Source-Channel Coding with Multiple Classes

Irina E. Bocharova[1], Albert Guillén i Fàbregas[234], Boris D. Kudryashov[1],
Alfonso Martinez[2], Adrià Tauste Campo[2], and Gonzalo Vazquez-Vilar[2]
[1]St. Petersburg Univ. of Information Technologies [2]Universitat Pompeu Fabra, [3]ICREA, [4]Univ. of Cambridge
Email: irina@eit.lth.se, guillen@ieee.org, boris@eit.lth.se {alfonso.martinez,atauste,gvazquez}@ieee.org

***Abstract*—We study a source-channel coding scheme in which source messages are assigned to classes and encoded using a channel code that depends on the class index. While each class code can be seen as a concatenation of a source code and a channel code, the overall performance improves on that of separate source-channel coding and approaches that of joint source-channel coding as the number of classes increases. The performance of this scheme is studied by means of random-coding bounds and validated by simulation of a low-complexity implementation using existing source and channel codes.**

I. Introduction

Jointly designed source-channel codes attain a lower error probability than separate source-channel coding [1]. This reduction in error probability has been quantified in terms of error exponents [1], [2]: joint coding has an error exponent at most twice that of separate codes [3]. This improvement justifies the interest in practical joint source-channel codes.

While a number of results treated this problem (see [4] and references therein), they do not realize the full gain over separate source-channel coding. In this paper, we analyze a scheme in which source messages are assigned to classes and encoded by different codes that depend on the class index. This scheme was shown to attain the joint source channel reliability function in those cases where it is known to be tight [5]. However, the analysis in [5] assumes that source and channel decoding are performed jointly at the receiver. In contrast, in this work we assume that the channel output is processed in parallel for each class using a maximum likelihood (ML) decoder. The decoded message is then selected from the outputs of the ML decoders based on a maximum a posteriori (MAP) criterion.

While this coding scheme fails to achieve the best performance of joint source-channel coding, it can be implemented with reduced complexity using existing source and channel codes. Moreover, this suboptimal decoding scheme improves on the error exponent of separate coding, and, as the number of classes increases, it approaches the error exponent of joint source-channel coding. The performance of this scheme is characterized through a random-coding analysis and validated by simulations of a reduced-complexity implementation.

This work has been funded in part by the European Research Council under ERC grant agreement 259663, by the European Union's 7th Framework Programme under grant agreement 303633 and by the Spanish Ministry of Economy and Competitiveness under grants RYC-2011-08150 and TEC2012-38800-C03-03.

II. System Model and Notation

We consider the transmission of a discrete memoryless source P over a memoryless channel W. The source message $\boldsymbol{v}$ has length k and is distributed according to $P^k(\boldsymbol{v}) \triangleq \prod_{j=1}^{k} P(v_j)$, where v_j is the j-th component. The channel input and output have length n and are respectively denoted by $\boldsymbol{x} = (x_1, \ldots, x_n)$, $\boldsymbol{y} = (y_1, \ldots, y_n)$. We define the channel law as $W^n(\boldsymbol{y}|\boldsymbol{x}) \triangleq \prod_{j=1}^{n} W(y_j|x_j)$.

A source-channel code is defined by an encoder and a decoder. The encoder maps the message $\boldsymbol{v}$ to a length-n codeword $\boldsymbol{x}(\boldsymbol{v})$. Based on the channel output $\boldsymbol{y}$, the decoder selects a message $\hat{\boldsymbol{v}}(\boldsymbol{y})$. Throughout the paper, random variables will be denoted by capital letters and the specific values they take are denoted by the corresponding lower case letters. The error probability of a source-channel code is thus given by

$$\epsilon_n = \Pr\{\boldsymbol{V} \neq \hat{\boldsymbol{V}}\}. \tag{1}$$

An exponent $E > 0$ is to said to be achievable if there exists a sequence of codes whose error probabilities ϵ_n satisfy

$$\epsilon_n \leq e^{-nE+o(n)}, \tag{2}$$

where $o(n)$ is a sequence such that $\lim_{n\to\infty} \frac{o(n)}{n} = 0$.

