$ 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

$$R_1^{(n)} \underset{n\to\infty}{\to} H(X|Y) + \delta, \qquad R_2^{(n)} \underset{n\to\infty}{\to} I(X;Y). \tag{12}$$

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

$$Pr\left(Q_{X'^n,y^n} = Q_{x^n,y^n}\right) = 2^{nH_{Q_{x^n,y^n}}(X|Y) - n\log|\mathcal{X}| + o(n)} \tag{13}$$

Proof. Since $P_{X'}$ is uniform:

$$\begin{aligned}\Pr\left(Q_{X'^n,y^n} = Q_{x^n,y^n}\right) &= |\mathcal{X}|^{-n}\left|\mathcal{T}_{Q_{x^n|y^n}}(y^n)\right| &(14)\\ &= 2^{-n\log|\mathcal{X}|}2^{nH_{Q_{x^n,y^n}}(X|Y)+o(n)},\end{aligned}$$

as [17, Lemma 2.5] gives the asymptotic size of the $Q_{x^n|y^n}$-shell $\mathcal{T}_{Q_{x^n|y^n}}(y^n)$.
□

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

$$\begin{aligned}&\Pr(E=1) \le \Pr(K=0) &(15)\\ &+ \Pr\left(\bigcup_{\substack{\alpha\in\mathcal{X}^K\\ \alpha\neq 0}}\left[Q_{\left(X^n+\sum_{i\le K}\alpha_i\mathbf{G}_K^{(i)}\right),Y^n} = Q_{X^n,Y^n}\right]\middle| K\neq 0\right).\end{aligned}$$

We provide an upper bound on the second term in (15). For all (x^n, y^n) such that $k \neq 0$, we have:

$$\begin{aligned}&\Pr\left(\bigcup_{\substack{\alpha\in\mathcal{X}^k\\ \alpha\neq 0}}\left[Q_{\left(x^n+\sum_{i\le k}\alpha_i\mathbf{G}_k^{(i)}\right),y^n} = Q_{x^n,y^n}\right]\right)\\ &\le \sum_{\substack{\alpha\in\mathcal{X}^k\\ \alpha\neq 0}}\Pr\left(Q_{\left(x^n+\sum_{i\le k}\alpha_i\mathbf{G}_k^{(i)}\right),y^n} = Q_{x^n,y^n}\right) &(16)\\ &\le |\mathcal{X}|^k 2^{nH_{Q_{x^n,y^n}}(X|Y) - n\log|\mathcal{X}| + o(n)} &(17)\\ &\le 2^{n\log|\mathcal{X}| - nH_{Q_{x^n,y^n}}(X|Y) - \delta n + o(n)}\\ &\quad\times 2^{nH_{Q_{x^n,y^n}}(X|Y) - n\log|\mathcal{X}| + o(n)} \le 2^{-\delta n + o(n)}, &(18)\end{aligned}$$

where (17) comes from Lemma 1 and (18) comes from (8). Therefore,

$$\begin{aligned}&\Pr\left(\bigcup_{\substack{\alpha\in\mathcal{X}^K\\ \alpha\neq 0}}\left[Q_{\left(X^n+\sum_{i\le K}\alpha_i\mathbf{G}_K^{(i)}\right),Y^n} = Q_{X^n,Y^n}\right]\middle| K\neq 0\right)\\ &= \sum_{x^n,y^n}\Pr\Big((X^n,Y^n) = (x^n,y^n)\Big| K\neq 0\Big)\\ &\times\Pr\left(\bigcup_{\substack{\alpha\in\mathcal{X}^K\\ \alpha\neq 0}}\left[Q_{\left(X^n+\sum_{i\le K}\alpha_i\mathbf{G}_K^{(i)}\right),Y^n} = Q_{X^n,Y^n}\right]\right.\\ &\qquad\qquad\left.\Big| K\neq 0, (X^n,Y^n) = (x^n,y^n)\right) &(19)\\ &\le \sum_{x^n,y^n}\Pr\Big((X^n,Y^n) = (x^n,y^n)\Big| K\neq 0\Big) 2^{-\delta n + o(n)} &(20)\\ &\le 2^{-\delta n + o(n)}, &(21)\end{aligned}$$

where (20) comes from (18) and the fact that $\mathbf{G}$ is independent of (X, Y).

