

Bounds on Information Combining

Ingmar Land, *Member, IEEE*, Simon Huettinger, Peter A. Hoeher, *Senior Member, IEEE*, and Johannes B. Huber, *Senior Member, IEEE*

Abstract—When the same data sequence is transmitted over two independent channels, or when a data sequence is transmitted twice but independently over the same channel, the independent observations can be combined at the receiver side. From an information-theory point of view, the overall mutual information between the data sequence and the received sequences represents a combination of the mutual information of the two channels. This concept is termed information combining. In this paper, a lower bound and an upper bound on the combined information is presented, and it is proved that these bounds are tight. Furthermore, this principle is extended to the computation of extrinsic information on single code bits for a repetition code and for a single parity-check code of length three, respectively.

For illustration of the concept and the bounds on information combining, two applications are considered. First, bounds on the information processing characteristic (IPC) of a parallel concatenated code are derived from its extrinsic information transfer (EXIT) chart. Second, bounds on the EXIT chart for an outer repetition code and for an outer single parity-check code of a serially concatenated coding scheme are computed.

Index Terms—Extrinsic information, extrinsic information transfer (EXIT) chart, information combining, information processing characteristic (IPC), iterative decoding, mutual information, parallel concatenated codes.

I. INTRODUCTION

WHEN binary linear channel codes are used for transmission over memoryless channels, several or even all observations of code bits contain information on a certain information bit (also denoted as info bit,¹ data bit, or source bit). The overall information on this info bit is formed by a combination of the available amounts of information.

If the information is represented by outputs of independent additive white Gaussian noise (AWGN) channels, the optimum

combining is given by maximum-ratio combining. A very convenient way of combining is possible, if the information is represented by independent log-likelihood ratios (LLRs). In this case, the optimum combination is simply the addition of the LLRs [1].

This combining of information is used for decoding of single codes, e.g., [2]–[4], as well as for iterative decoding of parallel and serially concatenated codes (turbo codes), e.g., [5]–[7], and for iterative decoding of low-density parity-check codes, e.g., [8], [9]. The general framework for iterative decoding is decoding on graphs, see, e.g., [10]–[12].

In this paper, “information combining” will be used in a very strict sense, namely, only for combining of mutual information. This notion of information combining was introduced in [13], [14] for design and analysis of parallel concatenated coding schemes. In such a coding scheme, the overall mutual information between the encoder inputs and the soft outputs of the iterative decoder is a combination of the mutual information between the encoder inputs and the extrinsic outputs of each constituent decoder. A similar concept is used in the extrinsic information transfer (EXIT) chart method [15], [16], in which iterative decoding is interpreted as processing of extrinsic (mutual) information.

The general scenario for information combining is as follows. Consider a certain number of code bits which are coupled by code constraints. Such a constraint can be that code bits are required to be equal, denoted as equality constraint, or that code bits are required to fulfill a parity-check equation, denoted as parity-check constraint. The code bits are transmitted over independent memoryless channels, and for each channel, the mutual information is known. Then the question is: What is the mutual information between a single code bit or info bit and all channel outputs? This mutual information is denoted as combined information, since it is in some sense a combination of the mutual informations of the individual channels. This general case of transmission over independent channels includes the special and practically relevant case of subsequent transmissions over the same memoryless channel.

If the statistical properties of the channels are completely known, the combined information can be computed exactly. On the other hand, if only one statistical parameter of each channel is known, namely, its mutual information, still an upper and a lower bound on the combined information can be given. Such bounds on information combining are addressed in this paper.

We restrict ourselves to binary-input symmetric memoryless channels (BISMCs). Examples for such channels are the binary-symmetric channel (BSC), the binary erasure channel (BEC), and the binary-input AWGN channel (assuming the input values “+1” and “−1”). Regarding information combining, we focus on three basic scenarios.