If we force the encoder (resp. the decoder) to be a concatenation of a source and channel encoder (resp. channel and source decoder), this model recovers separate source-channel coding. In this work, we propose a new scheme in which the source-message set is split into subsets, and different separate source-channel codes are used for each subset. At the receiver, each channel code is decoded in parallel, and the output is selected based on MAP criterion.

We define a partition $\mathcal{P}_k$ of the source-message set $\mathcal{V}^k$ into $N_k + 1$ disjoint subsets $\mathcal{A}_k^{(i)}$, $i = 0, 1, \ldots, N_k$. We shall refer to these subsets as *classes*. We assume that N_k grows at most subexponentially with k. All the messages belonging to the class $\mathcal{A}_k^{(0)}$ are encoded with the same codeword $\boldsymbol{x}(\boldsymbol{v}) = \boldsymbol{x}_0$ and are assumed to lead to a decoding error. The messages belonging to each remaining class, $\boldsymbol{v} \in \mathcal{A}_k^{(i)}$, are encoded with the codeword $\boldsymbol{x}(\boldsymbol{v}) \in \mathcal{C}_i$, $i = 1, \ldots, N_k$. $\mathcal{C}_i$ denotes a channel code of rate $R_i \triangleq \frac{1}{n} \log\bigl(\bigl|\mathcal{A}_k^{(i)}\bigr|\bigr)$, $i = 1, \ldots, N_k$.

At the receiver, we use a two-step decoder. Within each class $\mathcal{A}_k^{(i)}$, $i = 1, \ldots, N_k$, the decoder selects a message $\hat{\boldsymbol{v}}_i$ in $\mathcal{A}_k^{(i)}$ according to the ML criterion, i. e.,

$$\hat{\boldsymbol{v}}_i = \arg \max_{\boldsymbol{v} \in \mathcal{A}_k^{(i)}} \bigl\{ W^n\bigl(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{v})\bigr) \bigr\}. \tag{3}$$

For ease of presentation, we let the dependence of the decoder output on the channel output $\boldsymbol{y}$ be implicit. Next, the decoder selects the class index i with largest decoding metric

$$q(\boldsymbol{v},\boldsymbol{y}) \triangleq P^k(\boldsymbol{v}) W^n\big(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{v})\big), \tag{4}$$

that is,

$$\hat{\imath} = \arg\max_{i=1,\ldots,N_k} q(\hat{\boldsymbol{v}}_i,\boldsymbol{y}). \tag{5}$$

The final output is then given by $\hat{\boldsymbol{v}} \triangleq \hat{\boldsymbol{v}}_{\hat{\imath}}$.

The above scheme includes both separate and joint source-channel coding as special cases. For $N_k = 1$, the scheme corresponds to a concatenation of a source code and a channel code $\mathcal{C}_1$ with intermediate rate $\frac{1}{n}\log\big|\mathcal{A}_k^{(1)}\big|$. Similarly, when $\mathcal{A}_k^{(0)} = \emptyset$ and the subsets $\mathcal{A}_k^{(i)}$, $i = 1,\ldots,N_k$, coincide with the source-message types the overall decoder is MAP, and we recover the best joint source-channel coding scheme [5].

For later use, let the Gallager channel and source functions be given by

$$E_0(\rho,Q) \triangleq -\log\sum_y\left(\sum_x Q(x)W(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}, \tag{6}$$

and

$$E_{\mathrm{s}}(\rho) \triangleq \log\left(\sum_v P(v)^{\frac{1}{1+\rho}}\right)^{1+\rho}, \tag{7}$$

respectively. Also, for a sequence of partitions $\big\{\mathcal{A}_k^{(i)}\big\}$, $k = 1,2,\ldots$, we define

$$E_{\mathrm{s}}^{(i)}(\rho) \triangleq \lim_{k\to\infty}\frac{1}{k}\log\left(\sum_{\boldsymbol{v}\in\mathcal{A}_k^{(i)}} P^k(\boldsymbol{v})^{\frac{1}{1+\rho}}\right)^{1+\rho}. \tag{8}$$

