



HAL
open science

Excess rate for model selection in interactive compression using Belief-propagation decoding

Navid Mahmoudian Bidgoli, Thomas Maugey, Aline Roumy

► To cite this version:

Navid Mahmoudian Bidgoli, Thomas Maugey, Aline Roumy. Excess rate for model selection in interactive compression using Belief-propagation decoding. *Annals of Telecommunications - annales des télécommunications*, 2020, pp.1-18. 10.1007/s12243-020-00805-z . hal-02920417

HAL Id: hal-02920417

<https://inria.hal.science/hal-02920417>

Submitted on 24 Aug 2020

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.

Excess rate for model selection in interactive compression using Belief-propagation decoding

Navid Mahmoudian Bidgoli ·
Thomas Maugey · Aline Roumy

Received: date / Accepted: date

Abstract Interactive compression refers to the problem of compressing data while sending only the part requested by the user. In this context, the challenge is to perform the extraction in the compressed domain directly. Theoretical results exist, but they assume that the true distribution is known. In practical scenarios instead, the distribution must be estimated. In this paper, we first formulate the model selection problem for interactive compression and show that it requires to estimate the excess rate incurred by mismatched decoding. Then, we propose a new expression to evaluate the excess rate of mismatched decoding in a practical case of interest: when the decoder is the belief-propagation algorithm. We also propose a novel experimental setup to validate this closed-form formula. We show a good match for practical interactive compression schemes based on fixed-length Low-Density Parity-Check (LDPC) codes. This new formula is of great importance to perform model and rate selection.

Keywords source coding · interaction · model selection · mismatched decoding

1 Introduction

The way videos are consumed have considerably evolved in the last decade. With the arrival of new data formats and new streaming platforms, users have been enabled to interact with the content they watch, mostly by choosing part of the data they want to access. Compressing data so that users are able to extract only a part of it, called Interactive Compression/Coding (IC), requires new tools. More precisely, it has been proven that predictive coding,

This work was partially supported by the Cominlabs excellence laboratory with funding from the French National Research Agency (ANR-10-LABX-07-01) and by the Brittany Region (Grant No. ARED 9582 InterCOR).

The authors are with the Inria, Univ Rennes, CNRS, IRISA
E-mail: {navid.mahmoudian-bidgoli, thomas.maugey, aline.roumy}@inria.fr

widely used in standard video coders, can not be efficient in both storage and transmission [1]. Indeed, the challenge in IC is to deal with the uncertainty of the users' request upon compression. This can be formulated as a source coding problem, where a side information is available at the decoder, whereas the encoder has access to the set of possible side information [1, 2]. It differs from predictive coding, where the side information is available at both encoder and decoder. Therefore, the encoder in IC, relies on the statistics of the side information, and not on its realization, and belongs to the general class of model-based coding problems. Despite the efficiency of some proposed architectures to solve the IC problem [3–5], two key questions, related to IC (and thus model-based coding) remain: i) which statistical model should we select and send to the decoder for the data to be compressed? ii) at which encoding rate should we compress the data? These two questions require to determine the excess rate for mismatched decoding, *i.e.*, when an approximate model is used for decoding rather than the true model.

In mono source coding, *i.e.*, source coding without any Side Information (SI) source, the excess rate due to mismatched decoding is the KullbackLeibler (KL) divergence between the true distribution of the source and the one used to code/encode. This classical result holds for variable-length coding, and can be extended to several other coding schemes such as: fixed-length and predictive coding. However, all these generalizations don't tackle the case of IC. Indeed, IC is related to source coding with SI at the decoder. For this compression problem, evaluating the excess rate is still an open problem [6], since it is related to the mismatch capacity of a dual channel coding problem [7–9].

In this paper, we formulate the excess rate problem for different source coding schemes in Section 2 and propose to model this excess rate for IC using a closed form expression (relying on the KL divergence). Measuring the excess rate experimentally is not an easy task. For that purpose, we propose a code construction method in Section 3 that guarantees that the obtained rate is achievable while keeping low complexity. Finally, We validate the proposed model in Section 4 by comparing the rate using the true model and the one obtained experimentally under mismatch decoding.

Notation. Throughout this paper, a random scalar source is denoted by uppercase letters like X , and its realization is represented by the corresponding lowercase italic letter x . X^n denotes a random sequence of length n . Calligraphic letter \mathcal{X} represents the alphabet of random variable X . \oplus denotes the addition in the finite field of the operand.

2 Problem formulation

2.1 IC and excess rate for model selection

IC refers to the problem of compressing data, while allowing the user to access *any part* of the data *in the compressed domain*. Interactivity with a visual content occurs with several image modalities such as omnidirectional images [3], or texture maps of 3D models [10]. Common to both applications is that

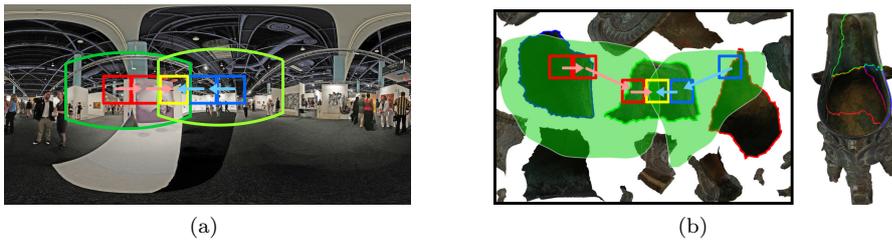


Fig. 1 Instances of users' requests. (a) Omnidirectional images. Two examples of requested viewports are shown in green. (b) 3D model and its texture map. Depending on user's navigation, users request different parts of the texture map (shown in green).

the image to be compressed is split into small blocks where the blocks are encoded/decoded one after the other. The user requests part of the data, and the server sends a compressed stream such that all blocks covering the requested part can be decoded. Fig. 1 shows two examples of such requests, depicted in green, for each image modality.

From the point of view of the encoder, each block must be compressed with the help of already decoded blocks such that, whatever the neighboring blocks available at the decoder, the transmitted compressed bitstream is sufficient to decode the block. For instance in Fig. 1, the current block, in yellow, must be encoded, whatever the request is, which means either if the already decoded blocks are the red ones or the blue ones.

More formally, let us denote x^n the current block, as the realization of a random vector X^n , see Fig. 2. Then, for a given request $k \in [1, L]$, the already decoded blocks produce an estimate of the current block, denoted y_k^n and called SI. This SI available at the decoder is not known in advance by the offline encoder, since it depends on the current request. However, this SI belongs to a set of SI sources, which is known to the encoder $\{y_k^n, k \in [1, L]\}$. Moreover, once a request is received, for each block, the SI available at the decoder is also known. Therefore, the online extractor can fetch from the compressed bitstream the necessary information. We refer to this source coding problem, introduced in [1, 2], as *IC*. IC differs from the compound coding problem [11, Sec. 3.1.9] by distinguishing the storage and transmission rates, denoted by R and R_k respectively. The optimal coding rates for independent and identically-distributed sources are:

$$R = \max_{i \in [1, L]} H(X|Y_i), \quad (1a)$$

$$R_k = H(X|Y_k), \quad (1b)$$

meaning that the source is encoded with respect to the worst-case correlation at rate R and extracted at rate R_k ($R_k \leq R$). This rate R_k is the same rate as if the SI was known in advance. Therefore, interactivity has no impact on the transmission rate, hence the advantage of the interactive problem formulation.

