

# Rate-storage regions for Extractable Source Coding with side information

Elsa Dupraz, Aline Roumy, Thomas Maugey, Michel Kieffer

► **To cite this version:**

Elsa Dupraz, Aline Roumy, Thomas Maugey, Michel Kieffer. Rate-storage regions for Extractable Source Coding with side information. *Physical Communication*, Elsevier, 2019, 37, pp.1-24. 10.1016/j.phycom.2019.100845 . hal-02925026v2

**HAL Id: hal-02925026**

**<https://hal.inria.fr/hal-02925026v2>**

Submitted on 3 Dec 2019

**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.

# Rate-Storage Regions for Extractable Source Coding with Side Information

Elsa Dupraz<sup>1</sup>, Aline Roumy<sup>2</sup>, Thomas Maugey<sup>2</sup>, Michel Kieffer<sup>3</sup>

<sup>1</sup> IMT Atlantique, Lab-STICC, UBL, <sup>2</sup> INRIA Rennes Bretagne-Atlantique, <sup>3</sup> L2S, UMR CNRS 8506; CentraleSupélec; Univ. Paris-Sud

## Abstract

This paper considers the coding of a source  $X$  with  $J$  decoders, each having access to a different side information  $Y_{(j)}$ . We define a new source coding problem called Extractable Source Coding with Side Information (ESC-SI). In this problem, the server stores one single coded description of the source, from which  $J$  descriptions can be extracted without re-encoding, depending on the side information available at the decoder. We want to minimize both the storage rate of the source on the server, and the  $J$  transmission rates from the server to the decoders. We provide the achievable storage-transmission rate regions for lossless source coding of general, non i.i.d., non-ergodic sources, and the achievable storage-transmission rate-distortion regions for lossy source coding for non i.i.d. Gaussian sources. The regions obtained for such general source models provide insightful design guidelines for practical applications.

## Index Terms

Source coding with side information, Lossless source coding, Lossy source coding, General sources, Gaussian sources, Storage-Transmission rate region

## I. INTRODUCTION

In the well-known Slepian-Wolf (SW) [1] and Wyner-Ziv (WZ) [2] coding schemes, a source  $X$  is encoded given that a side information  $Y$  is only available at the decoder. Several extensions have been considered in the literature. For instance, distributed source coding [3]–[5] refers to the case where several sources are encoded separately and must be recovered by one single decoder. On the opposite, Sgarro [6]

Section IV of this paper was presented at DCC 2018 and published as a one-page abstract in the proceedings of this conference.

introduced the case where one source  $X$  must be recovered by  $J$  decoders with different side informations  $Y_{(1)}, \dots, Y_{(J)}$ . In the broadcast setup considered in [6], [7], the source  $X$  is encoded into one single message at rate  $R$ , and each of the  $J$  decoders may be able to recover  $X$  from this message and from its own side information  $Y_{(j)}$ . By contrast, in successive refinement with side information [7]–[11], the source  $X$  is encoded into  $J$  descriptions  $\{M_{(1)}, \dots, M_{(J)}\}$  with respective rates  $R_{(1)}, \dots, R_{(J)}$ , and decoder  $j$  has to recover  $X$  from  $Y_{(j)}$  and from the first  $j$  descriptions  $M_{(1)}$  to  $M_{(j)}$ . The special case of successive refinement without side information was also considered in [12]–[15].

In this paper, we introduce a novel information-theoretic problem related to source coding with multiple side informations. We assume that the data generated by the source  $X$  is stored on a server to be later requested by many users. When requesting the data generated by  $X$ , user  $j$  already has access to some side information  $Y_{(j)}$ , which may come, e.g., from previous requests to the same server. In this context, we aim at minimizing both the *storage rate*  $S$  of the source  $X$  on the server, and the individual *transmission rates*  $R_{(j)}$  from the server to the  $J$  users. This objective differs from the ones in [6]–[11], where the focus is on the storage rate [6], [7] or on the transmission rates only [8]–[11]. For joint storage and transmission rates minimization, a straightforward strategy for the server when answering to a request by user  $j$  would be to decode the compressed description of  $X$  and to re-encode it accounting for the fact that a given  $Y_{(j)}$  is available at the decoder. Nevertheless, this strategy is often not desirable due to computational load on the server when the number of simultaneous user requests is large. As a result, in our formulation, we require the server to produce one single coded description of  $X$ , from which  $J$  different descriptions can be extracted *without re-encoding* depending on the  $Y_{(j)}$  available at the decoder. We therefore call this coding scheme Extractable Source Coding with Side Information (ESC-SI). By extraction without re-encoding, we mean that the source  $X$  is compressed into a set of  $I$  bits, and only a subset  $I_{(j)} \subseteq I$  of these bits are transmitted to user  $j$ . In the successive refinement setup, the subsets  $I_{(j)}$  are incremental in the sense that  $I_{(j)} \subseteq I_{(j+1)}$ . No such constraint appears in our coding scheme definition.

Applications of this problem include free-viewpoint television (FTV) [16]–[18], or access to data collected by a network of sensors [19]. For instance, in FTV, the source  $X$  represents one view, and  $Y_{(j)}$  represents the prediction of this view computed by user  $j$  based on its previously requested views. Most practical coding schemes for FTV store one coded description of  $X$  per possible prediction  $Y_{(j)}$  [16], [17]. These schemes hence give transmission rates  $R_{(j)} = H(X|Y_{(j)})$ , which are optimal, but a storage rate  $S = \sum_j H(X|Y_{(j)})$ , which is highly suboptimal due to the multiple stored representations of  $X$ . In contrast, the compound coding schemes proposed in [18], [20] store one single description of  $X$  that will

be sent whatever the  $Y_{(j)}$  available at the receiver. This gives  $S = R_{(j)} = \max_j H(X|Y_{(j)})$  for all  $j$ , which is optimal in terms of storage rate but suboptimal in terms of transmission rate. Applying the successive refinement results [7]–[11] to our problem may allow to achieve a storage rate  $S = \max_j H(X|Y_{(j)})$  (as in joint source encoding) and individual transmission rates  $R_{(j)} = H(X|Y_{(j)})$  (as in exhaustive source encoding), which was not recognized in the above practical schemes (see Theorem 1).

Most existing works [7]–[11] on successive refinement with side information only consider sources  $X$  and  $Y_{(j)}$  that generate independent and identically distributed (i.i.d.) symbols according to probability distributions  $P_X$  and  $P_{Y_{(j)}|X}$ . In addition, the conditional probability distributions  $P_{Y_{(j)}|X}$  are usually assumed to be physically degraded. In practical applications, however, the data to be compressed may be better described as realization of non-stationary sources or sources with memory. Moreover, the relation between  $X$  and the various side informations  $Y_{(j)}$  may be described by different models (erasure, noise, without or with memory). In these cases, the i.i.d. and degraded assumptions will not allow to capture all the statistical dependencies between successive symbols nor the side informations heterogeneity.

In this paper, after introducing a formal definition of the coding scheme, we provide the storage and transmission rate regions for lossless coding of non i.i.d. heterogeneous sources. The storage and transmission rate regions were already provided in [21] for i.i.d. sources, although [21] does not formally define the coding scheme. We show that, even with this general source model, one can always achieve the maximum of conditional entropies as the storage rate and the conditional entropies of each source as the transmission rates. As particular cases, we also provide the achievable storage and transmission rate regions for i.i.d. and ergodic sources. As an example, we consider a set of heterogeneous side informations  $Y_{(j)}$  such that the conditional probability distributions  $P(Y_{(j)}|X)$  correspond to one Binary Erasure Channel (BEC) model, one Binary Symmetric Channel (BSC) model, one Gilbert-Elliot (GE) channel model, and one Gaussian channel model, respectively. We then consider the lossy coding of non i.i.d., non-ergodic, Gaussian sources and we provide the corresponding storage and transmission rate-distortion regions. Here again, we describe the particular case of ergodic sources, and show an example of heterogeneous set of sources composed by first-order Markov sources and by sources with nearest neighbor correlations. This analysis provides insightful design guidelines for practical applications.

### A. Related problems

In our ESC-SI framework described in Fig. 1(a), a source  $X$  is encoded offline into one single message  $M$  at storage rate  $S$ . In order to reconstruct  $X$ , decoder  $j$  has access to side information  $Y_{(j)}$  and receives a part  $M_{(j)} \subseteq M$  of rate  $R_{(j)}$  of the message  $M$ .  $M_{(j)}$  is extracted from  $M$  by an online extractor.

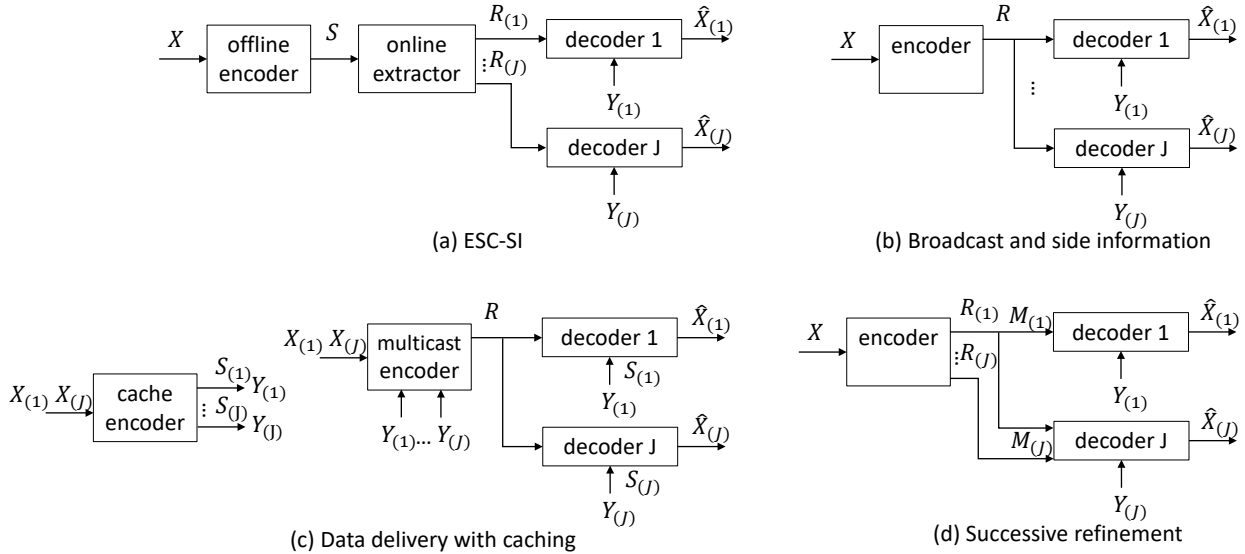


Fig. 1: (a) ESC-SI: problem setup studied in the paper. (b) Source coding problem with broadcast to multiple decoders with multiple side informations. (c) Data delivery with caching. (d) Successive refinement.