III. Error Exponent Analysis

In this section, we analyze the error probability for a given partition of the source-message set. First, we express the error probability of a given code as

$$\epsilon_n = \Pr\big\{\boldsymbol{V}\in\mathcal{A}_k^{(0)}\big\} + \sum_{i=1}^{N_k}\Pr\big\{\boldsymbol{V}\in\mathcal{A}_k^{(i)},\, \hat{\boldsymbol{V}}_i \neq \boldsymbol{V}\big\} + \sum_{i=1}^{N_k}\Pr\big\{\boldsymbol{V}\in\mathcal{A}_k^{(i)},\, \hat{\boldsymbol{V}}_i = \boldsymbol{V},\, \hat{I}\neq i\big\}, \tag{9}$$

where the first summand corresponds to the event that a source message belongs to the set $\mathcal{A}_k^{(0)}$; the second to the ML decoding error when the correct class is considered; and the last one to the event that the wrong class index is selected in the second decoding stage, i. e., a MAP decoding error.

A. Lower Bounds

Upper bounds on the error probability lead to lower bounds on the exponent. To this end, we start by upper bounding each of terms within the third summand in (9),

$$\begin{aligned}\Pr\big\{&\boldsymbol{V}\in\mathcal{A}_k^{(i)},\, \hat{\boldsymbol{V}}_i = \boldsymbol{V},\, \hat{I}\neq i\big\}\\ &\le \Pr\big\{\boldsymbol{V}\in\mathcal{A}_k^{(i)},\, \hat{I}\neq i \,\big|\, \hat{\boldsymbol{V}}_i = \boldsymbol{V}\big\} &(10)\\ &\le \Pr\left\{\boldsymbol{V}\in\mathcal{A}_k^{(i)},\, q(\boldsymbol{V},\boldsymbol{Y}) \le \max_{\bar{\boldsymbol{v}}\neq\boldsymbol{V},\bar{\boldsymbol{v}}\notin\mathcal{A}_k^{(0)}} q(\bar{\boldsymbol{v}},\boldsymbol{Y})\right\}, &(11)\end{aligned}$$

where (10) follows from the chain rule; and (11) follows from (5) by increasing the set of source messages over which the maximum is computed and assuming that ties are decoded as errors.

Substituting (11) into (9), we obtain

$$\begin{aligned}\epsilon_n \le{}& \Pr\big\{\boldsymbol{V}\in\mathcal{A}_k^{(0)}\big\} + \sum_{i=1}^{N_k}\Pr\big\{\boldsymbol{V}\in\mathcal{A}_k^{(i)},\, \hat{\boldsymbol{V}}_i\neq\boldsymbol{V}\big\}\\ &+ \Pr\left\{\boldsymbol{V}\notin\mathcal{A}_k^{(0)},\, q(\boldsymbol{V},\boldsymbol{Y}) \le \max_{\bar{\boldsymbol{v}}\neq\boldsymbol{V},\bar{\boldsymbol{v}}\notin\mathcal{A}_k^{(0)}} q(\bar{\boldsymbol{v}},\boldsymbol{Y})\right\}.\end{aligned} \tag{12}$$

Without loss of generality, assume that $P^k(\boldsymbol{v}) > 0$ for all $\boldsymbol{v}$. Consider the partition such that source messages are assigned depending on their probability, i. e.,

$$\mathcal{A}_k^{(i)} = \big\{\boldsymbol{v} \,\big|\, \gamma_i^k < P^k(\boldsymbol{v}) \le \gamma_{i+1}^k\big\},\ i = 0,\ldots,N_k, \tag{13}$$

with $\gamma_0 = 0 \le \gamma_1 \le \ldots \le \gamma_{N_k+1} = 1$. The thresholds $\gamma_1,\ldots,\gamma_{N_k-1}$ should be properly selected to optimize the system performance. Let assign a distribution $Q_i(x)$ to each class, $\mathcal{A}_k^{(i)}$, $i = 1,\ldots,N_k$. Then, for each source message $\boldsymbol{v}\in\mathcal{A}_k^{(i)}$, we may randomly generate the codeword $\boldsymbol{x}(\boldsymbol{v})$ according to $Q_i^n(\boldsymbol{x}) \triangleq \prod_{j=1}^n Q_i(x_j)$, $i = 1,\ldots,N_k$.