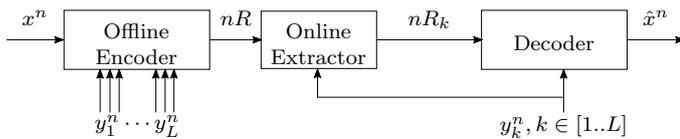


Fig. 2 IC scheme.

The optimal coding rates (1) have been derived under the assumption that the source statistics $\{P_{XY_k}, k \in [1, L]\}$ are perfectly known at both encoder and decoder. When the distributions are not known, a practical and optimal solution is the *two-stage code* [12, Chap. 6]. In the context of IC, it consists in 1) computing an estimate $\{Q_{Y_k}, k \in [1, L]\}$ of $\{P_{Y_k}, k \in [1, L]\}$ at both encoder and decoder (note that the realizations y_k^n are available at both encoder and decoder), 2) computing an estimate $\{Q_{X|Y_k}, k \in [1, L]\}$ of $\{P_{X|Y_k}, k \in [1, L]\}$ at the encoder and sending it to the decoder. Finally, data are encoded according to the estimated distributions.

These estimates have double impact on the compression performance. First, sending the distribution parameter adds an additional cost to the transmission rate, denoted by $\text{cost}_{Q_{X|Y_k}}$. Second, using an estimate rather than the true distribution increases the data compression rate by an additional factor called *excess rate* ΔR . The Minimum description length principle [13, 14] consists in choosing the estimate $Q_{X|Y_k}^*$, for a given estimate Q_{Y_k} , which minimizes the global cost:

$$Q_{X|Y_k}^* = \arg \min_{Q_{X|Y_k}} \text{cost}_{Q_{X|Y_k}} + H_P(X|Y_k) + \Delta R^{\text{IC}}(P_{XY_k}, Q_{X|Y_k}, Q_{Y_k}), \quad (2)$$

where $H_P(X|Y_k)$ stands for the conditional entropy computed with respect to the true distribution P_{XY_k} . Efficient distribution selection requires a closed form expression of the excess rate to avoid extensive simulations. The goal of this paper is to propose an analytical estimation of the excess rate $\Delta R^{\text{IC}}(P_{XY_k}, Q_{X|Y_k}, Q_{Y_k})$, for the IC scheme depicted in Fig. 2.

2.2 Restriction to a practical case of interest: linear codes with Belief Propagation decoding

IC shown in Fig. 2, is an extension of the SlepianWolf (SW) coding problem, depicted in Fig. 3(c). Indeed, as shown in [15, 2], the optimal code construction relies on a random binning argument for SW coding, and on an embedded random binning argument for IC. This is a consequence of the uncertainty at the encoder on the SI available at the decoder in Fig. 2, which is similar to the unavailability of the SI at the encoder in Fig. 3(c).

Unfortunately, the excess rate induced by using a wrong (approximate) distribution in SW coding is still an open problem [6]. This results from the duality between channel coding and SW coding [6], and from the fact that mismatched capacity is still an open problem [7–9]. Indeed, the excess rate in (2) is an information-theoretical measure and therefore includes an implicit

minimization over all possible decoding functions. Moreover, the excess rate depends very much on the decoder. For instance, in the case of the Binary Symmetric Channel (BSC) [16, p. 187], the excess rate is zero for maximum likelihood decoding but is non-zero for Belief Propagation (BP) decoding, as will be shown in Sec. 4. This is a consequence of the fact that maximum likelihood decoding is equivalent to minimizing the Hamming distance between codewords, which does not require the true distribution knowledge. Due to the prohibitive complexity of maximum likelihood decoding, BP decoding is extensively used. Therefore, we focus in the following on linear codes and BP decoding, which is of interest in practical scenarios.

Similar to SW coding, excess rate in IC remains an open and difficult problem. To overcome this issue and solve the model selection problem (2), we propose a closed formula for the excess rate in IC coding in one case of practical interest [3, 10], namely linear codes and BP decoding. For the sake of clarity, we denote $\Delta R_{\text{BP}}^{\text{IC}}(P_{XY_k}, Q_{X|Y_k}, Q_{Y_k})$ the excess rate, when BP decoding is applied. Then, we propose a novel code design method to show the accuracy of the conjectured formula.

2.3 Conjectured closed form formula and strategy for numerical evidence

To motivate our closed form estimate of the excess rate in IC, we first review various source coding problems. For the mono source compression problem (without having any SI), see Fig. 3(a), the excess rate induced by the use of a wrong distribution is derived in [16, Theorem 5.4.3]. In particular, when a single source X with distribution P_X is compressed with a variable-length code constructed with distribution Q_X , the excess rate is

$$\Delta R_{vl}^{mono} = D_{KL}(P_X || Q_X), \quad (3)$$

where subscript vl stands for variable-length code, $D_{KL}(P_X || Q_X)$ stands for the KL divergence between P_X and Q_X , and the variable-length code of rate R_{vl}^{mono} is defined by the following encoding and decoding functions:

$$f_{vl} : \mathcal{X}^n \rightarrow \{0, 1\}^*, \quad (4a)$$

$$g_{vl} : \{0, 1\}^* \rightarrow \mathcal{X}^n, \quad (4b)$$

$$R_{vl}^{mono} = \lim_{n \rightarrow +\infty} \mathbb{E}[l(f_{vl}(X^n))], \quad (4c)$$

where $l(u)$ is the length of the vector u and $\{0, 1\}^* = \{\emptyset, 0, 1, 00, 01, \dots\}$

In IC, fixed-length coding is of great interest because a practical implementation based on fixed-length coding for interactive image compression is proposed in [3]. For the mono source compression problem of Fig. 3(a), a fixed-length code of rate R_{fl}^{mono} is define as:

$$f_{fl} : \mathcal{X}^n \rightarrow \{1, 2, \dots, 2^{nR_{fl}^{mono}}\} = \mathcal{M}, \quad (5a)$$

$$g_{fl} : \mathcal{M} \rightarrow \mathcal{X}^n \cup \{\text{error}\}, \quad (5b)$$

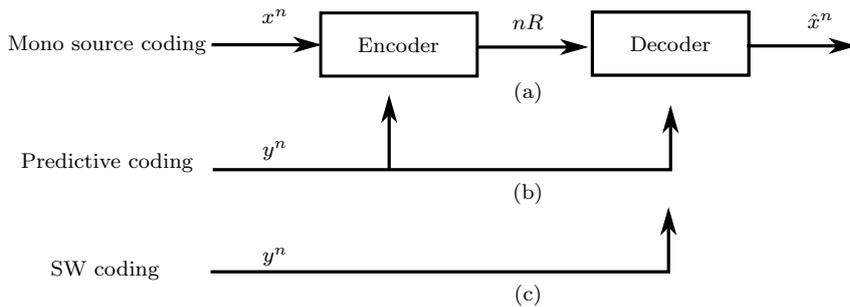


Fig. 3 Some source coding schemes based on the availability of SI. (a) Mono source coding. (b) Predictive coding. (c) SW coding.