In standard Slepian-Wolf [1] and Wyner-Ziv schemes [2], there is only one decoder with side information  $Y$ . In the lossless Slepian-Wolf setup, the achievable rate was given in [1] for i.i.d. sources, and in [22] for general non i.i.d. sources with memory. In this setup, the storage rate  $S$  necessarily equals the transmission rate  $R$ . In the lossy Wyner-Ziv setup, the achievable rate-distortion function was given in [2] for i.i.d. sources, in [23] for general non i.i.d. sources, and in [24] for non i.i.d. Gaussian sources. By contrast, in the proposed ESC-SI framework, there are several possible decoders, and each decoder has access to a different side information. The ESC-SI setup could be addressed by considering  $J$  independent Slepian-Wolf or Wyner-Ziv schemes in parallel. However, we show in the paper that this strategy leads to high storage rate compared to the coding scheme we propose.

In the broadcast setup shown in Fig. 1(b), the source  $X$  is encoded into one message  $M$  at rate  $S$  that is sent to  $J$  decoders. Decoder  $j \in \llbracket 1, J \rrbracket$  wishes to recover  $X$  from  $M$  and from the side information  $Y_{(j)}$ . Although this problem remains open in general, some special cases have been derived. For instance, the lossless case of this problem is solved in [6] for i.i.d. sources. In the lossy case, this problem is known as the Heegard-Berger problem and is solved for stochastically degraded side information [25, Thm. 2]. An outer bound is derived in [10] for the case of i.i.d. non-degraded sources. The broadcast problem has also been extended to the setup with encoder side information in [20] for lossless reconstruction and

in [26], [27] for lossy reconstruction of i.i.d. Gaussian sources with quadratic distortion measures. Non i.i.d. Gaussian sources were further considered in [11]. Nevertheless, in the broadcast problem, the same message is sent to all decoders, unlike in our framework.

In the problem of data delivery with caching described in Fig. 1(c), a server makes available a database, and each of the clients involved in the network stores a limited part of the content of the database. This limited content can be seen as different side informations  $Y_{(1)}, \dots, Y_{(J)}$  stored in the user's memories at storage rates  $S_{(1)}, \dots, S_{(J)}$ . Then, during the delivery phase, each user requests some data from the server, that has to satisfy these requests through a single link at delivery rate  $R$ . The objective is then to minimize the rate  $R$  on this single link, under constraints on the storage rates  $S_{(j)}$ . For this problem, theoretical rate-memory regions were derived in [28] for i.i.d. sources, and in [29] for correlated sources. However, in this problem again, the data is transmitted through one single link, unlike in our setup. In addition, in our setup, the storage rate  $S$  corresponds to data stored by the server, while in [28], [29], the  $J$  storage rates  $S_{(1)}, \dots, S_{(J)}$  correspond to data stored by the clients.

Then, in the problem of successive refinement described in Fig. 1(d), the source  $X$  is encoded into a set  $I$  of  $J$  messages  $\{M_{(1)}, \dots, M_{(J)}\}$  at rates  $R_{(1)}, \dots, R_{(J)}$ . Decoder  $j$  has to recover the source  $X$  from its side information  $Y_{(j)}$  and from the set  $I_{(j)} = \{M_{(1)}, \dots, M_{(j)}\} \subseteq I$  of the first  $j$  messages. This problem was first studied in [12]–[15] for i.i.d. sources in the case without side information. In this case, the results were also extended to general sources in [30], [31]. The problem of successive refinement with side information was then studied in [7], [9], [10], where achievable rate regions are derived for i.i.d. sources in both lossless and lossy setups. In successive refinement, the message construction is incremental in the sense that the sets  $I_{(j)}$  satisfy  $I_{(j)} \subseteq I_{(j+1)}$ . Our framework is slightly more general since the sets  $I_{(j)}$  are not constrained to be incremental in our coding scheme definition.

However, when deriving the proofs for our storage-transmission rate regions, it appears that incremental codeword construction achieves the optimal storage and transmission rates. Since we consider general non i.i.d. sources and since we do not consider any physical-degradation assumption, the main difficulty is then to find a source ordering for incremental code construction. In the context of successive refinement without physical-degradation assumption, [10] proposes a source ordering that holds for i.i.d. sources only. In the broadcast setup, [11] proposes an ordering in the case of one source and two side informations that generate sequences of i.i.d. Gaussian vectors, with correlation inside the vectors. However, the incremental criterion of [11] cannot be applied when more than two side informations are considered. In this paper, we propose one criterion for the ordering of general non i.i.d. sources for lossless coding and one criterion for the ordering of non i.i.d. Gaussian sources for lossy source coding. It is worth noting that the Gaussian

model we consider is more general than the model of [11] since it directly considers sequences of non i.i.d. Gaussian random variables.

## II. PROBLEM FORMULATION

In this section, we describe our notations and assumptions for the non i.i.d. sources considered in this paper. We then formally define the lossless and lossy coding schemes, as well as the corresponding storage and transmission rates.

### A. Notation

In this paper, a source  $\mathbf{X}$  is denoted using a bold uppercase letter. The source  $\mathbf{X}$  generates a sequence of random variables denoted  $X_i$  using an uppercase letter. The realization  $x_i$  of  $X_i$  is denoted using a lowercase letter. In addition,  $\mathbf{X}^n = (X_1, \dots, X_n)$  is a random vector of length  $n$  containing elements  $X_1$  to  $X_n$ . It is denoted using a boldface uppercase letter and its realization  $\mathbf{x}^n$  is denoted using a boldface lowercase letter. The alphabet  $\mathcal{X}$  of a random variable  $X$  is denoted with a calligraphic letter, the cardinality of the set  $\mathcal{X}$  is denoted  $|\mathcal{X}|$ , and the  $n$ -th extension of  $\mathcal{X}$  is represented by  $\mathcal{X}^n$ . If  $\mathcal{B}$  is a finite source alphabet, then  $\mathcal{B}^+$  is the set of all finite strings from  $\mathcal{B}$ .  $\forall \mathbf{b} \in \mathcal{B}^+$ ,  $|\mathbf{b}|$  stands for the length of  $\mathbf{b}$ . We use  $\llbracket 1, N \rrbracket$  to denote the set of integers between 1 and  $N$ .

### B. Source models

This section describes the source models we consider in this paper.

**Definition 1** (General sources). *Let  $\mathbf{X}$  be a source that generates symbols in an alphabet  $\mathcal{X}$  which can be either discrete or continuous. Let  $\{\mathbf{Y}_{(j)}\}_{1 \leq j \leq J}$  be a set of  $J$  sources that generate symbols in  $J$  discrete or continuous alphabets  $\mathcal{Y}_1, \dots, \mathcal{Y}_J$ , respectively. The set  $\{\mathbf{X}, \{\mathbf{Y}_{(j)}\}_{1 \leq j \leq J}\}$  is said to be a set of  $J + 1$  general sources if the  $J + 1$  sources generate random vectors of size  $n \geq 1$  according to the joint probability distribution  $P_{\mathbf{X}^n, \mathbf{Y}_{(1)}^n, \dots, \mathbf{Y}_{(J)}^n}$ .*

This source model is a particular case of the one introduced in [22, Section 1.3]. In particular, it satisfies the consistency condition described in [22, Section 1.3]. Since no assumption is made on the joint probability distribution  $P_{\mathbf{X}^n, \mathbf{Y}_{(1)}^n, \dots, \mathbf{Y}_{(J)}^n}$ , the general source model of Definition 1 encompasses a wide range of source models, including non i.i.d., non-stationary sources. In particular, the sources  $\{\mathbf{Y}_{(j)}\}_{1 \leq j \leq J}$  that will be considered as side information at the decoders are not assumed to be degraded in any sense. The alphabets  $\mathcal{Y}_1, \dots, \mathcal{Y}_J$  can be different from each others, which allows to consider

heterogeneous sources. For instance,  $\mathcal{Y}_1$  may be binary,  $\mathcal{Y}_2$  may contain erasure symbols,  $\mathcal{Y}_3$  may be continuous, etc. As a particular case of Definition 1, i.i.d. sources are such that  $\mathcal{Y}_1 = \dots = \mathcal{Y}_J = \mathcal{Y}$ , and the joint probability distribution  $P_{\mathbf{X}^n, \mathbf{Y}_{(1)}^n, \dots, \mathbf{Y}_{(J)}^n}$  can be expressed as

$$P_{\mathbf{X}^n, \mathbf{Y}_{(1)}^n, \dots, \mathbf{Y}_{(J)}^n} = \prod_{i=1}^n P_{X_i, Y_{(1),i}, \dots, Y_{(J),i}}. \quad (1)$$

For lossless source coding, we will derive the storage and transmission rates for the model of Definition 1, for a discrete alphabet  $\mathcal{X}$  and for whatever alphabets  $\mathcal{Y}_j$ . Lossy source coding problems with multiple side informations are known to be very difficult, and remain open in general [7]. This is why, in this paper, we restrict ourselves to the following non i.i.d. Gaussian model.

**Definition 2** (Gaussian sources). *Let  $\{\mathbf{X}, \{\mathbf{Y}_{(j)}\}_{1 \leq j \leq J}\}$  be a set of general sources that generate jointly Gaussian vectors  $(\mathbf{X}^n, \mathbf{Y}_{(1)}^n, \dots, \mathbf{Y}_{(J)}^n)$ , each of respective length  $n$ . The random vectors  $\mathbf{X}^n, \mathbf{Y}_{(1)}^n, \dots, \mathbf{Y}_{(J)}^n$  are all assumed to be centered, i.e.,  $E[\mathbf{X}^n] = \mathbf{0}^n$ , and  $\forall j \in \llbracket 1, J \rrbracket$ ,  $E[\mathbf{Y}_{(j)}^n] = \mathbf{0}^n$ . For all  $j, j' \in \llbracket 1, J \rrbracket$ , the random vectors  $\mathbf{Y}_{(j)}^n, \mathbf{Y}_{(j')}^n$  are assumed independent conditionally to  $\mathbf{X}^n$ . We denote by  $\Sigma_X = E[\mathbf{X}^n(\mathbf{X}^n)^T]$  the covariance matrix of  $\mathbf{X}^n$  and,  $\forall j \in \llbracket 1, J \rrbracket$ , by  $\Sigma_{(j)} = E[\mathbf{Y}_{(j)}^n(\mathbf{Y}_{(j)}^n)^T]$  the covariance matrix of  $\mathbf{Y}_{(j)}^n$ . We use  $\Sigma_{X, Y_{(j)}}$  to denote  $E[\mathbf{X}^n(\mathbf{Y}_{(j)}^n)^T]$ . All the matrices  $\Sigma_X, \Sigma_{(j)}, \Sigma_{X, Y_{(j)}}$ , are assumed to be positive-definite.*

We also denote by

$$\Sigma_{(X|j)}^n = \Sigma_X - \Sigma_{X, Y_{(j)}} \Sigma_{(j)}^{-1} \Sigma_{X, Y_{(j)}} \quad (2)$$

the conditional covariance matrix of  $\mathbf{X}^n$  knowing  $\mathbf{Y}_{(j)}^n$ , and by  $\lambda_i^{(X|j)}$ ,  $i \in \llbracket 1, n \rrbracket$ , the eigenvalues of  $\Sigma_{(X|j)}^n$ .