We define $t \triangleq \lim_{n\to\infty}\frac{k}{n}$ and $R \triangleq \max_{i=1}^{N_k} R_i$. The next result follows from (12) using the exponential bounds [6, Th. 5.2], [2, Th. 5.6.1] and [5, Th. 1] via the random-coding argument.

Theorem 1. *There exists a sequence of codes and partitions (13) such that, for any $\rho\ge 0$, $\rho_i\in[0,1]$, $i = 1,\ldots,N_\infty$, the following exponent is achievable*

$$\min\left\{\rho R - tE_s(\rho,P),\ \min_{i=1,\ldots,N_\infty}\Big\{E_0(\rho_i,Q_i) - \rho_i R_i - tE_s^{(i)}(0,P)\Big\}\right\}. \tag{14}$$

Further analysis involves optimization over R_i, parameters ρ,ρ_i, and distributions Q_i, for $i = 1,\ldots,N_\infty$.

For $N_k = 1$, it follows that $R_1 = R$ and $E_{\mathrm{s}}^{(1)}(0,P) = 0$. Hence, (14) leads to the separate source-channel coding exponent [1]

$$\min\Big\{\rho R - tE_{\mathrm{s}}(\rho),\ E_0(\rho',Q) - \rho' R\Big\}, \tag{15}$$

for $\rho\ge 0$ and $\rho'\in[0,1]$. Considering more than one class improves the exponent, as we will see later.

B. Upper Bounds

We now turn our attention to deriving upper bounds on the error exponent of our construction. Disregarding the last summand in (9) we may lower bound the error probability of a given code as

$$\epsilon_n \geq \Pr\{\boldsymbol{V} \in \mathcal{A}_k^{(0)}\} + \sum_{i=1}^{N_k} \Pr\{\boldsymbol{V} \in \mathcal{A}_k^{(i)}, \hat{\boldsymbol{V}}_i \neq \boldsymbol{V}\}. \quad (16)$$

While it is tempting to apply source and channel coding converses to the respective summands in (16), this must be done with care. A subtle point here is that channel coding converses are usually derived for the maximal error probability, and then extended to the average error. The required polynomial relationship between maximal and average error probabilities holds only for either equiprobable messages or certain codes and channels.

In particular, for linear codes with an appropriately defined randomized ML decoder and for channels such as discrete additive-noise channels and erasure channels, the average and maximal error probabilities coincide [7, Appx. A]. These channels include as particular examples the binary symmetric channel, binary erasure channel or phase-shift-keying modulated additive white Gaussian noise (AWGN) channel. If we restrict the codes $\mathcal{C}_i$ to be linear and the ML decoder (3) resolves the ties at random, the average error probability $\Pr\{\hat{\boldsymbol{V}}_i \neq \boldsymbol{V} \,|\, \boldsymbol{V} \in \mathcal{A}_k^{(i)}\}$ can thus be bounded by any converse bound on the maximal error probability of a channel code of rate R_i, even when the messages are non-equiprobable. In particular, applying the exponential bounds [6, Th. 5.6] and [8, Th. 19] via (16) we obtain the following result.

Theorem 2. *Let W be an additive-noise or erasure discrete channel. Then, for the code construction in Sec. II with $\{\mathcal{A}_k^{(i)}\}$ given in (13) and $\{\mathcal{C}_i\}$ a set of linear codes with randomized ML decoding, any achievable error exponent E is upper bounded by*

$$\min\Big\{ \max_{\rho\geq 0}\{\rho R - tE_s(\rho, P)\}, \min_{i=1,\ldots,N_\infty} \max_{\rho_i\geq 0} \Big\{E_0(\rho_i) - \rho_i R_i - tE_s^{(i)}(0,P)\Big\}\Big\}. \quad (17)$$

where $E_0(\rho) \triangleq \max_Q E_0(\rho, Q)$.