By enlarging the typical set to take into account the uncertainty on the true distribution, one can show [17, Section 3] that the excess rate remains the same as the one for variable-length code (3):

$$\Delta R_{fl}^{mono} = D_{KL}(P_X || Q_X). \quad (6)$$

Both results can be extended to the predictive coding scheme, see Fig. 3(b), where a source X is compressed with SI Y available at both encoder and decoder. The excess rate is

$$\Delta R_{fl}^{pred} = \Delta R_{vl}^{pred} = D_{KL}(P_{XY} || Q_{X|Y} P_Y), \quad (7)$$

where P_{XY} stands for the true joint distribution, and $Q_{X|Y} P_Y$ is the decoding metric (the distribution used at decoder side).

In IC, the excess rate is an open problem (see Sec. 2.2), and we conjecture that, in the case of linear codes decoded with BP, the excess rate for a specific SI k can be well approximated by

$$\Delta R_{fl}^{IC} = D_{KL}(P_{XY_k} || Q_{X|Y_k} Q_{Y_k}). \quad (8)$$

This formula is of great interest as it allows to solve the model selection problem (2), without the need for extensive tests. This formula holds for any SI Y_k and, to simplify the notation, we drop index k in the remaining of the paper. To show the accuracy of this conjectured formula (8), we will first optimize the linear code ensemble \mathcal{C}_n , parameterized by the fixed-length n , by solving numerically

$$R^*(P_{XY}, Q_{XY}) = \min_{\mathcal{C}_n: P_e(\mathcal{C}_n, P_{XY}, Q_{XY}) \xrightarrow{n \rightarrow \infty} 0} R(\mathcal{C}_n), \quad (9)$$

where P_e stands for the probability of error under BP decoding. Then, we will compute the achievable excess rate with

$$\Delta R_{fl}^{IC} = R^*(P_{XY}, Q_{X|Y} P_Y) - R^*(P_{XY}, P_{XY}), \quad (10)$$

and compare it with the conjectured formula (8). Numerical precision in (9) is a key issue because in the context of IC, only small variations of the estimated distribution Q_{XY} around the true distribution P_{XY} are of practical interest. Indeed, in IC, the encoder has access to both realization vectors x^n and y^n and can provide an accurate estimate Q_{XY} .

A numerical solution to (9) can be obtained by introducing the mismatch distribution into classical optimization approaches such as the quantized Density Evolution (DE) algorithm [18], or the Mutual-Information based algorithm [19]. On one hand, the rate obtained by quantized DE is achievable but the algorithm is very sensitive to its initialization, such that the obtained rate is not necessarily the best one. On the other hand, the Mutual-Information based approach solves optimally a simplified problem such that the obtained rate is not necessarily achievable. These algorithms are therefore not sufficient to test our conjecture and, in the next section, we propose a novel alternating algorithm to solve (9), which insures that the optimal rate is indeed achievable, without a need to resort to multiple random initializations or genetic algorithms such as differential evolution [20].

3 Code design under rate optimization and the case of mismatched decoding

3.1 Duality with binary-input channel code optimization problem

In this section, we establish the duality between IC and channel coding and write the optimization problem (9) as a binary-input channel code design problem. First, IC is an extension of the SW coding problems. More precisely, in IC and for a given SI $Y = Y_k$, the optimal code construction is the same as the one for SW coding, see [15, 2]. Second, channel coding and SW coding are dual problems [21–24]. In particular, for linear codes with general distribution, the duality between channel coding and SW coding problem has been established in [22] and is shown in Fig. 4. This duality holds at the level of each individual linear codebook and implies that when encoding X^n with SI Y^n at the decoder with a general distribution P_{XY} , the decoding error probability of any single linear coset code is exactly the decoding error probability of its dual channel coding problem under maximum likelihood decoding or BP decoding.

Therefore, as in [22], the IC of source X with SI Y can be turned into a channel coding problem, with channel input U , channel output \mathbf{V} and channel transition distribution $P_{\mathbf{V}|U}$ where

$$\bar{V} = U \oplus X, \quad \bar{\bar{V}} = Y, \quad \mathbf{V} = (\bar{V}, \bar{\bar{V}}). \quad (11)$$

Here U has uniform distribution with the same alphabet as X , but is independent of (X, Y) . This duality has been first formulated in [21] for binary variables, where X is uniform and X, Y follow the distribution of a BSC. It has then been extended in [22, 23] to arbitrary variables by turning the X, Y variables into an equivalent symmetric channel with uniform input distribution.

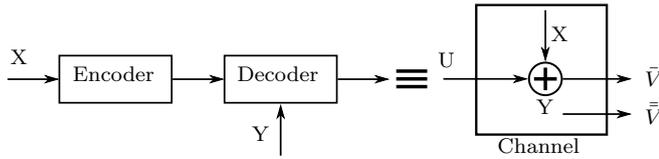


Fig. 4 Duality between SW coding and channel coding.

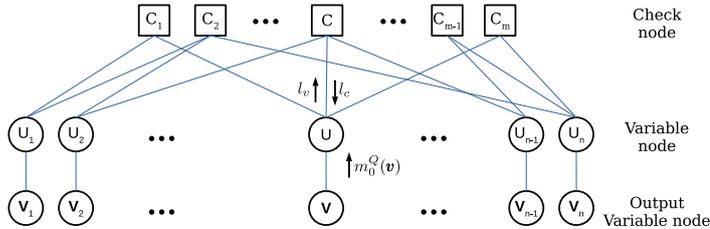


Fig. 5 LDPC bipartite graph.

Moreover, and without loss of generality, we restrict to the case of binary-input channels. Indeed, thanks to the chain rule, one can turn any finite input IC problem into a set of binary input IC subproblems, that can be solved separately, while still achieving optimality of the original problem [25]. Therefore, in the following, we assume that $\mathcal{X} = \mathcal{U} = \{0, 1\}$. As a consequence, source coding rate R and channel coding rate R_{ch} of the dual problems satisfy $R_{ch} = 1 - R$, and the optimization problem (9) becomes

$$C(P_{V|U}, Q_{V|U}) = \max_{\mathcal{C}_n: P_e(\mathcal{C}_n, P_{V|U}, Q_{V|U}) \xrightarrow{n \rightarrow \infty} 0} R_{ch}(\mathcal{C}_n), \quad (12)$$

where $Q_{V|U}$ stands for the dual mismatch decoding metric [6], \mathcal{C}_n represents the binary-input channel code with blocklength n , $R_{ch}(\mathcal{C}_n)$ is the channel rate, and $P_e(\mathcal{C}_n, P_{V|U}, Q_{V|U})$ is the decoding error probability.