Most existing works on source coding with multiple side informations only considered i.i.d. Gaussian models [7], [9], [10]. The above model cannot be reduced to the i.i.d. case by applying the conditional Karhunen-Loeve Transform (KLT) [24] since the conditional KLT varies with the side information  $\mathbf{Y}_{(j)}$ . In addition, [11] assumes that the sources  $\mathbf{X}$  and  $\mathbf{Y}_{(j)}$  generate sequences  $(\mathbf{X}_n^k, \mathbf{Y}_{(j),n}^k)_{n \geq 1}$  of Gaussian vectors of length  $k$ . For example, the components of  $\mathbf{X}_n^k$  are correlated, but independent from the components of previous and successive vectors  $\mathbf{X}_{n-1}^k, \mathbf{X}_{n+1}^k$ . For instance, Gauss-Markov sources cannot be represented by the model of [11]. On the opposite, the Gaussian model of Definition 2 captures the dependencies between components for a wide range of Gaussian sources, even with memory or non stationarity.



### C. Spectral conditional entropy and mutual information

Since our definition of general sources and Gaussian sources does not perform any assumption on e.g., stationarity and ergodicity, we need to consider spectral entropy and spectral mutual information introduced in [22] rather than standard definitions of entropy and mutual information.

For two general sources  $\mathbf{X}$  and  $\mathbf{Y}_{(j)}$ , the *spectral conditional entropy*  $\bar{H}(\mathbf{X}|\mathbf{Y}_{(j)})$  is defined from a lim sup in probability. The lim sup in probability of a sequence  $\{A_n\}_{n=1}^{+\infty}$  of random variables is denoted as  $\text{p-lim sup}_{n \rightarrow \infty} A_n$  and is given by

$$\text{p-lim sup}_{n \rightarrow \infty} A_n = \inf \left\{ \alpha \mid \lim_{n \rightarrow +\infty} \mathbb{P}(A_n > \alpha) = 0 \right\}. \quad (3)$$

The spectral conditional entropy  $\bar{H}(\mathbf{X}|\mathbf{Y}_{(j)})$  is then defined as in [22, Chapter 7]

$$\bar{H}(\mathbf{X}|\mathbf{Y}_{(j)}) = \text{p-lim sup}_{n \rightarrow \infty} -\frac{1}{n} \log P_{\mathbf{X}^n|\mathbf{Y}_{(j)}^n}(\mathbf{X}^n|\mathbf{Y}_{(j)}^n) \quad (4)$$

In (4), the vectors  $\mathbf{X}^n, \mathbf{Y}_{(j)}^n$  are random vectors distributed according to the joint probability distribution  $P_{\mathbf{X}^n, \mathbf{Y}_{(j)}^n}$  that can be obtained from  $P_{\mathbf{X}^n, \mathbf{Y}_{(1)}^n, \dots, \mathbf{Y}_{(J)}^n}$ . In the above definition, the lim sup in probability is required because general sources are not necessarily ergodic, and hence the random variable  $-\frac{1}{n} \log P_{\mathbf{X}^n|\mathbf{Y}_{(j)}^n}(\mathbf{X}^n|\mathbf{Y}_{(j)}^n)$  does not necessarily converge to a unique value.

According to [22], if the sources  $\mathbf{X}$  and  $\mathbf{Y}_{(j)}$  are i.i.d., the spectral conditional entropy reduces to the standard entropy, that is  $\bar{H}(\mathbf{X}|\mathbf{Y}_{(j)}) = H(X|Y_{(j)})$ . If the sources  $\mathbf{X}$  and  $\mathbf{Y}_{(j)}$  are ergodic, then the spectral conditional entropy reduces to the entropy rate, that is  $\bar{H}(\mathbf{X}|\mathbf{Y}_{(j)}) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n H(X_i|Y_{(j),i})$ .

For two general sources  $\mathbf{X}$  and  $\mathbf{Y}_{(j)}$ , the *spectral mutual information* is defined as [32]

$$\bar{I}(\mathbf{X}; \mathbf{Y}_{(j)}) = \text{p-lim sup}_{n \rightarrow \infty} \frac{1}{n} \log \frac{P_{\mathbf{X}^n|\mathbf{Y}_{(j)}^n}(\mathbf{X}^n|\mathbf{Y}_{(j)}^n)}{P_{\mathbf{X}^n}(\mathbf{X}^n)}. \quad (5)$$

As for spectral conditional entropy, spectral mutual information reduces to standard definitions of mutual information when i.i.d. and ergodic sources are considered.

### D. Definitions of Coding schemes

This section formally defines the ESC-SI coding scheme, as well as the lossless and lossy source coding problems we consider in this paper.

**Definition 3** (ESC-SI code). A  $((2^{nS}, (2^{nR_{(j)}})_{j \in \llbracket 1, J \rrbracket}), n)$  ESC-SI code for the set of general sources  $\{X, \{Y_{(j)}\}_{1 \leq j \leq J}\}$  consists of

- an offline encoder  $h^{\text{off}}$  that assigns a sequence of  $nS$  bits to each possible vector  $\mathbf{x}^n \in \mathcal{X}^n$

$$h^{\text{off}} : \mathcal{X}^n \rightarrow \{0, 1\}^{nS} \quad (6a)$$

$$\mathbf{x}^n \mapsto (b_1, \dots, b_{nS}). \quad (6b)$$

- a set of  $J$  online extractors  $h_{(j)}^{\text{on}}$ ,  $j \in \llbracket 1, J \rrbracket$ , that extract a subsequence of bits from the sequence of bits  $(b_1, \dots, b_{nS})$

$$h_{(j)}^{\text{on}} : \{0, 1\}^{nS} \rightarrow \{0, 1\}^{nR_{(j)}} \quad (7a)$$

$$(b_1, \dots, b_{nS}) \mapsto (b_k)_{k \in \mathcal{I}_{(j)}} \quad (7b)$$

where  $\mathcal{I}_{(j)} \subseteq \{1, \dots, nS\}$ , and  $|\mathcal{I}_{(j)}| = nR_{(j)}$ .

- a set of  $J$  decoders  $g_{(j)}$ ,  $j \in \llbracket 1, J \rrbracket$ , that, given the source realization  $\mathbf{y}_{(j)}^n$ , builds an estimate  $\hat{\mathbf{x}}_{(j)}^n$  from the received subsequence of bits

$$g_{(j)} : \{0, 1\}^{nR_{(j)}} \times \mathcal{Y}_j^n \rightarrow \mathcal{X}^n \quad (8a)$$

$$(b_k)_{k \in \mathcal{I}_{(j)}}, \mathbf{y}_{(j)}^n \mapsto \hat{\mathbf{x}}_{(j)}^n. \quad (8b)$$

The main novelty of the above definition resides in the combination of two encoding mappings: a standard offline encoder that produces the coded source description, and an online extractor that can only extract a part of the coded description. In this definition, the offline encoder provides one single common source description  $(b_1, \dots, b_{nS})$  stored on the server, while the online extractor can construct one codeword  $(b_k)_{k \in \mathcal{I}_{(j)}}$  per available side information  $\mathbf{Y}_{(j)}$ . This may allow to reduce both the storage rate  $S$  on the server, and the transmission rates  $R_{(j)}$  from the server to the decoders. In the successive refinement problem considered in [7], [9], [10], the code construction is incremental in the sense that the set of indices provided by the online extractors are incremental:  $\mathcal{I}_{(j)} \subseteq \mathcal{I}_{(j+1)}$ . Our problem is more general since no such constraint appears in the coding scheme definition.

We now provide definitions of the lossless and lossy source coding problems which we consider for ESC-SI codes.

**Definition 4** (Achievable storage and transmission rates region for a lossless ESC-SI code). *The probability of error for a ESC-SI code is defined as*

$$P_{\text{error}}^n = \mathbb{P} \left( \bigcup_{j \in \llbracket 1, J \rrbracket} (\hat{\mathbf{x}}_{(j)}^n \neq \mathbf{x}^n) \right) \quad (9)$$

A tuple of rates  $(S, (R_{(j)})_{j^* \in \llbracket 1, J \rrbracket})$  is said to be achievable for ESC-SI if there exists a sequence of  $(2^{nS}, (2^{nR_{(j)}})_{j^* \in \llbracket 1, J \rrbracket}, n)$  codes such that  $\lim_{n \rightarrow +\infty} P_{error}^n = 0$ . The achievable storage and transmission rate region is the closure of the set of achievable rate tuples.

**Definition 5** (Achievable storage, transmission rates, distortions region for a lossy ESC-SI code). Consider a quadratic distortion measure  $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  such that  $d(x, \hat{x}) = (x - \hat{x})^2$ . The tuple  $(S, (R_{(j)})_{j \in \llbracket 1, J \rrbracket}, (D_{(j)})_{j \in \llbracket 1, J \rrbracket})$  is said to be achievable for ESC-SI if there exists a sequence of ESC-SI codes such that

$$\forall j \in \llbracket 1, J \rrbracket, \lim_{n \rightarrow \infty} E \left[ \frac{1}{n} \sum_{i=1}^n d(X_i, \hat{X}_{(j),i}) \right] \leq D_{(j)}, \quad (10)$$

where the expectation is taken over  $(\mathbf{X}^n, \mathbf{Y}_{(j)}^n)$ .

### III. LOSSLESS CODING SCHEME

In this section, we provide the storage and transmission rate regions for lossless ESC-SI for general sources. The example of a set of heterogeneous side informations is then considered.

#### A. Storage and transmission rate regions for general sources

From the definition of the spectral conditional entropy in (4), the storage and transmission rates region is obtained as follows.