Manuscript received October 6, 2003; revised July 10, 2004. The material in this paper was presented in part at the International Symposium on Turbo Codes and Related Topics, Brest, France, September 2003.

I. Land was with the Information and Coding Theory Laboratory, Faculty of Engineering, University of Kiel, D-24143 Kiel, Germany. He is now with the Digital Communication Division, Department of Communication Technology, Aalborg University, 9220 Aalborg, Denmark (e-mail: il@kom.aau.dk).

S. Huettinger was with the Institute for Information Transmission, University Erlangen-Nürnberg, D-91058 Erlangen, Germany. He is now with Siemens AG, D-91065 Erlangen, Germany (e-mail: simon.huettinger@int.de).

P. A. Hoeher is with the Information and Coding Theory Laboratory, Faculty of Engineering, University of Kiel, D-24143 Kiel, Germany (e-mail: ph@tf.uni-kiel.de).

J. B. Huber is with the Institute for Information Transmission, University Erlangen-Nürnberg, D-91058 Erlangen, Germany (e-mail: huber@int.de).

Communicated by A. E. Ashikhmin, Associate Editor for Coding Theory.
Digital Object Identifier 10.1109/TIT.2004.840883

¹In the following, we employ the term “info bit” to avoid confusion with “mutual information.”

The first scenario is the parallel concatenation of two independent BSMCs $X \rightarrow Y_1$ and $X \rightarrow Y_2$, denoted as constituent channels, having the same input X . The overall channel, $X \rightarrow [Y_1 Y_2]$, is denoted as parallel concatenated channel² (PCCh). The main question in this context is: How is the mutual information of the PCCh (the combined information) related to the mutual informations of the two constituent channels.

The second and the third scenario comprises three code bits X_1, X_2 , and X_3 , which are transmitted over independent³ BSMCs $X_i \rightarrow Y_i, i = 1, 2, 3$. In the second scenario, the code bits are required to fulfill an equality constraint, and in the third scenario, the code bits are required to fulfill a parity-check constraint. In both scenarios, the question is: How is the extrinsic information on code bit $X_1, I(X_1; Y_2 Y_3)$, related to the mutual informations of the channels corresponding to code bits X_2 and $X_3, I(X_2; Y_2)$ and $I(X_3; Y_3)$.

For each of the three scenarios, upper and lower bounds on the combined information are given. Furthermore, it is shown that each bound is achieved by either the case that all BSMCs are BECs or the case that all BSMCs are BSCs; thus, the bounds are tight.

The notion of information combining was introduced in [13] in the sense of *optimal combining* of mutual information. The combined information is the mutual information between the respective info bit or code bit and all channel outputs. Motivated by [17], bounds on combined information were devised for the parallel concatenation of two channels (first scenario described above) in [18]. In the present paper, this concept is revised, and it is generalized to the second and the third scenario described above. For all three scenarios, we investigate the maximum and the minimum of the combined information if for all individual channels only their mutual informations are given.

Motivated by [18], results similar to that of the present paper were independently found in [19], using a different approach. Information combining is considered for equality and parity-check constraints involving an arbitrary number of channels, and the extremes of the combined information are determined with respect to single channels; the results are applied to analyze iterative decoders for low-density parity-check codes. Furthermore, a more general notion of information combining is introduced, but only optimal combining is considered. As opposed to that, the present paper focuses on the case of two and three channels, and the extremes are determined with respect to all channels; as applications, the information-processing characteristic (IPC) of a parallel concatenated code is computed based on its EXIT chart, and the EXIT charts for repetition codes and single parity-check codes of length 3 are considered.

The paper is organized as follows. In Section II, some definitions and properties for the single constituent channels are given. In Section III, the parallel-concatenated channel (PPCh) is addressed, and tight bounds on information combining are stated and proven. Section IV deals with the extrinsic information for the two simple codes given above and the corresponding

bounds. For each scenario, an example is given in Section V. Finally, conclusions are drawn in Section VI.