3.2 LDPC codes, optimization techniques and their limitation

We now restrict our discussion to the case that is of most practical interest [3], namely the LDPC codes decoded with BP algorithms. Indeed, the linear codebook-level duality holds very generally, and in particular for LDPC codes with BP decoding algorithm and mismatched decoding metric [6].

An LDPC code [26] is a linear code that can be depicted as a bipartite code Fig. 5, where a Variable Node (VN) represents a channel input variable U , a Check Node (CN) represents a parity check equation, and an output VN is the channel output variable \mathbf{V} . BP decoding [27] provides an estimate of the input vectors u^n given the output vector \mathbf{v}^n by exchanging messages. For an edge between node C and node U , a message (l_c from C to U , or l_v from U to C) is an estimate of the input variable U . Then, BP decoding

consists in exchanging these messages by some update rules at both CN and VN [27, Chapter 2], to provide an estimate of the input vectors u^n given the output vector \mathbf{v}^n . In the case of mismatched decoding, the same update rules are applied. The only change is the initialization $m_0^Q(\mathbf{v})$, which now depends merely on the mismatched decoding metric:

$$m_0^Q(\mathbf{v}) = \ln \frac{Q_{V|U}(\mathbf{V} = \mathbf{v}|U = 0)}{Q_{V|U}(\mathbf{V} = \mathbf{v}|U = 1)} = \ln \frac{Q_{X|Y}(X = \bar{v}|Y = \bar{v})}{Q_{X|Y}(X = \bar{v} \oplus 1|Y = \bar{v})}, \quad (13)$$

where $\mathbf{v} = (\bar{v}, \bar{v})$.

The design parameters of an LDPC code are the connection degrees of the VN and CN from the edge perspective. More precisely, we denote the proportion of edges connected to VN and CN of degree i by λ_i and ρ_i respectively. The design parameters are summarized with the degree distribution polynomials $\lambda(x) = \sum_{i=2}^{d_v} \lambda_i x^{i-1}$ and $\rho(x) = \sum_{i=2}^{d_c} \rho_i x^{i-1}$. It follows that the rate of the LDPC code is equal to:

$$R_{ch}(\mathcal{C}_n(\rho, \lambda)) = 1 - \frac{\sum_{i=2}^{d_c} \rho_i / i}{\sum_{i=2}^{d_v} \lambda_i / i}. \quad (14)$$

LDPC code optimization is performed for a random code ensemble (all codes that satisfy the degree distribution constraints form an ensemble), and for channels that have a monotonic behavior with respect to a scalar p . The greater the p , the harder the channel is. An example of such a channel is the BSC, where the parameter is the crossover probability $p \leq 0.5$.

LDPC code optimization can be classified into two categories. In the first category, the rate of the code is fixed and the goal is to find the hardest channel (i.e. with the maximum threshold) that can be achieved with vanishing error probability [20]. This leads to

$$\max_{(\rho, \lambda)} p, \quad (15a)$$

$$\text{subject to } R(\mathcal{C}_n(\rho, \lambda)) = R_0, \quad (15b)$$

$$P_e(\mathcal{C}_n, P_{V|U}(p)) \xrightarrow[n \rightarrow \infty]{} 0, \quad (15c)$$

where $P_{V|U}(p)$ stands for the channel distribution $P_{V|U}$ with parameter p , and where the first constraint is linear in the design parameter (14). The second category of code design consists in fixing the distribution and optimizing the rate:

$$\max_{(\rho, \lambda)} R(\mathcal{C}_n(\rho, \lambda)), \quad (16a)$$

$$\text{subject to } P_e(\mathcal{C}_n, P_{V|U}(p)) \xrightarrow[n \rightarrow \infty]{} 0. \quad (16b)$$

In both problems (15) and (16), the difficulty lies in the evaluation of the asymptotic error probability as a function of the design parameters (ρ, λ) in (15c) and (16b). The accurate evaluation is called DE [20, 27] and consists in

tracking the evolution of the densities of the messages involved in the iterative BP algorithm. When the number of iterations of BP goes to infinity, the messages have a continuous density, and (15c) becomes an infinite dimensional constraint. To solve (15), one can use a genetic algorithm such as differential evolution at the price of a very high complexity, since at each iteration (of the global optimizer), several DE are performed (where one DE consists in a great number of iterations up to the convergence of the BP algorithm). A way to simplify the infinite dimensional constraint (16b) is to quantize the densities and add a slow-convergence constraint [18, Constraint 2 in Sec. III]. This leads to multiple linear constraints [18], one per BP iteration. Therefore, the solution is an iterative algorithm, where each iteration solves a linear programming problem. Both approaches are quite accurate, but suffer a very high complexity and the need for an accurate first estimation.

A faster solution consists in replacing the whole density by a scalar parameter [27]. A popular approach, called Extrinsic Information Transfer Chart (EXIT) chart, consists in computing a mutual information to approximate the density, and (15c) and (16b) become a one dimensional equation linear in the design parameters [19] and the whole problem is a linear programming problem. The price to pay for this simplification is a lack of accuracy.

In the following, we propose a novel algorithm for solving problem (16) since our goal is to perform rate optimization (12). We first propose a way to solve the inaccuracy problem of the EXIT chart analysis and combine the EXIT chart based optimization (Section 3.3.1) with a novel channel hardening approach (Section 3.3.2) to provide an efficient initialization to start the optimization problem proposed in [18]. This way, the obtained rate is guaranteed to be achievable since it is obtained with quantize DE algorithm [18]. Second, the accurate proposed initialization insures that the final rate is closed to the optimum one.

In the following, we detail our approach for the case of mismatch decoding, but note that this can also efficiently apply for the case without mismatch. The only difference between the mismatched and classical approaches lies in the initialization. In the case of mismatched decoding, the distribution of the initial message $m_0^Q(\mathbf{v})$ (13) is

$$P_0 = \sum_{\mathbf{v} \in \mathcal{X} \times \mathcal{Y}}^{|\mathcal{X}| \cdot |\mathcal{Y}|} P_{X,Y}(X = \bar{v}, Y = \bar{\bar{v}}) \cdot \delta_{m_0^Q(\mathbf{v})}, \quad (17)$$

where $\mathbf{v} = (\bar{v}, \bar{\bar{v}})$ and δ_t is a Dirac delta function at point t .