We now provide an upper bound on the first term in (15).

$$\mathcal{S} \doteq \left\{\pi \in \mathcal{P}(\mathcal{X}\times\mathcal{Y}), \quad 1 - \frac{H_\pi(X|Y) + \delta}{\log|\mathcal{X}|} \le 0\right\}. \tag{22}$$

Then we have:

$$\begin{aligned}&\Pr(K=0) &(23)\\ &= \Pr\left(\left\lceil n - n\frac{H_{Q_{X^n,Y^n}}(X|Y)+\delta}{\log|\mathcal{X}|}\right\rceil^+ = 0\right) &(24)\\ &= \Pr\left(n - n\frac{H_{Q_{X^n,Y^n}}(X|Y)+\delta}{\log|\mathcal{X}|} \le 0\right) &(25)\\ &= \Pr(Q_{X^n,Y^n} \in \mathcal{S}) &(26)\\ &= \sum_{\pi\in\mathcal{S}\cap\mathcal{P}_n(\mathcal{X}\times\mathcal{Y})}\Pr(Q_{X^n,Y^n} = \pi) &(27)\\ &\le |\mathcal{S}\cap\mathcal{P}_n(\mathcal{X}\times\mathcal{Y})| \sup_{\pi\in\mathcal{S}\cap\mathcal{P}_n(\mathcal{X}\times\mathcal{Y})}\Pr(Q_{X^n,Y^n} = \pi) &(28)\\ &\le |\mathcal{S}\cap\mathcal{P}_n(\mathcal{X}\times\mathcal{Y})| \sup_{\pi\in\mathcal{S}\cap\mathcal{P}_n(\mathcal{X}\times\mathcal{Y})} 2^{-nD(\pi\|P_{X,Y})} &(29)\\ &\le |\mathcal{S}\cap\mathcal{P}_n(\mathcal{X}\times\mathcal{Y})| \sup_{\pi\in\mathcal{S}} 2^{-nD(\pi\|P_{X,Y})} &(30)\\ &\le 2^{-n\inf_{\pi\in\mathcal{S}} D(\pi\|P_{X,Y}) + o(n)}, &(31)\end{aligned}$$

where (29) comes from [17, Lemma 2.6]. Since $P_{X,Y} \notin \mathcal{S}$ by definition of δ, we have $\inf_{\pi\in\mathcal{S}} D(P_{X,Y}\|\pi) > 0$. Thus there exists a positive constant $\beta > 0$ such that

$$\Pr(K=0) \le 2^{-\beta n + o(n)}. \tag{32}$$

Thus by combining (15), (21), (32), we have:

$$\Pr(E=1) \le 2^{-\delta n + o(n)} + 2^{-\beta n + o(n)}. \tag{33}$$

Rate on the common channel. The encoding function f_1 defined in (9) returns Q_{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,

$$\begin{aligned}nR_1^{(n)} &= 1 + |\mathcal{X}||\mathcal{Y}|\log_2(n+1) + \Pr(E=1)n\log|\mathcal{X}|\\ &\quad + \Pr(E=0)\sum_{x^n,y^n}\Pr\big((X^n,Y^n) = (x^n,y^n)\big|E=0\big)\\ &\quad\times(n-k)\log|\mathcal{X}| &(34)\\ &\le 1 + |\mathcal{X}||\mathcal{Y}|\log_2(n+1) + \Pr(E=1)n\log|\mathcal{X}|\end{aligned}$$

$$+ (n - \mathbb{E}[K]) \log|\mathcal{X}| \quad (35)$$

$$\leq 1 + |\mathcal{X}||\mathcal{Y}| \log_2(n+1) + \Pr(E=1) n \log|\mathcal{X}| + n\mathbb{E}\Big[H_{Q_{X^n,Y^n}}(X|Y)\Big] + n\delta + 1, \quad (36)$$

where (35) comes from $n - k \geq 0$ for all (x^n, y^n), and (36) comes from (8).

By the law of large numbers [18, Theorem 11.2.1] $\mathbb{E}\Big[H_{Q_{X^n,Y^n}}(X|Y)\Big] \underset{n\to\infty}{\to} H(X|Y)$, and by using (33), we obtain

$$\lim_{n\to\infty} R_1^{(n)} \leq H(X|Y) + \delta. \quad (37)$$

Rate on the secondary channel. The encoding function f_2 is defined in (10). 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

$$R_2^{(n)} \leq \frac{1}{n}\Big(1 + \sum_{k\neq 0} \Pr(K=k|E=0) \times H(X^n|\mathbf{H}_k X^n, K=k, \mathcal{C}, E=0)\Big) \quad (38)$$

$$= \frac{1}{n} + \frac{1}{n} H(X^n|K,\mathcal{C},E=0) - \frac{1}{n} H(\mathbf{H}_K X^n|K,\mathcal{C},E=0), \quad (39)$$

where (39) 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 (39). 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, the decoder 1 retrieves X^n with zero error (see Sec. III-C), and the entropy coder is also lossless. Thus r is greater than the rate achieved by a conditional entropy coder that compresses X^n knowing the side information Y^n, whose rate is lower bounded by $nH(X|Y)$.