II. CONSTITUENT CHANNELS

In this section, the constituent channels will be considered separately. Since all constituent channels are BSMCs and thus have the same properties, we will label the variables with the generic subindex i , where $i \in \{1, 2, 3\}$.

In the following, it will be shown how a BSMC can be separated into subchannels which are BSCs. For this separation, the absolute value of the channel output will be employed.

Let $X_i \rightarrow Y_i$ denote a BSMC with $X_i \in \mathbb{X} := \{-1, +1\}$ and $Y_i \in \mathbb{Y}_i \subset \mathbb{R}$, where \mathbb{X} and \mathbb{Y}_i denote the input and the output alphabet of the channel, respectively. The channel inputs are assumed to be independent and uniformly distributed. The transition probabilities are given by $p_{Y_i|X_i}(y|x)$, denoting the probability density function for continuous output alphabets and the probability mass function for discrete output alphabets. Since the channel is symmetric, we can assume

$$p_{Y_i|X_i}(y|x) = p_{Y_i|X_i}(-y|-x)$$

for all $x \in \mathbb{X}$ and $y \in \mathbb{Y}_i$ without loss of generality. The mutual information of a constituent channel is defined as

$$I_i := I(X_i; Y_i).$$

Let the magnitude of Y_i be defined as the random variable $J_i \in \mathbb{J}_i := \{y \in \mathbb{Y}_i : y \geq 0\}$

$$J_i := |Y_i|.$$

Using J_i , the elements of the output alphabet \mathbb{Y}_i can be grouped into pairs

$$\mathbb{Y}_i(j) := \begin{cases} \{+j, -j\}, & \text{for } j \in \mathbb{J}_i \setminus \{0\} \\ \{0, 0\}, & \text{for } j = 0. \end{cases}$$

(The special treatment of the case $j = 0$ will be explained below.) With these definitions, J_i indicates which output set $\mathbb{Y}_i(j)$ the output symbol Y_i belongs to.

The random variable J_i separates the symmetric channel $X_i \rightarrow Y_i$ into strongly symmetric subchannels $X_i \rightarrow Y_i|J_i = j$; it will therefore be denoted as *subchannel indicator*. The subchannels are BSCs and occur with probability

$$q_i(j) := p_{J_i}(j), \text{ for } j \in \mathbb{J}_i.$$

Their conditional crossover probabilities $\epsilon_i(j)$ are defined as

$$\epsilon_i(j) := \begin{cases} p_{Y_i|X_i, J_i}(-j|+1, j), & \text{for } j \in \mathbb{J}_i \setminus \{0\} \\ \frac{1}{2}, & \text{for } j = 0. \end{cases}$$

Let

$$h(x) := -x \text{ ld } x - (1-x) \text{ ld } (1-x) x \in [0, 1],$$

denote the binary entropy function, and let $h^{-1}(y), y \in [0, 1]$, denote its inverse for $x \in [0, \frac{1}{2}]$. Then, the mutual information of subchannel j is given as

$$I_i(j) := I(X_i; Y_i|J_i = j) = 1 - h(\epsilon_i(j)). \quad (1)$$

²The term ‘‘parallel concatenated channel’’ follows the term ‘‘parallel concatenated code.’’

³As mentioned above, this includes the special and practically relevant case that all code bits are transmitted over the same memoryless channel.

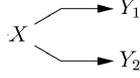


Fig. 1. Parallel concatenation of two channels.

As mentioned earlier, $j = 0$ was treated as a special case in the definitions. The actual subchannel, a BEC with erasure probability 1, was transformed into a BSC with crossover probability $\frac{1}{2}$. This transformation does not change the mutual information $I_i(0) = 0$, but will simplify the following derivations, since thus all subchannels are BSCs without exceptions.