**Theorem 1.** Let  $\{\mathbf{X}, \{\mathbf{Y}_{(j)}\}_{1 \leq j \leq J}\}$  be a set of  $J + 1$  general sources (see Definition 1), and assume that the alphabet  $\mathcal{X}$  is discrete. The achievable storage and transmission rate region for ESC-SI (see Definition 4) is the set of rate tuples  $(S, (R_{(j)})_{j \in \llbracket 1, J \rrbracket})$  such that  $\forall j \in \llbracket 1, J \rrbracket$ ,

$$\begin{aligned} S &\geq \max_{j \in \llbracket 1, J \rrbracket} \bar{H}(\mathbf{X} | \mathbf{Y}_{(j)}), \\ R_{(j)} &\geq \bar{H}(\mathbf{X} | \mathbf{Y}_{(j)}). \end{aligned} \quad (11)$$

*Proof:* See Appendix. ■

This result shows that it is possible to store one single description of the source  $\mathbf{X}$  at a rate given by the worst-case spectral conditional entropy  $\max \bar{H}(\mathbf{X} | \mathbf{Y}_{(j)})$ , from which we can extract  $J$  descriptions at rates given by the individual spectral conditional entropies  $\bar{H}(\mathbf{X} | \mathbf{Y}_{(j)})$ . In the proof of Theorem 1, the coding scheme that achieves (11) is incremental. Since the side informations  $\mathbf{Y}_{(1)}, \dots, \mathbf{Y}_{(J)}$  are not

assumed to be physically degraded, the  $\mathbf{Y}_{(j)}$  are ordered according to their spectral conditional entropies  $\bar{H}(\mathbf{X}|\mathbf{Y}_{(j)})$ . This generalizes the incremental criterion that was proposed in [10] for the problem of successive refinement with i.i.d. sources.

The next two corollaries provide the storage and transmission rate region for the particular cases of i.i.d. and ergodic sources.

**Corollary 2.** *For a set  $\{\mathbf{X}, \{\mathbf{Y}_{(j)}\}_{1 \leq j \leq J}\}$  of  $(J+1)$  i.i.d. sources, the achievable storage and transmission rate region for ESC-SI is given by*

$$\begin{aligned} S &\geq \max_{j \in \llbracket 1, J \rrbracket} H(X|Y_{(j)}), \\ R_{(j)} &\geq H(X|Y_{(j)}). \end{aligned} \quad (12)$$

**Corollary 3.** *For a set  $\{\mathbf{X}, \{\mathbf{Y}_{(j)}\}_{1 \leq j \leq J}\}$  of  $(J+1)$  ergodic sources, the achievable storage and transmission rate region for ESC-SI is given by*

$$\begin{aligned} S &\geq \max_{j \in \llbracket 1, J \rrbracket} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n H(X_i|Y_{(j),i}), \\ R_{(j)} &\geq \lim_{n \rightarrow \infty} \frac{1}{n} H(X_i|Y_{(j),i}). \end{aligned} \quad (13)$$

To finish, since no assumption is performed on the alphabets  $\mathcal{Y}_{(1)}, \dots, \mathcal{Y}_{(J)}$ , Theorem 1 allows one to consider heterogeneous sets of side informations, as will be seen in the following example.

### B. Lossless scheme considering a set of heterogeneous side informations

We now describe an example of a set of general sources. We consider a binary alphabet  $\mathcal{X} = \{0, 1\}$ , and an i.i.d. source  $X$  with  $P(X = 0) = 0.5$ . We also consider four different side informations  $Y_{(1)}, \dots, Y_{(4)}$  as follows:

- $Y_{(1)}$  is obtained from  $X$  via a BEC of parameter  $\epsilon$ , and  $H(X|Y_{(1)}) = 2\epsilon - H(\epsilon)$ .
- $Y_{(2)}$  is obtained from  $X$  via a BSC of parameter  $p$ , and  $H(X|Y_{(2)}) = H(p)$ .
- $Y_{(3)}$  is obtained from  $X$  via the following GE channel. Consider two sequences of binary random variables  $\{U_i\}_{i \geq 1}$ ,  $\{Z_i\}_{i \geq 1}$  such that  $P(U_i = 1|U_{i-1} = 0) = \alpha$ ,  $P(U_i = 0|U_{i-1} = 1) = \beta$ ,  $P(Z_i = 1|U_i = 0) = p_0$ ,  $P(Z_i = 1|U_i = 1) = p_1$ . The GE channel is described by  $Y_{(3),i} = X_i + Z_i$ . In this case, the conditional entropy rate  $H(X|Y_{(3)})$  can be calculated using [33] or evaluated numerically.

- $Y_{(4)}$  is obtained from  $X$  via a Gaussian channel such that  $(Y|X = x) \sim \mathcal{N}(1 - 2x, \sigma^2)$ , and  $H(X|Y_{(4)}) = 1 - \frac{1}{2} \log_2(2\pi e\sigma^2) + h(Y_{(4)})$ , where  $h(Y_{(4)})$  is the differential entropy of a Gaussian mixture. The term  $h(Y_{(4)})$  can be approximated using [34] or evaluated numerically.

These four side informations are not degraded versions of each other, since they correspond to different alphabets  $\mathcal{Y}_j$ . Applying Theorem 1 gives that the transmission rates  $R_{(1)}, \dots, R_{(4)}$  are given by the conditional entropies  $H(X|Y_{(1)}), \dots, H(X|Y_{(4)})$ , and that the storage rate is given by the maximum of these terms. The source ordering and the storage rate depend on the respective values of the parameters of each considered model. In order to construct an incremental coding scheme for ESC-SI, it suffices to order the four conditional entropies.

#### IV. LOSSY CODING SCHEME

We now provide the rate-distortions regions for lossy ESC-SI for non i.i.d. Gaussian sources.

##### A. Rate-distortion region for non i.i.d. Gaussian sources

The following theorem provides the achievable tuple of rates and distortions for non i.i.d. Gaussian sources for ESC-SI.

**Theorem 4.** *Let  $\{\mathbf{X}, \{\mathbf{Y}_{(j)}\}_{1 \leq j \leq J}\}$  be a set of  $J + 1$  Gaussian sources (see Definition 2). The rate-distortion tuple  $(S, (R_{(j)}, D_{(j)})_{j \in \llbracket 1, J \rrbracket})$  is achievable for Gaussian sources for ESC-SI if there exists  $\delta > 0$  such that*

$$S \geq \max_{j \in \llbracket 1, J \rrbracket} \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \max \left( 0, \frac{1}{2} \log_2 \frac{\lambda_i^{(X|j)}}{d_{(j),i}} \right) \quad (14)$$

$$\forall j \in \llbracket 1, J \rrbracket, R_{(j)} \geq \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \max \left( 0, \frac{1}{2} \log_2 \frac{\lambda_i^{(X|j)}}{d_{(j),i}} \right) \quad (15)$$

$$\forall j \in \llbracket 1, J \rrbracket, D_{(j)} \geq \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \min \left( \lambda_i^{(X|j)}, d_{(j),i} \right), \quad (16)$$

where  $d_{(j),i} = \frac{\lambda_i^{(X|j)} \delta}{\lambda_i^{(X|j)} + \delta}$ .

*Proof:* See Appendix. ■

In the above theorem, the storage rate  $S$  is given by the worst possible transmission rate  $R_{(j)}$ . We notice that the transmission rate  $R_{(j)}$  corresponds to the Wyner-Ziv rate-distortion function for a given target distortion  $D_{(j)}^* = \limsup \frac{1}{n} \sum_{i=1}^n \min \left( \lambda_i^{(X|j)}, d_{(j),i} \right)$  when  $\mathbf{Y}_{(j)}$  is the only possible side information. The parameter  $\delta$  is the distortion of an individual component when no side information is available at

the decoder. When a side information  $\mathbf{Y}_{(j)}$  is available at the decoder, the source  $\mathbf{X}$  can be reconstructed with a distortion  $D_{(j)}^* \leq \delta$  that depends on the parameter  $\delta$  and on the statistics between  $\mathbf{X}$  and  $\mathbf{Y}_{(j)}$ .

In the proof of Theorem 4, the sources are ordered according to the spectral mutual information terms  $\bar{I}(\mathbf{Y}_{(j)}; \mathbf{U})$ , where  $\mathbf{U}$  is an auxiliary source that generates vectors  $\mathbf{U}^n = \mathbf{X}^n + \mathbf{\Psi}^n$ , where  $\mathbf{\Psi}^n$  is an i.i.d. centered Gaussian random variable with variance  $\delta$ . For Gaussian sources, the spectral mutual information  $\bar{I}(\mathbf{Y}_{(j)}; \mathbf{U})$  can be expressed as

$$\bar{I}(\mathbf{Y}_{(j)}; \mathbf{U}) = \lim_{n \rightarrow +\infty} \sup \frac{1}{n} \sum_{i=1}^n \frac{1}{2} \log \frac{\lambda_i^{(X)} \delta}{\lambda_i^{(X|j)} + \delta}, \quad (17)$$

where  $\lambda_i^{(X)}$  are the eigenvalues of the covariance matrix  $\Sigma_X$ . It is worth noting that this incremental criterion depends on the eigenvalues of the two covariance matrices  $\Sigma_X$  and  $\Sigma_{(X|j)}^n$ , and also on the parameter  $\delta$ .

From the proof of Theorem 4, we can also derive the achievable tuple of rates and distortions for non i.i.d. discrete sources for ESC-SI as follows.

**Theorem 5.** *Let  $\{\mathbf{X}, \{\mathbf{Y}_{(j)}\}_{1 \leq j \leq J}\}$  be a set of  $J + 1$  general sources (see Definition 1), and assume that the alphabet  $\mathcal{X}$  is discrete. The rate-distortion tuple  $(S, (R_{(j)}, D_{(j)})_{j \in \llbracket 1, J \rrbracket})$  is achievable for ESC-SI if there exists  $\mathbf{U}^n$  and  $g_{(j)} : \mathcal{U}^n \times \mathcal{Y}_{(j)}^n \rightarrow \mathcal{X}^n$  such that: 1)  $\forall j \in \llbracket 1, J \rrbracket, \mathbf{Y}_{(j)}^n \leftrightarrow \mathbf{X}^n \leftrightarrow \mathbf{U}^n$  forms a Markov chain, 2)  $\hat{\mathbf{X}}_{(j)}^n = g_{(j)}(\mathbf{Y}_{(j)}^n, \mathbf{U}^n)$  satisfies  $E[d(\mathbf{X}^n, \hat{\mathbf{X}}_{(j)}^n)] \leq D_{(j)}$ , and 3) the following rate conditions are satisfied:*

$$S \geq \max_{j \in \llbracket 1, J \rrbracket} (\bar{I}(\mathbf{X}; \mathbf{U}) - \bar{I}(\mathbf{Y}_{(j)}; \mathbf{U})) \quad (18)$$

$$\forall j \in \llbracket 1, J \rrbracket, R_{(j)} \geq \bar{I}(\mathbf{X}; \mathbf{U}) - \bar{I}(\mathbf{Y}_{(j)}; \mathbf{U}). \quad (19)$$

$$(20)$$

*Proof:* See Appendix. ■

Note that obtaining a converse result for Theorems 4 and 5 is a difficult problem, due to the extractable constraint in Definition 3. For instance, due to this constraint, in the achievability result of Theorem 4, all the distortions  $d_{(j),i}$  are linked to each other through the parameter  $\delta$ .