3.3 Code design through rate optimization with rate-achievability guarantee

3.3.1 Rough solution with consistency or channel decomposition

An EXIT chart [28], is a technique which tracks the mutual information between the transmitted bit U and the soft Log-likelihood ratio (LLR) messages

L_v corresponding to this bit. This mutual information can be computed as [28][29, chapter 9.6]:

$$I(U; L_v) = 1 - \sum_{u=\pm 1} \frac{1}{2} \int_{-\infty}^{\infty} p_{L_v|U}(l_v|u) \cdot \log_2 \left(\frac{p_{L_v|U}(l_v|U = -1) + p_{L_v|U}(l_v|U = 1)}{p_{L_v|U}(l_v|u)} \right) dl_v. \quad (18)$$

The technique relies on the Gaussian approximation of messages exchanged between VN and CN processors, in which the output extrinsic information of one processor is the input *a priori* information for the other one and vice versa. Here the same notation of [28] is used, i.e., the mutual information between the extrinsic (a priori) information coming out of (into) a processor and the code bit associated with that processor is denoted by I_E (I_A). In a binary input additive white Gaussian noise (BIAWGN) channel, the input LLR messages have also Gaussian distribution and are consistent. Assuming consistency condition is also valid for other Gaussian messages exchanged between VN and CN, (18) for VN will be simplified to:

$$I(U; L_v) = J(\sigma) = 1 - \int_{-\infty}^{+\infty} \frac{e^{-\frac{(l_v - \frac{\sigma^2}{2})^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} \log_2(1 + e^{-l_v}) dl_v.$$

Let σ_A and σ_E denote the standard deviations of the consistent-Gaussian distribution of the messages coming into and out of a VN of degree d_v , respectively. The extrinsic and a priori mutual information of the VN is equal to:

$$I_{AC} = J(\sigma_A) \text{ and } I_{EV} = J\left(\sqrt{\sigma_{in}^2 + \sigma_0^2}\right), \quad (19)$$

where σ_0^2 represents the standard deviation of the initial LLR message distribution and $\sigma_{in}^2 = (d_v - 1)\sigma_A^2$ is the variance of the Gaussian input messages coming from the neighboring CN to VN of degree d_v .

For non-Gaussian SW coding problems which have discrete LLR distribution as in (17), the output distribution of VN update with consistent-Gaussian input messages is indeed a mixture of Gaussian as:

$$\begin{aligned} p_{L_v|U}(l_v|U = +1) &= P_0 \otimes \mathcal{N}\left(\frac{\sigma_{in}^2}{2}, \sigma_{in}^2\right) \\ &= \sum_{\mathbf{v} \in \mathcal{X} \times \mathcal{Y}}^{|\mathcal{X}| \cdot |\mathcal{Y}|} P_{X,Y}(X = \bar{v}, Y = \bar{\bar{v}}) \cdot \mathcal{N}\left(\frac{\sigma_{in}^2}{2} - m_0^Q(\mathbf{v}), \sigma_{in}^2\right), \end{aligned}$$

where \otimes represents the convolution operation. Therefore, when we have discrete output, the messages are no more consistent (they are still symmetric) and thus (19) is no longer valid for I_{EV} . We propose two solutions to compute I_{EV} , first by assuming that the distribution is still consistent, and second by decomposing the output distribution to BSC [30].

1- *Assuming consistency assumption:* Inspired by [19, eq. 35] one can assume the distribution is still consistent, let $J(\sigma, t)$ be:

$$J(\sigma, t) = 1 - \int_{-\infty}^{\infty} \frac{e^{-\frac{(l_v - \frac{\sigma^2}{2} - t)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} \log_2(1 + e^{-l_v}) dl_v.$$

Then, I_{EV} becomes

$$I_{EV} = \sum_{d=2}^{d_v} \lambda_d \cdot \sum_{\mathbf{v} \in \mathcal{X} \times \mathcal{Y}}^{|\mathcal{X}| \cdot |\mathcal{Y}|} P_{XY}(\mathbf{v}) J\left(\sqrt{(d-1)[J^{-1}(I_{AV})]^2}, m_0^Q(\mathbf{v})\right). \quad (20)$$

We approximate this mutual information with Gauss-Hermite quadrature.

2- *Decomposition into BSCs:* A binary-input symmetric memoryless channel can be separated into sub-channels which are BSCs [30]. Since $p_{L_v|U}(l_v|u)$ is symmetric, using a quantizer we can decompose it to $(W+1)$ BSC with intervals $0 < \zeta_0 < \zeta_1 < \dots < \zeta_W < +\infty$. We have

$$\begin{aligned} P_w &= \int_{\zeta_{w-1}}^{\zeta_w} p_{L_v|U}(l_v|U=+1) dl_v + \int_{-\zeta_w}^{-\zeta_{w-1}} p_{L_v|U}(l_v|U=+1) dl_v, \\ \epsilon_w &= \frac{1}{P_w} \int_{-\zeta_w}^{-\zeta_{w-1}} p_{L_v|U}(l_v|U=+1) dl_v, \\ I_w &= 1 - h_b(\epsilon_w), \end{aligned}$$

where P_w is the probability of sub-channel w , ϵ_w is its corresponding cross-over probability of the BSC sub-channel and I_w is the corresponding mutual information of the sub-channel. $h_b(\epsilon)$ denotes the binary entropy function. Without loss of generality, the sub-channel $w=0$ can be interpreted as a BSC with crossover probability 0.5 [30]. The mutual information of I_{EV} can be obtained by taking the expectation of the mutual information of the subchannels:

$$I_{EV} = \sum_{d=2}^{d_v} \lambda_i \mathbb{E}_w\{I_w^d\} = \sum_{d=2}^{d_v} \lambda_i \sum_{w=0}^W P_w^{(d)} I_w^{(d)}. \quad (21)$$

As it can be seen in (20) or (21), the EXIT function for the VNs depends on the initial messages and I_{AV} . To obtain the CN EXIT function, the approximate duality property is exploited [31,32]. This states that a degree- d single parity-check code and that of a degree- d repetition code are related as

$$I_{E,SPC}(d, I_A) = 1 - I_{E,REP}(d, 1 - I_A).$$

As can be seen, the EXIT function for CN only depends on I_{AC} .

EXIT curves can be used to design LDPC codes [32]. We will consider only check-regular LDPC codes. In order to converge to a vanishing probability of error for decoding, the EXIT chart of the VN has to lie above the inverse of the EXIT chart for the CN. The target in the code optimization is to maximize

the rate (14) while considering a fixed CN degree distribution and a fixed P_0 . Therefore, we obtain the optimization method as

$$\begin{cases} \text{maximize} & \sum_{i=2}^{d_v} \lambda_i/i \\ \text{subject to} & I_{EC}^{-1}(I_{AC}) < I_{EV}(I_{AV}, P_0) \\ \text{and to} & \sum_{i=2}^{d_v} \lambda_i = 1, \lambda_i \geq 0, i = 2, 3, \dots, d_v \end{cases}, \quad (22)$$

where $I_{AV} = I_{EC}$ and the optimization is solved by discretization of $I_{AV} \in (0, 1)$ and applying linear programming. Using this optimizer we can have a first rough estimate of the code parameters which satisfies (12) approximately.