Using the above definitions, the mutual information of a constituent channel can be written as the expected value of the mutual informations of its subchannels

$$I_i = \mathbb{E}_{j \in \mathcal{J}_i} \{I_i(j)\}. \quad (2)$$

The concept of separating a BISMCh into BSCs can easily be generalized to include channels with vector-valued outputs, using definitions similar to those for channels with scalar outputs. This is not further considered as this paper addresses only BISMChs with scalar outputs.

The separation of the constituent channels into binary-symmetric subchannels will be exploited in Sections III–V.

III. PARALLEL-CONCATENATED CHANNEL (PCCh)

The PCCh $X \rightarrow [Y_1 Y_2]$ is composed of the two constituent channels $X \rightarrow Y_1$ (CCh1) and $X \rightarrow Y_2$ (CCh2), as depicted in Fig. 1. The channel inputs X are assumed to be independent and uniformly distributed, and the constituent channels are assumed to be BISMChs. It can easily be seen that then the PCCh is a symmetric memoryless channel (with vector-valued outputs).

In the following, it will be investigated how the mutual information of the PCCh

$$I := I(X; Y_1 Y_2) \quad (3)$$

is related to the mutual informations $I_1 = I(X; Y_1)$ and $I_2 = I(X; Y_2)$ of CCh1 and CCh2, respectively.

First, I can be written as

$$\begin{aligned} I &= I(X; Y_1 Y_2) \\ &= I(X; Y_1) + I(X; Y_2) - I(Y_1; Y_2) \\ &= I_1 + I_2 - I(Y_1; Y_2). \end{aligned} \quad (4)$$

Since $I \leq I_1 + I_2$, the value of $I(Y_1; Y_2)$ can be regarded as the *information defect*⁴ with respect to the combination of the mutual informations I_1 and I_2 .

For illustration, consider the following two examples (cf. [14]). These examples will turn out to represent bounds on information combining.

Example 1 (Two BECs): Let the two constituent channels be BECs having erasure probabilities δ_1 and δ_2 , respectively. Then their mutual informations are given by $I_1 = 1 - \delta_1$ and $I_2 = 1 - \delta_2$. If the output of at least one constituent channel

is no erasure, the input can be recovered without error; accordingly, the mutual information for this case is equal to 1. On the other hand, if the outputs of both channels are erasures, the input cannot be recovered; accordingly, the mutual information for this case is equal to 0. Since the former case occurs with probability $1 - \delta_1 \delta_2$, the combined information is given by

$$\begin{aligned} I(X; Y_1 Y_2) &= 1 - \delta_1 \delta_2 \\ &= 1 - (1 - I_1)(1 - I_2) \\ &= I_1 + I_2 - I_1 \cdot I_2. \end{aligned}$$

Using (4), the information defect

$$I(Y_1; Y_2) = I_1 \cdot I_2$$

is obtained.

Example 2 (Two BSCs): Let the two constituent channels be BSCs having crossover probabilities ϵ_1 and ϵ_2 , respectively. Then their mutual informations are given by $I_1 = 1 - h(\epsilon_1)$ and $I_2 = 1 - h(\epsilon_2)$. For this example, it is easier to derive the information defect first. The information defect is given by the end-to-end mutual information of the serial concatenation of the two BSCs $Y_1 \rightarrow X$ and $X \rightarrow Y_2$. Since X is uniformly distributed, the crossover probability of the reversed constituent channel $Y_1 \rightarrow X$ is also given by ϵ_1 . Thus, the crossover probability of the BSC $Y_1 \rightarrow Y_2$ can be computed as

$$\epsilon := [1 - \epsilon_1]\epsilon_2 + \epsilon_1[1 - \epsilon_2]$$

and we have

$$I(Y_1; Y_2) = 1 - h(\epsilon).$$

Note that $\epsilon_1 = h^{-1}(1 - I_1)$ and $\epsilon_2 = h^{-1}(1 - I_2)$. Using (4), the combined information can be written as

$$I(X; Y_1 Y_2) = I_1 + I_2 + 1 - h(\epsilon).$$

In both examples, the combined information and the information defect are expressed solely by the mutual informations of the constituent channels.