If the optimizing values of ρ_i, $i = 1, \ldots, N_k$, in (17) are smaller than 1 the bounds from Theorems 1 and 2 coincide. While this converse result only applies to a specific family of channels and codes, it shows the tightness of the lower bound in Theorem 1 under certain assumptions.

C. Example

A binary memoryless source (BMS) is to be transmitted over a binary-input complex AWGN channel with signal-to-noise ratio (SNR) E_s/N_0. For a fair comparison, we normalize E_s/N_0 with respect to the number of transmitted *information* bits, i. e., the source entropy $H(V) = h(p)$, where $h(p) \triangleq -p\log_2 p - (1-p)\log_2(1-p)$ denotes the binary entropy function. We define the signal-to-noise ratio per source bit as

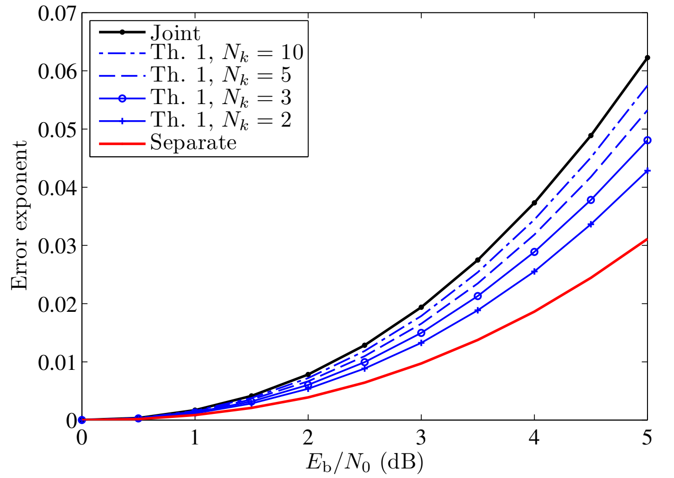

Fig. 1. Error exponent bounds. BMS with $P(1) = 0.1$ transmitted over a binary input AWGN channel for $t = 1$.

$$\frac{E_b}{N_0} \triangleq \frac{n}{kh(p)} \frac{E_s}{N_0}. \quad (18)$$

Figure 1 shows the achievable error exponents for different coding schemes as a function of E_b/N_0. The schemes considered are joint source-channel coding [1], separate source-channel coding (15), and our multi-class scheme for the partition (13) with $N_k = 2, 3, 5, 10$ when optimized over the thresholds γ_i, intermediate rates R_i, and parameters ρ, ρ_i, $i = 1, \ldots, N_k$. From Fig. 1 we can see that the multi-class construction approaches the best source-channel error exponent as the number of classes increases. Moreover, with just a small increase in complexity the scheme with $N_k = 2$ classes, shows a 0.4-0.7 dB improvement over separate source coding. In the following, we restrict ourselves to a low-complexity scheme with $N_k = 2$.

For a BMS, it is possible to obtain a closed-form expression for the source terms in (16). We assume without loss of generality that $p \triangleq P(1) \leq 1/2$ and we define

$$b_n(p, w) \triangleq \binom{n}{w} p^w (1-p)^{n-w}, \quad (19)$$

$$B_n(p, w_1, w_2) \triangleq \sum_{w=w_1}^{w_2} b_n(p, w), \quad w_1 \leq w_2, \quad (20)$$

to be the binomial distribution with parameter w and the probability of $w \in [w_1, w_2]$, respectively. The best source encoder strategy with $N_k = 2$ is to encode the sequences of Hamming weight $w \in \{w_1, \ldots, w_2\}$ with the first (lower-rate) channel code and the sequences of weight $w \in \{w_2+1, \ldots, w_3\}$ with the second (higher-rate) code. All other sequences are transmitted by some fixed codeword which leads to decoding