Finally, we compare the region obtained in Theorem 4 to existing results in the literature. First, the broadcast setup described in Figure 1(b) was studied theoretically in [26], [27] and applied to FTV

in [18], [20]. Applying these results to our setup gives the following rate-distortion region:

$$S \geq \max_{j \in \llbracket 1, J \rrbracket} \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \max \left( 0, \frac{1}{2} \log_2 \frac{\lambda_i^{(X|j)}}{d_{(j),i}} \right) \quad (21)$$

$$\forall j \in \llbracket 1, J \rrbracket, R_{(j)} \geq \max_{j \in \llbracket 1, J \rrbracket} \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \max \left( 0, \frac{1}{2} \log_2 \frac{\lambda_i^{(X|j)}}{d_{(j),i}} \right) \quad (22)$$

$$\forall j \in \llbracket 1, J \rrbracket, D_{(j)} \geq \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \min \left( \lambda_i^{(X|j)}, d_{(j),i} \right). \quad (23)$$

The broadcast solution gives the same storage rate as in Theorem 4 but it is sub-optimal in terms of transmission rates, since  $R_{(j)} = S$  for all  $j$ . An opposite strategy consists of applying standard Wyner-Ziv results [2] by assuming  $J$  independent encoders and decoders. This strategy was applied into the practical coding schemes of [16], [17] for FTV. In this case, the rate-distortion region is given by

$$S \geq \sum_{j \in \llbracket 1, J \rrbracket} \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \max \left( 0, \frac{1}{2} \log_2 \frac{\lambda_i^{(X|j)}}{d_{(j),i}} \right) \quad (24)$$

$$\forall j \in \llbracket 1, J \rrbracket, R_{(j)} \geq \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \max \left( 0, \frac{1}{2} \log_2 \frac{\lambda_i^{(X|j)}}{d_{(j),i}} \right) \quad (25)$$

$$\forall j \in \llbracket 1, J \rrbracket, D_{(j)} \geq \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \min \left( \lambda_i^{(X|j)}, d_{(j),i} \right). \quad (26)$$

This allows to reach the same transmission rates as in Theorem 4 but the storage rate is the sum of all transmission rates and not the maximum as derived in Theorem 4. Therefore, the coding schemes of [16], [17] applied to our framework are sub-optimal in terms of storage rate. At the end, we identified that the successive refinement scheme with side information [7]–[11] is optimal for our setup in the sense that it achieves the same storage rate as in the broadcast setup and the same transmission rates as in the Wyner-Ziv setup. Previous results on successive refinement however mainly consider i.i.d. sources [7]–[10] and do not permit to provide the rate-distortion regions for the ergodic Gaussian sources we now consider.

### B. Examples for ergodic Gaussian sources

When the Gaussian sources are ergodic, the  $\limsup$  in Theorem 4 become a  $\lim$ . We consider an example of such ergodic sources, and, for simplicity, we directly define the conditional covariance matrix  $\Sigma_{(X|j)}^n$  whose eigenvalues participate to the rate-distortions regions defined in Theorem 4. In order to get

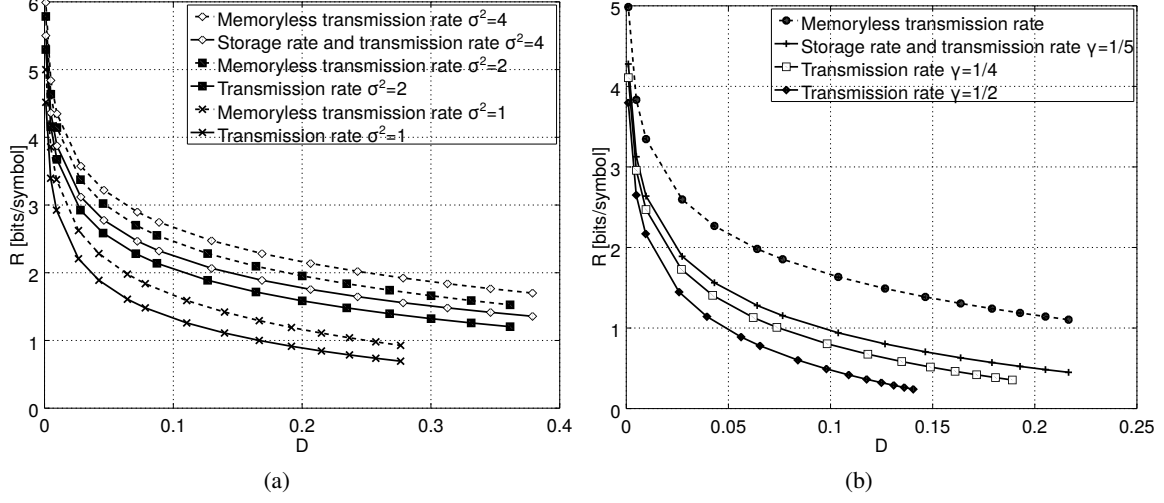


Fig. 2: Rates  $S$  and  $R_{(j)}$  with respect to distortions  $D_{(j)}$  for three side informations  $Y_{(j)}$ : (a) Nearest Neighbor correlation model, (b) First-order Markov model

an ergodic model, we consider positive-definite Toeplitz matrices with expressions

$$\Sigma_{(X|j)}^n = \begin{bmatrix} \rho_1^{(X|j)} & \rho_2^{(X|j)} & \rho_3^{(X|j)} & \cdots & \rho_n^{(X|j)} \\ \rho_2^{(X|j)} & \rho_1^{(X|j)} & \rho_2^{(X|j)} & \cdots & \rho_{n-1}^{(X|j)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_n^{(X|j)} & \rho_{n-1}^{(X|j)} & \cdots & \rho_2^{(X|j)} & \rho_1^{(X|j)} \end{bmatrix}. \quad (27)$$

With this model, finite memory can be considered by setting  $\rho_i^{(X|j)} = 0$  for all  $i$  greater than a given integer. Since the matrices  $\Sigma_{(X|j)}^n$  are Toeplitz matrices, we can apply [35, Theorem 4.5.2] in order to show that the tuple  $(S, (R_{(j)}, D_{(j)})_{(j) \in \llbracket 1, J \rrbracket})$  is achievable if

$$R_{(j)} \geq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \max \left( 0, \frac{1}{2} \log_2 \frac{\lambda_i^{(X|j)} + \delta}{\delta} \right) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \max \left( 0, \log \frac{\Phi^{(X|j^*)}(x) + \delta}{\delta} \right) dx \quad (28)$$

$$D_{(j)} \geq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \min \left( \lambda_i^{(X|j)}, \frac{\lambda_i^{(X|j)} \delta}{\lambda_i^{(X|j)} + \delta} \right) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min \left( \Phi^{(X|j)}(x), \frac{\Phi^{(X|j)}(x) \delta}{\Phi^{(X|j)}(x) + \delta} \right) dx, \quad (29)$$

and  $S \geq \max_j R_{(j)}$ , where the function  $\Phi^{(X|j)}$  is defined as

$$\Phi^{(X|j)}(x) = \rho_0^{(X|j)} + 2 \sum_{n=1}^{+\infty} \rho_n^{(X|j)} \exp(-inx), \quad (30)$$

and  $i$  is the imaginary unit.



As examples, we consider two Gaussian models that are particular cases of the model defined in (27). These two models have been introduced in [35] for the standard case of lossy source coding without side information. They cannot be considered in the Gaussian model of [11] since they contain some sliding memory effect. In order to obtain approximations of the rate and distortion expressions provided in Theorem 4, we computed the eigenvalues of the three Toeplitz matrices obtained with each considered model at length  $N = 1000$ .

*a) Nearest Neighbor correlation:* We first consider the Gaussian model of Definition 2 and (27) with  $J = 3$  sources, and  $\rho_1^{(X|j)} = \sigma_j^2$ ,  $\rho_2^{(X|j)} = \sigma_j^2/2$ , and  $\rho_i^{(X|j)} = 0$ ,  $\forall i > 2$ . This model assumes that each component  $i$  in  $\mathbf{X}^n$  is only correlated with the components  $i - 1$ ,  $i$ , and  $i + 1$  of  $\mathbf{Y}_{(j)}^n$ . In order to completely define the model for each of the three side information vectors, we set  $\sigma_1^2 = 1$ ,  $\sigma_2^2 = 2$ ,  $\sigma_3^2 = 4$ . The rate-distortion functions for ESC-SI for this model are represented in Figure 2 (a). As expected, the storage rate is superimposed with the worst possible transmission rate. Figure 2 (a) also shows the rate-distortion functions for the memoryless Gaussian model ( $\rho_1^{(X|j)} = \sigma_j^2$  and  $\rho_i^{(X|j)} = 0$ ,  $\forall i > 1$ ), which illustrates the potential gain at taking the memory into account.

*b) First-order Markov source:* We now consider the Gaussian model with  $J = 3$  sources and  $\rho_i^{(X|j)} = \sigma^2 |\gamma_j|^{i-1}$ , with  $-1 < \gamma_j < 1$ , for all  $j \in \llbracket 1, J \rrbracket$ . This model assumes that all the components of  $\mathbf{X}^n$  and  $\mathbf{Y}_{(j)}^n$  are correlated with a level  $\sigma^2 |\gamma_j|^i$  decreasing with  $i$ . For this model, we set  $\sigma^2 = 1$ ,  $\gamma_1 = 1/2$ ,  $\gamma_2 = 1/4$ , and  $\gamma_3 = 1/5$ . The corresponding rate-distortion functions for ESC-SI are represented in Figure 2 (b), as well as the rate-distortion functions for the memoryless case. In this case, we observe that the memoryless rate-distortion functions are the same whatever the previously requested source  $Y_{(j)}$  available at the decoder, while taking the memory into account permits a decrease in the transmission rates and distortions.