For the subsequent discussion, the following function will show to be useful.

Definition 1 (Information Defect Function): For $x_1, x_2 \in [0, 1]$, the information defect function (IDF) is defined as

$$f(x_1, x_2) := 1 - h([1 - \epsilon_1]\epsilon_2 + \epsilon_1[1 - \epsilon_2])$$

with $\epsilon_1 = h^{-1}(1 - x_1)$ and $\epsilon_2 = h^{-1}(1 - x_2)$.

Thus, for Example 2 (two BSCs) the information defect may be written as $I(Y_1; Y_2) = f(I_1, I_2)$.

Using this function, the main theorem of this paper can be stated as follows.

Theorem 1 (Bounds on Information Combining): Given the mutual informations $I_1 = I(X; Y_1)$ and $I_2 = I(X; Y_2)$ of the constituent channels, the mutual information of the PCCh, $I = I(X; Y_1 Y_2)$, is bounded as

$$\underbrace{I_1 + I_2 - f(I_1, I_2)}_{\text{two BSCs}} \leq I \leq \underbrace{I_1 + I_2 - I_1 \cdot I_2}_{\text{two BECs}}$$

⁴The term “information defect” follows the term “mass defect” used in nuclear physics.

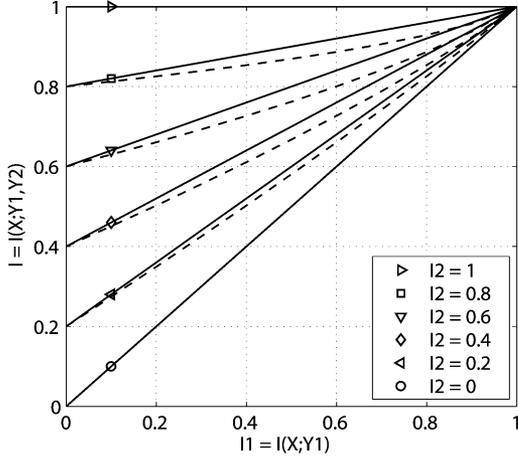


Fig. 2. Bounds on combined information $I = I(X; Y_1 Y_2)$ (PCCh) versus $I_1 = I(X; Y_1)$ (CCh1) for several $I_2 = I(X; Y_2)$ (CCh2). The lower bounds (dashed lines) correspond to the case of two BSCs, the upper bounds (solid lines) correspond to the case of two BECs.

As shown in the examples, the lower bound is achieved if both channels are BSCs, and the upper bound is achieved if both channels are BECs. Thus, for each bound, we have an example actually achieving this bound. Consequently, these two bounds are tight.

In Fig. 2, the bounds on I are plotted versus I_1 with I_2 as parameter (see also [14], [20]). As the gap between the two bounds is very small, it can be concluded that the mutual informations of the constituent channels dominate the value of the combined information, rather than the actual structures of the constituent channels.

Theorem 1 will be proven in three steps. First, the information defect for the general case will be written as an expected value which includes only the IDF. Second, two properties of the IDF will be stated. Finally, these properties will be used to get an upper and a lower bound on the information defect and thus on the mutual information of the PCCh.

Taking the subchannel indicators J_1 and J_2 into account, the information defect can be written as

$$\begin{aligned}
 I(Y_1; Y_2) &= I(Y_1 J_1; Y_2 J_2) \\
 &= \underbrace{I(J_1; Y_2 J_2)}_{=0} + \underbrace{I(Y_1; J_2 | J_1)}_{=0} + I(Y_1; Y_2 | J_1 J_2) \\
 &= \mathbb{E}_{\substack{j_1 \in \mathcal{J}_1 \\ j_2 \in \mathcal{J}_2}} \{I(Y_1; Y_2 | J_1 = j_1, J_2 = j_2)\} \\
 &= \mathbb{E}_{j_1 \in \mathcal{J}_1} \left\{ \mathbb{E}_{j_2 \in \mathcal{J}_2} \{f(I_1(j_1), I_2(j_2))\} \right\}. \quad (5)
 \end{aligned}$$