3.3.2 Refined solution for a fake harder channel

In general, the EXIT chart optimization is optimistic in the sense that the optimized degree distribution might not have vanishing probability of error [19]. To make sure that the optimized degree distribution is valid, after the EXIT chart optimization we evaluate the decoding error probability of its output degree distribution with DE [18]. If the optimized degree distribution coming out of the EXIT chart optimization does not have a zero decoding error probability, we optimize the degree distribution for a more difficult channel.

Keeping in mind that for the initial messages, the position of the LLRs in the space and their associated probabilities will be determined by the decoding metric Q_{XY} and the source distribution P_{XY} , respectively, the term difficult channel means that with the same decoding metric we assume that the probability of receiving the wrong symbol is higher than the true one. This affects the probabilities of the initial messages and can be achieved by decreasing the probability of positive LLRs and increasing the negative LLRs probabilities in P_0 of (17) by a gap ϵ :

$$P_{XY}(X = \bar{v}, Y = \bar{v}) = \begin{cases} P_{XY}(X = \bar{v}, Y = \bar{v}) - \epsilon \cdot P_Y(Y = \bar{v}) & \text{if } m_0^Q(\mathbf{v}) > 0 \\ P_{XY}(X = \bar{v}, Y = \bar{v}) + \epsilon \cdot P_Y(Y = \bar{v}) & \text{if } m_0^Q(\mathbf{v}) < 0 \end{cases}. \quad (23)$$

This way we are sure that the decoding metric is fixed and only its probability distribution is changing. For a binary-input binary-output source, this corresponds to increasing the crossover probabilities ($P_{X|Y}(X = 1|Y = 0)$ and $P_{X|Y}(X = 0|Y = 1)$) by ϵ .

We increase the difficulty of the channel until the optimized degree can converge to zero decoding error by DE test. In the EXIT curve computations, we apply both VN EXIT curves of (20) and (21) and pick the one which provides the higher channel rate (with vanishing error probability).

3.3.3 Final solution

With the EXIT chart optimization we are able to have a rough estimation of the LDPC degree distributions, but this estimation is not optimal. Indeed, the assumption of having Gaussian density for the messages used in EXIT chart, was to simplify and stabilize the numerical computation of the evolution of the message densities. In order to relax this Gaussian assumption, we tune this

rough estimation by another optimization, proposed by Chung et al. [18], which uses discretized DE in its “inner” loop and takes as input the rough estimation of degree distribution and tries to optimize it iteratively. The details of the algorithm can be found in [18]. For the case of SW coding, again the initial message density of (17) is used during the calculation of message densities.

To design LDPC codes using discretized DE, again the channel rate is maximized using linear programming which results in optimizing the $\lambda(x)$ for a fixed $\rho(x)$. For that, unlike EXIT chart optimization, the algorithm must be initiated with a $\lambda(x)$ that results the channel code rate lower than the desired rate. Then the optimization is run to update the degree distributions to $\lambda'(x)$ and increase the channel code rate maintaining the following constrains:

$$\left\{ \begin{array}{l} \text{maximize } \sum_{i=2}^{d_v} \lambda_i/i \\ \text{subject to } \sum_{i=2}^{d_v} \lambda_i = 1, \lambda_i \geq 0, i = 2, 3, \dots, d_v \\ \text{and to } \lambda'(x) \text{ is not significantly different from } \lambda(x) \\ \text{and to } \lambda'(x) \text{ produces smaller probability of error} \end{array} \right. \quad (24)$$

We recursively tune the output of (24) until the output $\lambda(x)$ does not change significantly. The details of the algorithm is given in Algorithm 1. Here $EXIT_opt(m_0^Q, P, d_c)$ is the EXIT chart optimization function discussed in Section 3.3.1 which takes as input the initial messages and their associated probabilities for a fixed check-regular LDPC codes with $\rho(x) = x^{d_c-1}$. The output of this function is the optimized variable degree distribution λ . The function $DE(m_0^Q, P_0, (\lambda, \rho))$ analyzes the performance of LDPC code ensemble for a pair of degree distribution (λ, ρ) using DE algorithm and the output of this function is the probability of error p_e . The function $QDE_opt(m_0^Q, P_0, \lambda_{init}, d_c)$ optimize the degree distribution starting from λ_{init} for fixed m_0^Q, P, d_c using quantized/discretized DE discussed here.

4 Experimental results

We first show the significance of the model selection problem encountered in IC. Indeed, Table 1 shows the KL divergence for typical values of distributions observed in [3]. More precisely, for IC of images shown in Fig. 1, if the block to be encoded is of size 8×8 pixels, for each SI provided for the block (predictions generated from neighboring blocks), the cost to encode a distribution with 1 bit is $1/64 = 0.156$ bit per pixel, which is on the same order of magnitude as the values in Table 1. Therefore, as explained in (2), there is a trade-off between the cost to encode a distribution and its corresponding excess rate caused by using the approximate distribution.

Second, we show that the code optimization algorithm proposed in Section 3 allows to get better codes than state of the art methods and is therefore an accurate method to evaluate the best possible compression rate, and thus the excess rate. For that, we compare the output of our code optimization with the code optimization of [33]. For the distribution defined in [33, section IV-A], the best compression rate in [33] is 0.6 while our method achieves 0.589 (the lower the rate, the better the compression) with the same maximum VN

Algorithm 1 Algorithm for LDPC code design**Input:**

P_{XY} as the true joint distribution and $Q_{XY} = Q_{X|Y}.P_Y$ as the decoding metric.

Degree d_c for the check-regular nodes, i.e. $\rho(x) = x^{d_c-1}$

Maximum degree of the variable nodes d_v^{max}

The maximum value of permitted gap ϵ to ϵ_{max} and increase step of the gap to Δ_ϵ

Output: λ^{opt}

1: Initialize $\epsilon \leftarrow 0$

2: **while** $\epsilon \leq \epsilon_{max}$ **do**

3: $m_0^Q \leftarrow$ using Q_{XY} in (13)

4: Assign probabilities P_0 to m_0^Q using (23)

5: $\lambda^{opt} \leftarrow EXIT_{opt}(m_0^Q, P_0, d_c)$. Note: apply both consistency assumption & decomposition into BSCs methods and choose the one that provides higher channel rate.