*c) Combination of the two models:* According to the assumptions of Definition 2 and of Theorem 4, the above two sets can be combined into a unique set of 6 side informations following either one or the other model. In this case, the individual transmission rates and distortions do no change. The storage rate can be obtained from (16) by combining the worst rates from component to component.

## V. CONCLUSION

This paper defines the ESC-SI source coding problem, in which we want to minimize both the storage rate of the source  $\mathbf{X}$  on the server, and the individual transmission rates from the server to  $J$  users, each having access to a different side information  $\mathbf{Y}_{(j)}$ . It provides the achievable storage and transmission rate regions for lossless source coding of general sources, and the achievable storage-transmission rate-

distortion regions for lossy coding of non i.i.d. Gaussian sources. Future works will be dedicated to the design of practical coding schemes for this problem.

## APPENDIX

### PROOF OF THEOREM 1

**Achievability.** We show the existence of a sequence of codes that achieves the set of rates (11).

*Source ordering.* The first step consists in ordering the set  $\{\mathbf{Y}_{(j)}\}_{1 \leq j \leq J}$  of side informations. Let

$$\pi : \llbracket 1, J \rrbracket \rightarrow \llbracket 1, M \rrbracket \quad (31a)$$

$$j \mapsto \pi(j) \quad (31b)$$

be an ordering mapping of the  $Y_j$  such that

$$\bar{H}(\mathbf{X}|\mathbf{Y}_{(\pi^{-1}(1))}) \leq \dots \leq \bar{H}(\mathbf{X}|\mathbf{Y}_{(\pi^{-1}(m))}) \leq \dots \leq \bar{H}(\mathbf{X}|\mathbf{Y}_{(\pi^{-1}(M))}),$$

where  $m = \pi(j)$  stands for the order of a particular index  $j$ .

*Random code generation* is performed recursively. The idea is to first partition the space  $\mathcal{X}^n$  into  $2^{nr_1}$  bins, and then to further partition each bin into  $2^{nr_2}$  bins. This process is repeated  $M$  times and creates an embedded random binning.

- First, assign every  $\mathbf{x}^n \in \mathcal{X}^n$  to one of  $2^{nr_1}$  bins independently according to a uniform distribution on  $\llbracket 1, 2^{nr_1} \rrbracket$ . The binning mapping is defined as  $f_1(\mathbf{x}^n) = i_1$ , where  $i_1$  is the index of the bin assigned to  $\mathbf{x}^n$ . Let the bin  $\mathcal{B}_{i_1} = f_1^{-1}(i_1)$  denote the preimage of  $i_1$  under the map  $f_1$ .
- At the second stage, for each bin index  $i_1 \in \llbracket 1, 2^{nr_1} \rrbracket$ , assign every  $\mathbf{x}^n \in \mathcal{B}_{i_1}$  independently according to a uniform distribution on  $\llbracket 1, 2^{nr_2} \rrbracket$ . The binning mapping is defined as  $f_2(\mathbf{x}^n) = (i_1, i_2)$ , where  $i_1$  and  $i_2$  are the indexes assigned to  $\mathbf{x}^n$  at the first and second stage respectively. Let the bin  $\mathcal{B}_{i_1, i_2} = f_2^{-1}(i_1, i_2)$  denote the preimage of  $(i_1, i_2)$  under the map  $f_2$ .
- Repeat the sub-partitioning up to stage  $M$ . At the last stage, the mapping is  $f_M(\mathbf{x}^n) = (i_1, i_2, \dots, i_M)$ .

Finally, the binning is defined by a set of mappings  $\{f_m\}_{1 \leq m \leq M}$ , which is revealed to both the encoder and the decoder.

*Encoding* is performed in two parts, see Definition 3. The *offline encoder* computes a sequence of bin indexes for the source sequence  $\mathbf{x}^n$ :

$$h^{\text{off}}(\mathbf{x}^n, (\mathbf{y}_{(j)}^n)_{j \in \mathcal{J}}) = f_M(\mathbf{x}^n) = (i_1, i_2, \dots, i_M).$$

The  $j$ -th *online extractor* extracts the  $\pi(j)$  first bin indexes from the full sequence of bin indexes:

$$h_{(j)}^{\text{on}}(i_1, i_2, \dots, i_M) = (i_1, i_2, \dots, i_{\pi(j)}).$$

*Decoding.* Define the set  $T_\epsilon^n(\mathbf{X}, \mathbf{Y}_{(j)})$  as

$$T_\epsilon^n(\mathbf{X}, \mathbf{Y}_{(j)}) = \left\{ (\mathbf{x}^n, \mathbf{y}_{(j)}^n) \text{ s.t. } \frac{1}{n} \log \frac{1}{P_{\mathbf{X}^n | \mathbf{Y}_{(j)}^n}(\mathbf{x}^n | \mathbf{y}_{(j)}^n)} < \sum_{i=1}^{\pi(j)} r_i - \epsilon \right\}. \quad (32)$$

Given the received index sequence  $(i_1, \dots, i_{\pi(j)})$  and the side information  $\mathbf{y}_{(j)}^n$ , declare  $\hat{\mathbf{x}}^n = \mathbf{x}^n$  if there is a unique pair of sequences  $(\mathbf{x}^n, \mathbf{y}_{(j)}^n)$  such that  $f_{\pi(j)}(\mathbf{x}^n) = (i_1, \dots, i_{\pi(j)})$  and  $(\mathbf{x}^n, \mathbf{y}_{(j)}^n) \in T_\epsilon^n(\mathbf{X}, \mathbf{Y}_{(j)})$ . Otherwise, declare an error.

*Probability of error.* We define the events

$$E_0 = \{(\mathbf{X}^n, \mathbf{Y}_{(j)}^n) \notin T_\epsilon^n(\mathbf{X}, \mathbf{Y}_{(j)})\} \quad (33)$$

$$E_j = \{ \exists \mathbf{x}'^n \neq \mathbf{X}^n : f_{\pi(j)}(\mathbf{x}'^n) = f_{\pi(j)}(\mathbf{X}^n) \\ \text{and } (\mathbf{x}'^n, \mathbf{Y}_{(j)}^n) \in T_\epsilon^n(\mathbf{X}, \mathbf{Y}_{(j)}) \}, \quad \forall j \in \mathcal{J}.$$

$\forall j \geq 1$ , the error probability  $\mathbb{P}(E_j)$  can be derived as

$$\begin{aligned} \mathbb{P}(E_j) &= \sum_{(\mathbf{x}^n, \mathbf{y}_{(j)}^n)} P_{\mathbf{X}^n, \mathbf{Y}_{(j)}^n}(\mathbf{x}^n, \mathbf{y}_{(j)}^n) \dots \\ &\quad \mathbb{P}(\exists \mathbf{x}'^n \neq \mathbf{x}^n : f_{\pi(j)}(\mathbf{x}'^n) = f_{\pi(j)}(\mathbf{x}^n) \text{ and } (\mathbf{x}'^n, \mathbf{y}_{(j)}^n) \in T_\epsilon^n) \\ &\leq \sum_{(\mathbf{x}^n, \mathbf{y}_{(j)}^n)} P_{\mathbf{X}^n, \mathbf{Y}_{(j)}^n}(\mathbf{x}^n, \mathbf{y}_{(j)}^n) \sum_{\substack{\mathbf{x}'^n \neq \mathbf{x}^n \\ (\mathbf{x}'^n, \mathbf{y}_{(j)}^n) \in T_\epsilon^n}} \mathbb{P}(f_{\pi(j)}(\mathbf{x}'^n) = f_{\pi(j)}(\mathbf{x}^n)) \end{aligned} \quad (34)$$

$$\leq \sum_{(\mathbf{x}^n, \mathbf{y}_{(j)}^n)} P_{\mathbf{X}^n, \mathbf{Y}_{(j)}^n}(\mathbf{x}^n, \mathbf{y}_{(j)}^n) \cdot |T_\epsilon^n(\mathbf{X} | \mathbf{y}_{(j)}^n)| \cdot 2^{-n \left( \sum_{i=1}^{\pi(j)} r_i \right)} \quad (35)$$

where the set  $T_\epsilon^n(\mathbf{X} | \mathbf{y}_{(j)}^n)$  is defined as

$$T_\epsilon^n(\mathbf{X} | \mathbf{y}_{(j)}^n) = \left\{ \mathbf{x}^n \text{ s.t. } \frac{1}{n} \log \frac{1}{P_{\mathbf{X}^n | \mathbf{Y}_{(j)}^n}(\mathbf{x}^n | \mathbf{y}_{(j)}^n)} < \sum_{i=1}^{\pi(j)} r_i - \epsilon \right\}. \quad (36)$$

Here,  $|T_\epsilon^n(\mathbf{X} | \mathbf{y}_{(j)}^n)|$  can be bounded using the fact that

$$1 \geq \sum_{\mathbf{x}'^n \in T_\epsilon^n(\mathbf{X} | \mathbf{y}_{(j)}^n)} P_{\mathbf{X}^n | \mathbf{Y}_{(j)}^n}(\mathbf{x}'^n | \mathbf{y}_{(j)}^n) \geq |T_\epsilon^n(\mathbf{X} | \mathbf{y}_{(j)}^n)| \exp \left( -n \sum_{i=1}^{\pi(j)} r_i + n\epsilon \right) \quad (37)$$

where  $P_{\mathbf{X}^n|\mathbf{Y}_{(j)}^n}(\mathbf{x}^n|\mathbf{y}_{(j)}^n) \geq \exp(-n \sum r_i + n\varepsilon)$  comes from the definition of  $T_\varepsilon^{(n)}(\mathbf{X}|\mathbf{y}_{(j)}^n)$  in (36). Equation (37) gives

$$|T_\varepsilon^{(n)}(\mathbf{X}|\mathbf{y}_{(j)}^n)| \leq \exp\left(n \sum_{i=1}^{\pi(j)} r_i - n\varepsilon\right) \quad (38)$$

and as a consequence the error probability  $\mathbb{P}(E_j)$  becomes  $\mathbb{P}(E_j) \leq \exp(-n\varepsilon)$ .

Next, according to the definition of  $T_\varepsilon^{(n)}(\mathbf{X}, \mathbf{Y}_{(j)})$  in (45), the error probability  $\mathbb{P}(E_0)$  can be expressed as

$$\mathbb{P}(E_0) = \mathbb{P}\left(\frac{1}{n} \log \frac{1}{P_{\mathbf{X}^n|\mathbf{Y}_{(j)}^n}(\mathbf{X}^n|\mathbf{Y}_{(j)}^n)} \geq \sum_{i=1}^{\pi(j)} r_i + \varepsilon\right) \quad (39)$$

If  $\sum_{i=1}^{\pi(j)} r_i \geq \bar{H}(\mathbf{X}|\mathbf{Y}_{(j)})$ , then from the definition of  $\bar{H}(\mathbf{X}|\mathbf{Y}_{(j)})$  in (4),  $\mathbb{P}(E_0)$  goes to 0 when  $n$  goes to infinity.