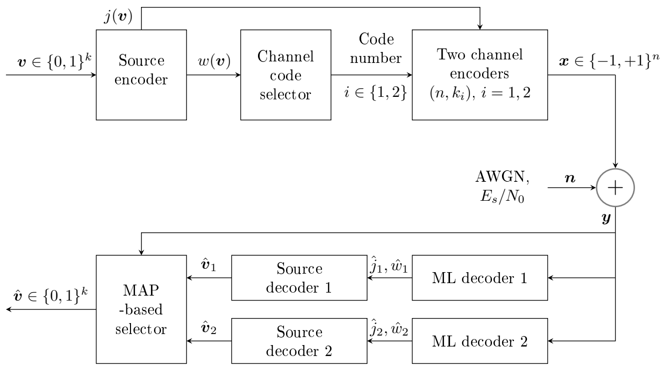

Fig. 2. Implementation of a two-class JSCC system.

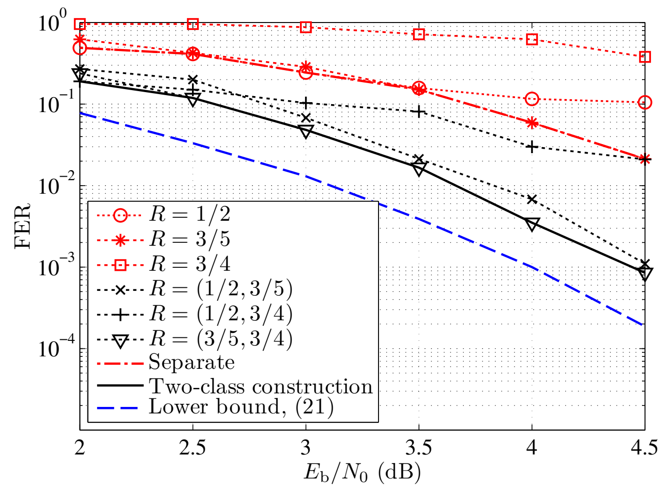

Fig. 3. Error rates for separate coding, JSCC using different pairs of linear codes, and optimized JSCC using two linear codes, $n = 100$, $k = 80$.

error. Then, from (16) we obtain

$$\epsilon_n \geq \min_{w_1,w_2,w_3=0,\ldots,k} \Big\{ B_k(p, w_1, w_2)\underline{\epsilon}_B(n, R(w_1, w_2)) + B_k(p, w_2+1, w_3)\underline{\epsilon}_B(n, R(w_2+1, w_3)) + B_k(p, 0, w_1-1) + B_k(p, w_3+1, k) \Big\}, \tag{21}$$

where

$$R(w_1, w_2) = \frac{1}{n} \left\lceil \log_2 \sum_{w=w_1}^{w_2} \binom{k}{w} \right\rceil, \tag{22}$$

and $\underline{\epsilon}_B(n, R)$ is a lower bound to the error probability of a block channel code with rate $R \geq 0$ and length n. This expression will be used in the next section to lower-bound the frame error rate (FER). As $\underline{\epsilon}_B(n, R)$ we use Shannon's sphere-packing bound [9], since it is accurate for relatively short codes and low SNRs [10]. In order to compute the bound in [9] we use the approximation from [11], very accurate for error probability below 0.1.

IV. Practical Code Design

In this section, we describe a practical source-channel code with two classes and illustrate its performance by means of simulations. A block diagram of the scheme is shown in Fig. 2. We consider a fixed-to-variable lossless source code followed by two linear (n, k_i)-codes $\mathcal{C}_i$, $i = 1, 2$.

A. Encoding

The Hamming weight (type) of the source sequence determines which one of two available codes will be used to encode each source message. For a given source sequence $\boldsymbol{v} \in \{0,1\}^k$ a binary enumerative encoder [12] first computes a pair of integer numbers $(w(\boldsymbol{v}), j(\boldsymbol{v}))$, where $w = w(\boldsymbol{v}) \in \{0, 1, \ldots, k\}$ is the Hamming weight of $\boldsymbol{v}$, and $j = j(\boldsymbol{v}) \in \left\{1, 2, \ldots, \binom{k}{w}\right\}$ is the index of the actual sequence $\boldsymbol{v}$ in the lexicographically ordered list of all possible sequences with weight w. With very small loss of optimality both w and j can be losslessly encoded by codewords of lengths $L_w = \lceil \log_2(k+1) \rceil$ and $L_j = \left\lceil \log_2 \binom{k}{w} \right\rceil$, producing the codeword of overall length $L(\boldsymbol{v}) = L_w + L_j$. Enumerative coding can be implemented with linear complexity via arithmetic coding [12].