6: $p_e \leftarrow DE(m_0^Q, P_0, (\lambda^{opt}, \rho))$

7: **if** $p_e \approx 0$ **then**

8: break

9: **end if**

10: $\epsilon \leftarrow \epsilon + \Delta_\epsilon$

11: **end while**

12: **repeat**

13: $\lambda^{opt} \leftarrow QDE_{opt}(m_0^Q, P_0, \lambda^{opt}, d_c)$

14: **until** λ^{opt} or the rate resulting from (λ^{opt}, ρ) converge

Table 1 The KL divergence between the true joint distribution and the approximate distribution when crossover probability $P_{X|Y}(0|1) = P_{X|Y}(1|0) = p$ is quantized using different number of bits.

| # bits | $p = 0.05$ | $p = 0.05$ | $p = 0.2$ | $p = 0.2$ | $p = 0.3$ | $p = 0.3$ |
|--------|----------------|----------------|----------------|----------------|----------------|----------------|
| | $P_Y(0) = 0.1$ | $P_Y(0) = 0.5$ | $P_Y(0) = 0.1$ | $P_Y(0) = 0.5$ | $P_Y(0) = 0.1$ | $P_Y(0) = 0.5$ |
| 1 | 0.208 | 0.208 | 0.01 | 0.01 | 0.009 | 0.009 |
| 2 | 0.047 | 0.047 | 0.032 | 0.032 | 0.018 | 0.018 |
| 3 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 |
| 4 | 0.007 | 0.007 | 0.002 | 0.002 | 0.001 | 0.001 |

degree and check regular node degree. The degree distributions of our code are:

$$\lambda(x) = 0.238796x + 0.210703x^2 + 0.117978x^5 + 0.125822x^6 + 0.306701x^{19}, \quad \text{and} \quad \rho(x) = x^6.$$

Finally, we test our conjecture in (8) by computing the excess rate using LDPC codes with BP decoding, based on Algorithm 1, and comparing it with KL divergence. We first optimize the code without mismatch and then with mismatch to find the rate. Tests have been carried out for binary variables under different configurations of joint distribution $P_{X,Y}$ as the true distribution, from symmetric or asymmetric $P_{X|Y}$ to uniform or non-uniform marginals P_Y . For each joint distribution $P_{X,Y}$, 30 decoding metrics $Q_{XY} = Q_{X|Y}.P_Y$ are considered as approximate distributions in which $Q_{X|Y}$ is a BSC with parameter q ranging from 0.009 to 0.26 with step 0.0087. Indeed, we assume that an accurate estimate of the SI distribution P_Y is available at both encoder and decoder at no cost, $Q_Y = P_Y$, since in IC the SI is available at encoder and decoder.

The maximum VN degree is set to 100 in all experiments, and 11 bit is used to quantize DE. Let $r = \log |\mathcal{X}| - r_{ch} = 1 - r_{ch}$ denotes the compression rate in which r_{ch} is the rate of optimized channel code. At high channel code rates (low source code rates), with a fixed maximum VN degree, LDPC codes with higher check-regular degrees perform better. Therefore, for designing code with respect to the true distribution P_{XY} we increase the degree of the check nodes until the optimized code produced by the code optimizer framework (Algorithm 1) has less than 0.01 difference in rate compared to the theoretical limit $H_P(X|Y)$ given in (1b), i.e., conditional entropy between the source and SI given the true distribution P_{XY} . We denote this CN degree with D_c^* and the corresponding compression rate with r_P . For each approximate decoding metric, we optimize the code for all CN degrees which are less than or equal to D_c^* and pick the one which has a lower compression rate and denote it by r_Q . Finally, we compare the excess rate $\Delta R_{BP}^{IC} = r_Q - r_P$ with our conjecture formulated in (8), i.e., $D_{KL}(P_{XY}||Q_{X|Y}P_Y)$.

Results are shown in Fig. 6. We can see that in almost all cases of interest (see Table 1), the KL is a good measure to estimate the excess rate. This means concretely that one can use the KL metric for estimating the excess rate ΔR_{BP}^{IC} in (2) instead of computing it in practice, reducing at the same time the computational complexity.

5 Conclusion

In IC, selecting the model to use for decoding and thus to transmit to the decoder is an important task. In particular, it requires to evaluate the impact of an approximate model on the compression rate. We characterized this excess rate in terms of the mismatched capacity of the dual channel, for model based source coding, which is an open problem. For these coding schemes, we showed experimentally that the KL divergence between the true and the approximate model is a proper estimate of the excess rate experienced by mismatched BP decoding of LDPC codes. This was evidenced by a new algorithm that allows to design LDPC codes decoded with BP according to any decoding metric.

References

1. Roumy, A., Maugey, T.: Universal lossless coding with random user access: The cost of interactivity. In: 2015 IEEE International Conference on Image Processing (ICIP), pp. 1870–1874 (2015)
2. Dupraz, E., Roumy, A., Maugey, T., Kieffer, M.: Rate-storage regions for extractable source coding with side information. *Physical Communication* **37**, 100845 (2019)
3. Mahmoudian Bidgoli, N., Maugey, T., Roumy, A.: Compression de contenus 360 et transmission adaptée à la navigation de l'utilisateur. In: Actes de la 27ème édition du colloque GretsI (2019)
4. Maugey, T., Roumy, A., Dupraz, E., Kieffer, M.: Incremental coding for extractable compression in the context of massive random access. *IEEE Transactions on Signal and Information Processing over Networks* **6**, 251–260 (2020)
5. Draper, S.C., Martinian, E.: Compound conditional source coding, slepian-wolf list decoding, and applications to media coding. In: 2007 IEEE International Symposium on Information Theory, pp. 1511–1515 (2007)