From the analysis of  $\mathbb{P}(E_0)$  and  $\mathbb{P}(E_j)$ , and from the fact that the error probability  $P_{\text{error}}^n$  can be bounded as  $P_{\text{error}}^n \leq \mathbb{P}(E_0) + \sum_j \mathbb{P}(E_j)$ , there exists at least one code  $\{f_{\pi(j)}\}_j$  with probability of error less than  $(J+1)\varepsilon$ . Thus, if  $J$  is finite, we can construct a sequence of codes with  $P_{\text{error}}^n \rightarrow 0$ .

In other words, for any  $j \in \llbracket 1, J \rrbracket$ , there exists a code with vanishing error rate and rate  $r_1 + \dots + r_{\pi(j)}$ , provided that  $r_1 + \dots + r_{\pi(j)} > \bar{H}(\mathbf{X}|\mathbf{Y}_{(j)})$ . Finally, to be able to handle any  $\mathbf{Y}_{(j)}$ , the storage rate is such that  $S > \max_{j \in \mathcal{J}} \bar{H}(\mathbf{X}|\mathbf{Y}_{(j)})$ . Moreover, for side information  $\mathbf{Y}_{(j)}$ , the transmission rate  $R_{(j)}$  needs to satisfy  $R_{(j)} = r_1 + \dots + r_{\pi(j)} > \bar{H}(\mathbf{X}|\mathbf{Y}_{(j)})$ , which completes the achievability proof.

**Converse.** The converse is done by contradiction. If  $R_{(j)} < \bar{H}(\mathbf{X}|\mathbf{Y}_{(j)})$ , the decoder fails to recover the source comes from [22, Theorem 7.2.1]. The same result [22, Theorem 7.2.1] shows that if  $S < \max_{j \in \mathcal{J}} \bar{H}(\mathbf{X}|\mathbf{Y}_{(j)})$ , then with non zero probability, the source indexed by  $\bar{j} = \arg \max_{j \in \mathcal{J}} \bar{H}(\mathbf{X}|\mathbf{Y}_{(j)})$  can not be recovered at the decoder.

#### PROOF OF THEOREM 4

*Test-channel:* We consider the following test channels

$$\mathbf{U}^n = \mathbf{X}^n + \boldsymbol{\Psi}^n \quad (40)$$

$$\forall j \in \llbracket 1, J \rrbracket, \hat{\mathbf{X}}_{(j)}^n = A_{(j)} \mathbf{U}^n + B_{(j)} \mathbf{Y}_{(j)}^n \quad (41)$$

where  $\Psi^n \sim \mathcal{N}(0, \delta I_n)$  is independent of  $\mathbf{X}^n$ , and  $I_n$  is the identity matrix of dimension  $n \times n$ .  $A_{(j)}$  and  $B_{(j)}$  are  $n \times n$  matrices such that

$$A_{(j)} = \Sigma_{(X|j)}^n (\delta I_n + \Sigma_{(X|j)}^n)^{-1} \quad (42)$$

$$B_{(j)} = (I_n + \frac{1}{\delta} \Sigma_{(X|j)}^n)^{-1} \Sigma_{(X|j)}^n (\Sigma_{(j)}^n)^{-1}. \quad (43)$$

$\hat{\mathbf{X}}_{(j)}^n$  is the lossy reconstruction of  $\mathbf{X}^n$  at the  $j$ -th decoder, obtained from a Minimum Mean Square Error (MMSE) estimator. For all  $i \in \{1, \dots, n\}$  and  $j \in \llbracket 1, J \rrbracket$ , this test channel gives individual distortions

$$E \left[ (X_i - \hat{X}_{(j),i})^2 \right] = \left( \left( \frac{1}{\delta} I_n + (\Sigma_{(X|j)}^n)^{-1} \right)^{-1} \right)_{i,i} = \frac{\lambda_i^{(X|j)} \delta}{\lambda_i^{(X|j)} + \delta}. \quad (44)$$

*Random code generation:* Generate  $2^{nr_0}$  sequences  $\mathbf{U}^n$  at random according to (40). The distribution of  $\mathbf{U}^n$  does not depend on the possible side information  $\mathbf{Y}_{(j)}$ . Denote by  $\mathcal{C}$  the set of generated sequences  $\mathbf{u}^n$  and index them with  $s \in \{1, \dots, 2^{nr_0}\}$ . Assign each  $\mathbf{u}^n(s) \in \mathcal{C}$  to  $M$  incremental bins, following the same process as in the proof of Theorem 1.

In order to construct the incremental bins, consider the source reordering function  $\pi : \llbracket 1, J \rrbracket \rightarrow \llbracket 1, M \rrbracket$ ,  $j \mapsto \pi(j)$ . We denote  $m = \pi(j)$ , and the reordering function  $\pi$  is such that  $\bar{I}(\mathbf{Y}_{\pi^{-1}(m)}; \mathbf{U}) \leq \bar{I}(\mathbf{Y}_{\pi^{-1}(m-1)}; \mathbf{U})$ ,  $\forall m \in \llbracket 2, M \rrbracket$ , where  $\bar{I}(\cdot; \cdot)$  is the spectral mutual information defined in Section II-C. The number of incremental bins is defined by values  $r_m$  such that at the  $m$ -th level, there are  $2^{n(r_1 + \dots + r_m)}$  bins. This defines  $M$  mappings  $f_{(j)}(\mathbf{u}^n) = (i_1, \dots, i_{\pi(j)})$ ,  $j \in \{1, \dots, M\}$  where the  $(i_1, \dots, i_{\pi(j)})$  are the indices of the successive bins to which  $\mathbf{u}^n$  belongs.

*Encoding:* Given a sequence  $\mathbf{x}^n$ , find a sequence  $\mathbf{u}^n(s) \in \mathcal{C}$  such that  $(\mathbf{x}^n, \mathbf{u}^n(s)) \in T_{\varepsilon, n}^{(1)}(\mathbf{X}, \mathbf{U})$ , where

$$T_{\varepsilon, n}^{(1)}(\mathbf{X}, \mathbf{U}) = \left\{ (\mathbf{x}^n, \mathbf{u}^n) \text{ s.t. } \frac{1}{n} \log \frac{P_{\mathbf{U}^n | \mathbf{X}^n}(\mathbf{u}^n | \mathbf{x}^n)}{P_{\mathbf{U}^n}(\mathbf{u}^n)} < r_0 - \varepsilon \right\}. \quad (45)$$

The offline encoder then sends to the storage unit the index sequence  $(i_1, \dots, i_M)$  obtained for  $\mathbf{u}^n(s)$ . Upon request of the source  $X$  and side information  $Y_{(j)}$ , the online extractor sends to the user the index sequence  $(i_1, \dots, i_{\pi(j)})$  for  $\mathbf{u}^n(s)$ .

*Decoding:* Given the received index sequence  $(i_1, \dots, i_{\pi(j)})$  and the side information  $\mathbf{y}_{(j)}^n$ , declare  $\hat{\mathbf{u}}^n = \mathbf{u}^n(s)$  if there is a unique pair of sequences  $(\mathbf{y}_{(j)}^n, \mathbf{u}^n(s))$  such that  $f_{(j)}(\mathbf{u}^n(s)) = (i_1, \dots, i_{\pi(j)})$  and  $(\mathbf{y}_{(j)}^n, \mathbf{u}^n(s)) \in T_{\varepsilon, n}^{(2)}(\mathbf{Y}_{(j)}, \mathbf{U})$  where

$$T_{\varepsilon, n}^{(2)}(\mathbf{Y}_{(j)}, \mathbf{U}) = \left\{ (\mathbf{y}_{(j)}^n, \mathbf{u}^n) \text{ s.t. } \frac{1}{n} \log \frac{P_{\mathbf{U}^n | \mathbf{Y}_{(j)}^n}(\mathbf{u}^n | \mathbf{y}_{(j)}^n)}{P_{\mathbf{U}^n}(\mathbf{u}^n)} < \sum_{i=1}^j r_i - \varepsilon \right\}. \quad (46)$$

Then compute  $\hat{\mathbf{x}}_{(j)}^n$  from  $\hat{\mathbf{u}}^n$  and  $\mathbf{y}_{(j)}^n$  according to (41).

*Probability of error:* We define the error events:

$$E_{0,1} = \{\nexists s \text{ such that } (\mathbf{X}^n, \mathbf{u}^n(s)) \in T_{\varepsilon,n}^{(1)}(\mathbf{X}, \mathbf{U})\}$$

$$E_{0,2} = \{(\mathbf{X}^n, \mathbf{u}^n(s)) \in T_{\varepsilon,n}^{(1)}(\mathbf{X}, \mathbf{U}) \text{ but } (\mathbf{Y}_{(j)}^n, \mathbf{u}^n(s)) \notin T_{\varepsilon,n}^{(2)}(\mathbf{Y}_{(j)}, \mathbf{U})\}$$

$$E_j = \{\exists s' \neq s : f_{(j)}(\mathbf{u}^n(s')) = f_{(j)}(\mathbf{u}^n(s)) \text{ and } (\mathbf{Y}_{(j)}^n, \mathbf{u}^n(s')) \in T_{\varepsilon,n}^{(2)}(\mathbf{Y}_{(j)}, \mathbf{U})\}, \quad \forall j \in \mathcal{J}$$

By the same derivation as in the proof of Theorem 1, we show that  $\mathbb{P}(E_j) \rightarrow 0$  as  $n \rightarrow \infty$ ,  $\forall j \in \mathcal{J}$ . By the definitions of the spectral mutual information  $\bar{I}(\mathbf{X}; \mathbf{U})$ , see [22, Section 5.4], and of the set  $T_{\varepsilon,n}^{(1)}$  in (45) we show that if  $r_0 \geq \bar{I}(\mathbf{X}; \mathbf{U})$ , then  $\mathbb{P}(E_{0,1}) \rightarrow 0$  as  $n \rightarrow \infty$ . With the same arguments and from the definition of  $T_{\varepsilon,n}^{(2)}$  in (46), we show that  $\mathbb{P}(E_{0,2}) \rightarrow 0$  as  $n \rightarrow \infty$  if  $\sum_{i=1}^j r_i \geq \bar{I}(\mathbf{X}; \mathbf{U}) - \bar{I}(\mathbf{Y}_{(j)}; \mathbf{U})$ . Finally, from the two above rate conditions, the decoding error probability  $P_{error}^n = \mathbb{P}(E_{0,1} \cup E_{0,2} \cup_{j \in \mathcal{J}} E_j) \rightarrow 0$  as  $n \rightarrow \infty$ .