If either $w(\boldsymbol{v}) < w_{\min}$ or $w(\boldsymbol{v}) > w_{\max}$ holds, then a source coding error is reported. If not, if $w(\boldsymbol{v}) \leq w_{\text{th}}$ then the source codeword length $L(\boldsymbol{v}) \leq k_1$ and the channel code $\mathcal{C}_1$ is used for transmission, otherwise $L(\boldsymbol{v}) \leq k_2$ and the second code, $\mathcal{C}_2$, is used. In this encoding scheme $l = k_i - L(\boldsymbol{v})$ leftover bits may appear; these are used for error detection. To do this, we choose a pseudo-random binary matrix $H(w, j)$ of size $L \times l$ and we assume that this matrix is known to the decoder. Multiplying the information vector of length L by $H(w, j)$ we obtain a parity-check sequence $\boldsymbol{c}$ of l bits which is included in the information payload of the selected code.

B. Decoding

At the decoder, two decoding attempts are performed in parallel. In each of the branches corresponding to each of the two possible codes, the receiver performs ML decoding. The decoding result of each of the branches is then checked to be compatible with the encoding scheme and the additional parity check written on the leftover positions. If only one of two codes —which is the typical case— passes this compatibility test, then the corresponding output is used as the decoder output. If both decoders fail, a predetermined decoder output, for example all-zero data sequence, is used. Finally, if both decoders report decoding success, the decision with larger a posteriori likelihood is selected and the corresponding message is accepted as a final decision.

C. Simulation Results

The performance of this coding scheme has been evaluated for the transmission of a BMS with $p = 0.1$ over an AWGN channel in two scenarios.

In the first scenario we consider tail biting (TB) block codes of length $n = 100$. The length of the source sequence is chosen to be $k = 80$. The proposed scheme requires a set of codes that fit to a particular source realization and E_b/N_0. We have chosen TB block codes of three rates $R = 1/2, 3/5$ and $3/4$. The code of rate $R = 1/2$ was taken from [13], that contains

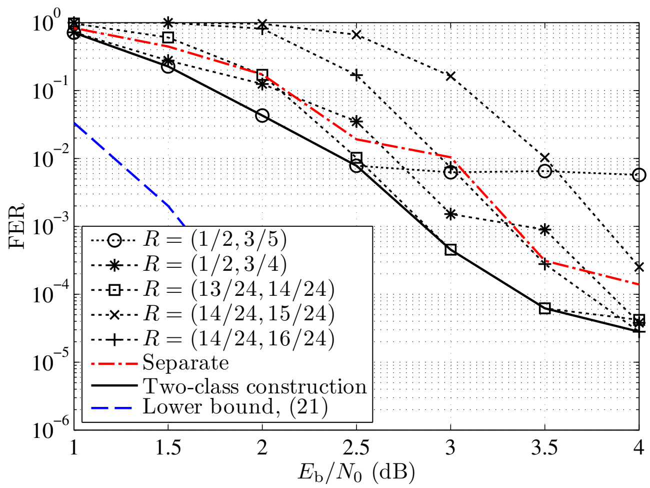

Fig. 4. Error rates for separate coding, JSCC using different pairs of linear codes, and optimized JSCC using two linear codes, $n = 1008$, $k = 1000$.

tables of low-rate low-complexity TB codes. For the other rates we did a short search for high-rate convolutional codes using techniques from [14], [15]. Among the most efficient near ML decoding algorithms we have selected BEAST [16] which allows near ML decoding for codes of length 100 with acceptable complexity. The simulated FER performance is shown in Fig. 3. The curves "Separate" and "Two-class code" show the best performance obtained within the corresponding family of codes.