Therefore, we have

$$nH(X|Y) \leq r < H(f_1(X^n,Y^n)|\mathcal{C}) + 1 \quad (40)$$

$$= 1 + H(Q_{X^n,Y^n}, E|\mathcal{C}) + \Pr(E=0) H(\mathbf{H}_K X^n|Q_{X^n,Y^n},\mathcal{C},E=0) + \Pr(E=1) H(X^n|Q_{X^n,Y^n},\mathcal{C},E=1) \quad (41)$$

$$\leq H(\mathbf{H}_K X^n|Q_{X^n,Y^n},\mathcal{C},E=0) + o(n) \quad (42)$$

$$= H(\mathbf{H}_K X^n|Q_{X^n,Y^n},K,\mathcal{C},E=0) + o(n) \quad (43)$$

$$\leq H(\mathbf{H}_K X^n|K,\mathcal{C},E=0) + o(n) \quad (44)$$

where $o(n)$ in (42) corresponds to the term $1 + H(Q_{X^n,Y^n}, E|\mathcal{C}) + \Pr(E=1)H(X^n|Q_{X^n,Y^n},\mathcal{C},E=1)$, and (43) follows from the fact that K is a deterministic function of Q_{X^n,Y^n}.

We now provide an upper bound on the second term of (39).

$$\frac{1}{n}H(X^n|K,\mathcal{C},E=0) \leq \frac{1}{n\Pr(E=0)}\Big(H(X^n|K,\mathcal{C},E) - \Pr(E=1)H(X^n|K,\mathcal{C},E=1)\Big)$$

$$\leq \frac{1}{n}H(X^n|K,\mathcal{C},E) + o(1) \quad (45)$$

$$\leq H(X) + o(1). \quad (46)$$

By combining (39), (44) and (46), we obtain

$$\lim_{n\to\infty} R_2^{(n)} \leq I(X;Y). \quad (47)$$

Conclusion. The rates in (37) and (47) 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 $\Big(n, R_1^{(n)}, R_2^{(n)}\Big)$-zero-error source codes, such that

$$\Big(R_1^{(n)}, R_2^{(n)}\Big) \underset{n\to\infty}{\to} \big(H(X|Y), I(X;Y)\big). \quad (48)$$

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] I. Csiszar, "Linear codes for sources and source networks: Error exponents, universal coding," *IEEE Transactions on Information Theory*, vol. 28, no. 4, pp. 585–592, 1982.

[4] J. Chen, D.-K. He, and A. Jagmohan, "The equivalence between slepian-wolf coding and channel coding under density evolution," *IEEE Transactions on Communications*, vol. 57, no. 9, pp. 2534–2540, 2009.

[5] J. Chen, D.-K. He, A. Jagmohan, L. A. Lastras-Montaño, and E.-H. Yang, "On the linear codebook-level duality between slepian–wolf coding and channel coding," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5575–5590, 2009.

[6] L. Wang and Y.-H. Kim, "Linear code duality between channel coding and slepian-wolf coding," in *2015 53rd IEEE Annual Allerton Conference on Communication, Control, and Computing (Allerton)*.

[7] A. H. 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, 1994.

[8] R. Timo, T. Chan, and A. Grant, "Rate distortion with side-information at many decoders," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5240–5257, 2011.

[9] C. Tian and S. N. Diggavi, "Side-information scalable source coding," *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5591–5608, 2008.

[10] E. Akyol, U. Mitra, E. Tuncel, and K. Rose, "On scalable coding in the presence of decoder side information," in *2014 IEEE International Symposium on Information Theory (ISIT)*.

[11] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner–Ziv problem," *IEEE Transactions on Information Theory*, vol. 50, no. 8, pp. 1636–1654, 2004.

[12] A. El Gamal and A. Orlitsky, "Interactive Data Compression," in *Annual Symposium on Foundations of Computer Science*, 1984, p. 9.

[13] 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.

[14] E. Tuncel, "Kraft inequality and zero-error source coding with decoder side information," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4810–4816, 2007.

[15] E. Tuncel, J. Nayak, P. Koulgi, and K. Rose, "Zero-error distributed source coding," in *Distributed source coding: theory, algorithms, and applications*. Elsevier, 2009.

[16] R. Ma and S. Cheng, "Zero-error Slepian–Wolf coding of confined-correlated sources with deviation symmetry," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 8195–8209, 2013.

[17] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[18] T. M. Cover, *Elements of information theory*. John Wiley & Sons, 1999.