*Distortion and rate computation:* First, from the individual distortions (44) and from the error probability analysis, the overall distortion for the sequence  $\mathbf{X}^n$  can be calculated for all  $j \in \llbracket 1, J \rrbracket$  as

$$\frac{1}{n} \sum_{i=1}^n E[(X_i - \hat{X}_{(j),i})^2] \leq (1 - P_{error}^n) \frac{1}{n} \sum_{i=1}^n \frac{\lambda_i^{(X|j)} \delta}{\lambda_i^{(X|j)} + \delta} + P_{error}^n d_{max} \quad (47)$$

where  $d_{max}$  is a constant that represents the maximum possible distortion over a given component, and the expectation is calculated given that  $\mathbf{Y}_{(j)}$  is available at the decoder. Then, from the definition of the spectral mutual information in [22, Section 5.4] and from the fact that the Gaussian model parameters in Definition 2 are deterministic,  $\bar{I}(\mathbf{X}; \mathbf{U}) - \bar{I}(\mathbf{Y}_{(j)}; \mathbf{U}) = \lim_{n \rightarrow \infty} \sup \frac{1}{n} h(\mathbf{U}^n | \mathbf{X}^n) - \frac{1}{n} h(\mathbf{U}^n | \mathbf{Y}_{(j)}^n)$ . From [24], we can then express

$$\frac{1}{n} h(\mathbf{U}^n | \mathbf{X}^n) - \frac{1}{n} h(\mathbf{U}^n | \mathbf{Y}_{(j)}^n) = \frac{1}{2n} \log_2 \frac{\det(\Sigma_{(X|j)}^n + \Delta)}{\det(\Delta)} = \frac{1}{n} \sum_{i=1}^n \frac{1}{2} \log_2 \frac{\lambda_i^{(X|j)}}{d_{(j),i}}, \quad (48)$$

where  $\det(\cdot)$  is the determinant of the matrix in argument, and  $d_{(j),i} = \frac{\lambda_i^{(X|j)} \delta}{\lambda_i^{(X|j)} + \delta}$ . At the end, taking the limits when  $n \rightarrow \infty$  in (47) and (48), and expressing the rate-allocation optimization between the individual components  $X_i$  of  $\mathbf{X}^n$  gives the rate and distortion expressions in (14). It can be seen from (47) and (48) that the rate and the distortion are allocated component by component. This operation leads to the expression of  $S$  in (14) in which the maximum over the  $j \in \llbracket 1, J \rrbracket$  is taken component by component.

## PROOF OF THEOREM 5

*Random code generation* is the same as in the proof of Theorem 4, except that the sequences  $\mathbf{U}^n$  are generated according to a probability distribution  $P_{\mathbf{U}^n}$  that satisfies the conditions defined in Theorem 5. *Encoding* is the same as in the proof of Theorem 4. *Decoding* is also the same, except that  $\hat{\mathbf{x}}_{(j)}^n$  is computed as  $\hat{\mathbf{x}}_{(j)}^n = g_{(j)}(\mathbf{y}_{(j)}^n, \hat{\mathbf{u}}^n)$ , where  $g_{(j)}$  satisfies the conditions defined in Theorem 5.

The decoding error probability  $P_{error}^n$  is calculated as in Theorem 4. From this error probability, we can calculate the distortion as

$$\frac{1}{n} \sum_{i=1}^n E[(X_i - \hat{X}_{(j),i})^2] \leq (1 - P_{error}^n)D_{(j)} + P_{error}^n d_{max}. \quad (49)$$

According to the proof of Theorem 4, when the conditions (18) are satisfied, we have that  $P_{error}^n$  tends to 0 as  $n$  goes to  $\infty$ . This gives that  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E[(X_i - \hat{X}_{(j),i})^2] \leq D_{(j)}$ , which concludes the proof.

## ACKNOWLEDGEMENT

This work has received a French government support granted to the Cominlabs excellence laboratory and managed by the National Research Agency in the ‘‘Investing for the Future’’ program under reference ANR-10-LABX-07-01.

## REFERENCES

- [1] D. Slepian and J. Wolf, ‘‘Noiseless coding of correlated information sources,’’ *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, July 1973.
- [2] A. Wyner and J. Ziv, ‘‘The rate-distorsion function for source coding with side information at the decoder,’’ *IEEE Transactions on Inf. Th.*, vol. 22, no. 1, pp. 1–10, Jan 1976.
- [3] A. Argyriou, Ö. Alay, and P. Palantas, ‘‘Modeling the lossy transmission of correlated sources in multiple access fading channels,’’ *Physical Communication*, vol. 24, pp. 34–45, 2017.
- [4] M. Martalò and R. Raheli, ‘‘Models, statistics, and rates of binary correlated sources,’’ *Physical Communication*, vol. 19, pp. 70–80, 2016.
- [5] Z. Xiong, A. Liveris, and S. Cheng, ‘‘Distributed source coding for sensor networks,’’ *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 80–94, Sep 2004.
- [6] A. Sgarro, ‘‘Source coding with side information at several decoders,’’ *IEEE Transactions on Information Theory*, vol. 23, no. 2, pp. 179–182, 1977.
- [7] 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.
- [8] S. C. Draper, ‘‘Universal incremental slepian-wolf coding,’’ in *Allerton Conference on Communication, control and computing*, 2004, pp. 1757 – 1761.

- [9] C. Tian and S. N. Diggavi, "On multistage successive refinement for wyner-ziv source coding with degraded side informations," *IEEE Transactions on Information Theory*, vol. 53, no. 8, pp. 2946–2960, 2007.
- [10] R. Timo, T. J. Oechtering, and M. Wigger, "Source coding problems with conditionally less noisy side information," *IEEE Transactions on Information Theory*, vol. 60, no. 9, pp. 5516–5532, 2014.
- [11] S. Unal and A. B. Wagner, "Vector gaussian rate-distortion with variable side information," *IEEE Transactions on Information Theory*, vol. 63, no. 8, pp. 5162–5178, 2017.
- [12] W. H. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Transactions on Information Theory*, vol. 37, no. 2, pp. 269–275, 1991.
- [13] B. Rimoldi, "Successive refinement of information: Characterization of the achievable rates," *IEEE Transactions on Information Theory*, vol. 40, no. 1, pp. 253–259, 1994.
- [14] A. No, A. Ingber, and T. Weissman, "Strong successive refinability and rate-distortion-complexity tradeoff," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3618–3635, 2016.
- [15] L. Zhou, V. Y. Tan, and M. Motani, "Second-order and moderate deviations asymptotics for successive refinement," *IEEE Transactions on Information Theory*, vol. 63, no. 5, pp. 2896–2921, 2017.
- [16] H. Kimata, M. Kitahara, K. Kamikura, and Y. Yashima, "Free-viewpoint video communication using multi-view video coding," *NTT Technical Review*, vol. 2, no. 8, pp. 21–26, Aug. 2004.
- [17] S. Shimizu, M. Kitahara, H. Kimata, K. Kamikura, and Y. Yashima, "View scalable multiview video coding using 3-d warping with depth map," *IEEE Trans. on Circ. and Syst. for Video Technology*, vol. 17, no. 11, pp. 1485–1495, Nov. 2007.
- [18] G. Cheung, A. Ortega, and N. Cheung, "Interactive streaming of stored multiview video using redundant frame structures," *IEEE Transactions on Image Processing*, vol. 3, no. 3, pp. 744–761, Mar. 2011.
- [19] C.-Y. Chong and S. P. Kumar, "Sensor networks: evolution, opportunities, and challenges," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1247–1256, 2003.
- [20] S. C. Draper and E. Martinian, "Compound conditional source coding, Slepian-Wolf list decoding, and applications to media coding," in *IEEE International Symposium on Information Theory*, 2007.
- [21] A. Roumy and T. Maugey, "Universal lossless coding with random user access: the cost of interactivity," in *Proceedings IEEE International Conference on Image Processing*, Quebec, Canada, Sep. 2015.
- [22] T. Han, *Information-spectrum methods in information theory*. Springer, 2003.
- [23] K.-I. Iwata and J. Muramatsu, "An information-spectrum approach to rate-distortion function with side information," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 85, no. 6, pp. 1387–1395, 2002.
- [24] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed Karhunen–Loeve transform," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5177–5196, 2006.
- [25] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Transactions on Information Theory*, vol. 31, no. 6, pp. 727–734, Nov 1985.
- [26] E. Perron, S. N. Diggavi, and I. E. Telatar, "On the role of encoder side-information in source coding for multiple decoders," in *Information Theory, 2006 IEEE International Symposium on*. IEEE, 2006, pp. 331–335.
- [27] E. Perron, S. Diggavi, and E. Telatar, "Lossy source coding with gaussian or erased side-information," in *2009 IEEE International Symposium on Information Theory*, June 2009, pp. 1035–1039.



- [28] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2856–2867, 2014.
- [29] P. Hassanzadeh, A. Tulino, J. Llorca, and E. Erkip, "Cache-aided coded multicast for correlated sources," *arXiv preprint arXiv:1609.05831*, 2016.
- [30] M. Effros, "Distortion-rate bounds for fixed-and variable-rate multiresolution source codes," *IEEE Transactions on Information Theory*, vol. 45, no. 6, pp. 1887–1910, 1999.
- [31] T. Matsuta and T. Uyematsu, "Non-asymptotic bounds and a general formula for the rate-distortion region of the successive refinement problem," *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 101, no. 12, pp. 2110–2124, 2018.
- [32] K. Iwata, "An information-spectrum approach to rate-distortion function with side information," in *IEEE International Symposium on Information Theory, Proceedings.*, 2002, p. 156.
- [33] M. Mushkin and I. Bar-David, "Capacity and coding for the Gilbert-Elliot channels," *IEEE Transactions on Information Theory*, vol. 35, no. 6, pp. 1277–1290, 1989.
- [34] M. F. Huber, T. Bailey, H. Durrant-Whyte, and U. D. Hanebeck, "On entropy approximation for gaussian mixture random vectors," in *Multisensor Fusion and Integration for Intelligent Systems, 2008. MFI 2008. IEEE International Conference on.* IEEE, 2008, pp. 181–188.
- [35] T. Berger, "Rate distortion theory: A mathematical basis for data compression," 1971.