In the second scenario, we fix $k = 1000$, $n = 1008$ and as channel codes we use quasi-cyclic (QC) LDPC codes with base matrix containing $c = 24$ columns and rates $R = 12/24, 13/24, \ldots, 16/24$. For constructing these parity-check matrices we used the optimization algorithm from [17]. The only exception is the code of rate $R = 18/24 = 2/3$ which is borrowed from [18, code A]. The FER after 50 iterations of belief propagation decoding is shown in Fig. 4.

We conclude from the presented plots that the proposed scheme outperforms separate coding by 1 dB and by about 0.4-0.7 dB in the first and second scenario, respectively, in agreement with the values predicted by the random coding analysis.

References

[1] I. Csiszár, "Joint source-channel error exponent," *Probl. Contr. Inf. Theory*, vol. 9, pp. 315–328, 1980.

[2] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, Inc., 1968.

[3] Y. Zhong, F. Alajaji, and L. L. Campbell, "On the joint source-channel coding error exponent for discrete memoryless systems," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1450–1468, Apr. 2006.

[4] M. Fresia, F. Peréz-Cruz, H. Poor, and S. Verdú, "Joint source and channel coding," *IEEE Signal Process. Mag.*, vol. 27, no. 6, pp. 104–113, Nov. 2010.

[5] A. Tauste Campo, G. Vazquez-Vilar, A. Guillén i Fàbregas, T. Koch, and A. Martinez, "A derivation of the source-channel error exponent using non-identical product distributions," *IEEE Trans. Inf. Theory*, 2014, in press.

[6] F. Jelinek, *Probabilistic Information Theory*. New York: McGraw-Hill, 1968.

[7] Y. Polyanskiy, H. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[8] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory*, vol. 20, no. 4, pp. 405–417, Jul. 1974.

[9] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, no. 3, pp. 611–656, May 1959.

[10] I. Sason and S. Shamai (Shitz), *Performance analysis of linear codes under maximum-likelihood decoding: a tutorial*. Foundations and Trends Commun. and Inf. Theory, now Publishers, 2006.

[11] S. Vialle and J. Boutros, "Performance of optimal codes on Gaussian and Rayleigh fading channels: a geometrical approach," in *37th Ann. Allerton Conf. Comm., Contr. and Comp.*, Allerton, IL, USA, Sep. 1999.

[12] T. Cover, "Enumerative source encoding," *IEEE Trans. Inf. Theory*, vol. 19, no. 1, pp. 73–77, Jan. 1973.

[13] I. E. Bocharova, R. Johannesson, B. D. Kudryashov, and P. Stahl, "Tailbiting codes: Bounds and search results," *IEEE Trans. Inf. Theory*, vol. 48, no. 1, pp. 137–148, Jan. 2002.

[14] F. Hug, I. E. Bocharova, R. Johannesson, and B. D. Kudryashov, "Searching for high-rate convolutional codes via binary syndrome trellises," in *2009 IEEE Int. Symp. Inf. Theory*, Seoul, Korea, Jun.–Jul. 2009.

[15] I. E. Bocharova and B. D. Kudryashov, "Rational rate punctured convolutional codes for soft-decision Viterbi decoding," *IEEE Trans. Inf. Theory*, vol. 43, no. 4, pp. 1305–1313, Jul. 1997.

[16] I. E. Bocharova, R. Johannesson, B. D. Kudryashov, and M. Lonĉar, "BEAST decoding for block codes," *Eur. Tran. Tel.*, vol. 15, no. 4, pp. 297–305, Jul.–Aug. 2004.

[17] I. Bocharova, B. Kudryashov, and R. Johannesson, "Combinatorial optimization for improving QC LDPC codes performance," in *2013 IEEE Int. Symp. Inf. Th.*, Istanbul, Turkey, Jul. 2013.

[18] *Air Interface for Fixed and Mobile Broadband Wireless Access Systems*, IEEE P802.16e/D12 Draft, Oct. 2005.