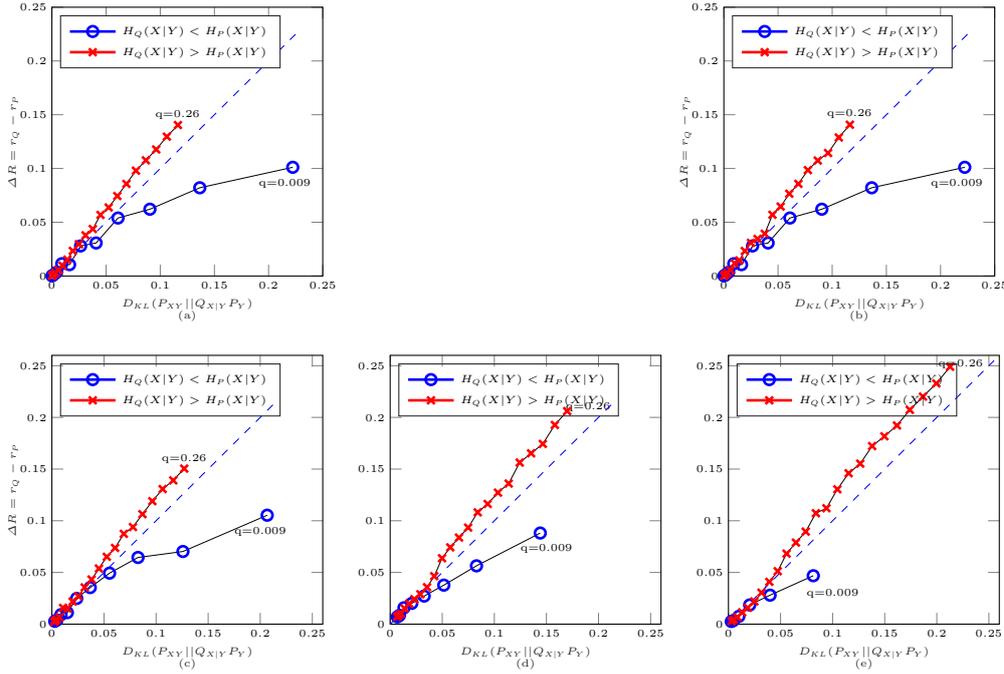


Fig. 6 Excess rate under mismatched decoding metric Q_{XY} vs $D_{KL}(P_{XY} || Q_{X|Y} P_Y)$. (a-b) symmetric $P_{X|Y}$, $P_{X|Y}(0|1) = 0.1$. (c-e) asymmetric $P_{X|Y}$, $P_{X|Y}(0|1) = 0.05$, $P_{X|Y}(1|0) = 0.1$. $H_P(X|Y)$ and $H_Q(X|Y)$ represent conditional entropy of source X conditioned on SI Y using P_{XY} and Q_{XY} respectively. (a) $P_Y(0) = P_Y(1) = 0.5$. (b) $P_Y(0) = 0.1$. (c) $P_Y(0) = 0.1$. (d) $P_Y(0) = P_Y(1) = 0.5$. (e) $P_Y(0) = 0.9$.

6. Chen, J., He, D., Jagmohan, A.: On the duality between Slepian–Wolf coding and channel coding under mismatched decoding. *IEEE Transactions on Information Theory* **55**(9), 4006–4018 (2009)
7. Merhav, N., Kaplan, G., Lapidoth, A., Shamai Shitz, S.: On information rates for mismatched decoders. *IEEE Transactions on Information Theory* **40**(6), 1953–1967 (1994)
8. Ganti, A., Lapidoth, A., Telatar, I.E.: Mismatched decoding revisited: general alphabets, channels with memory, and the wide-band limit. *IEEE Transactions on Information Theory* **46**(7), 2315–2328 (2000)
9. Scarlett, J., Martinez, A., i. Fabregas, A.G.: Mismatched decoding: Error exponents, second-order rates and saddlepoint approximations. *IEEE Transactions on Information Theory* **60**(5), 2647–2666 (2014)
10. Mahmoudian Bidgoli, N., Maugey, T., Roumy, A., Nasiri, F., Payan, F.: A geometry-aware compression of 3d mesh texture with random access. In: 2019 Picture Coding Symposium (PCS), pp. 1–5 (2019)
11. Han, T.S., Kobayashi, K.: Mathematics of information and coding, *Translations of mathematical monographs*, vol. 203. Providence, R.I. : American Mathematical Society (2002)
12. Csiszar, I., Shields, P.C.: Information Theory and Statistics: A Tutorial. *Foundations and Trends in Communications and Information Theory* **1**(4), 417–528 (2004). Publisher: Now Publishers, Inc.
13. Rissanen, J.: Modeling by shortest data description. *Automatica* **14**(5), 465 – 471 (1978)
14. Gruenwald, P.D.: The Minimum Description Length Principle. MIT Press (2007)
15. Dupraz, E., Maugey, T., Roumy, A., Kieffer, M.: Transmission and Storage Rates for Sequential Massive Random Access. arXiv:1612.07163 [cs, math] (2017). URL <http://arxiv.org/abs/1612.07163>. ArXiv: 1612.07163

16. Cover, T.M., Thomas, J.A.: Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing). Wiley-Interscience, USA (2006)
17. Bidgoli, N.M., Maugey, T., Roumy, A.: Correlation model selection for interactive video communication. In: 2017 IEEE International Conference on Image Processing (ICIP), pp. 2184–2188 (2017)
18. Sae-Young Chung, Forney, G.D., Richardson, T.J., Urbanke, R.: On the design of low-density parity-check codes within 0.0045 db of the shannon limit. *IEEE Communications Letters* **5**(2), 58–60 (2001)
19. Roumy, A., Guemghar, S., Caire, G., Verdu, S.: Design methods for irregular repeat-accumulate codes. *IEEE Transactions on Information Theory* **50**(8), 1711–1727 (2004)
20. Richardson, T.J., Shokrollahi, M.A., Urbanke, R.L.: Design of capacity-approaching irregular low-density parity-check codes. *IEEE Transactions on Information Theory* **47**(2), 619–637 (2001)
21. Wyner, A.: Recent results in the shannon theory. *IEEE Transactions on Information Theory* **20**(1), 2–10 (1974)
22. Chen, J., He, D., Yang, E.: On the codebook-level duality between slepian-wolf coding and channel coding. In: 2007 Information Theory and Applications Workshop, pp. 84–93 (2007)
23. Wang, L., Kim, Y.: Linear code duality between channel coding and slepian-wolf coding. In: 2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp. 147–152 (2015)
24. Liveris, A.D., Zixiang Xiong, Georghiadis, C.N.: Compression of binary sources with side information at the decoder using ldpc codes. *IEEE Communications Letters* **6**(10), 440–442 (2002)
25. Westerlaken, R.P., Borchert, S., Gunnewiek, R.K., Lagendijk, R.L.: Analyzing symbol and bit plane-based ldpc in distributed video coding. In: 2007 IEEE International Conference on Image Processing, vol. 2, pp. II – 17–II – 20 (2007)
26. Gallager, R.: Low-density parity-check codes. *IRE Transactions on Information Theory* **8**(1), 21–28 (1962)
27. Richardson, T., Urbanke, R.: Modern coding theory. Cambridge Univ Press (2008)
28. ten Brink, S.: Convergence behavior of iteratively decoded parallel concatenated codes. *IEEE Transactions on Communications* **49**(10), 1727–1737 (2001)
29. Ryan, W., Lin, S.: Channel Codes: Classical and Modern. Cambridge University Press (2009)
30. Land, I., Huber, J.: Information combining. *Found. Trends Commun. Inf. Theory* **3**(3), 227–330 (2006)
31. Ashikhmin, A., Kramer, G., ten Brink, S.: Extrinsic information transfer functions: model and erasure channel properties. *IEEE Transactions on Information Theory* **50**(11), 2657–2673 (2004)
32. ten Brink, S., Kramer, G., Ashikhmin, A.: Design of low-density parity-check codes for modulation and detection. *IEEE Transactions on Communications* **52**(4), 670–678 (2004)
33. Sun, Z., Tian, C., Chen, J., Wong, K.M.: Ldpc code design for asynchronous slepian-wolf coding. *IEEE Transactions on Communications* **58**(2), 511–520 (2010)