The following relations were applied: First line: J_1 is a function of Y_1 , and J_2 is a function of Y_2 . Second line: Chain rule of mutual information [21]. Third line: $I(Y_1; Y_2 | J_1, J_2)$ is written as an expectation with respect to realizations of J_1 and J_2 . Last line: For given realizations j_1 and j_2 , the argument of the expectation corresponds to the case of two BSCs and can thus be written using the information defect function; as J_1 and J_2 are statistically independent, the expectation can be separated.

Equation (5) provides a means for the exact evaluation of the information defect. First, the two constituent channels have to be separated into their binary-symmetric subchannels, and

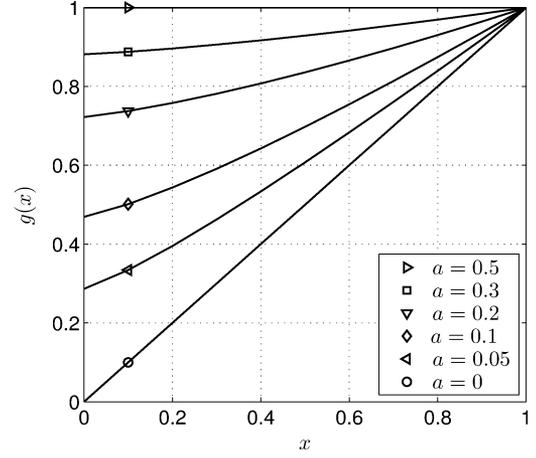


Fig. 3. Function $g(x)$ versus x for several values of parameter a (cf. Lemma 1).

the corresponding values of the subchannel probabilities $q_1(j)$ and $q_2(j)$, and the subchannel crossover probabilities $\epsilon_1(j)$ and $\epsilon_2(j)$ have to be computed. Then, the mutual informations of the subchannels can be determined according to (1). In the final step, (5) can be evaluated.

In the following, (5) will be employed for derivation of an upper and a lower bound for the information defect. Two important properties of the IDF are stated in the following lemma.

Lemma 1 (Properties of the IDF): The information defect function $f(x_1, x_2)$, $x_1, x_2 \in [0, 1]$, has the following two properties:

- $f(x_1, x_2)$ is convex- \cap in x_1 for constant x_2 , and vice versa;
- $f(x_1, x_2)$ is lower-bounded as

$$f(x_1, x_2) \geq x_1 \cdot x_2.$$

Proof:

a) Since the IDF is symmetric in x_1 and x_2 , it is sufficient to consider the IDF as a function of x_1 with constant parameter x_2 . For simplification, let us define the function

$$\begin{aligned}
 g(x) &:= 1 - f(1 - x, 1 - h(a)) \\
 &= h([1 - 2a]h^{-1}(x) + a)
 \end{aligned}$$

$x \in [0, 1]$, with parameter $a \in [0, \frac{1}{2}]$. (The range of a is chosen such that the equality $h^{-1}(h(a)) = a$ can be applied when evaluating the above equation.) Then, $f(x_1, x_2)$ is convex- \cap in x_1 for constant x_2 if and only if $g(x)$ is convex- \cup in x for constant a . For illustration, the function $g(x)$ is plotted versus x for several values of a in Fig. 3. The plot indicates that $g(x)$ is convex- \cup for all a . A formal proof may be found in [22, Lemma 2], as noticed in [19].

b) For the time being, let x_2 be constant. Furthermore, let $x_2 = 1 - h(a)$ and let $x_1 = 1 - x$. Then $g(x)$ can be used to write the (one-dimensional) bound equivalently as

$$h([1 - 2a]h^{-1}(x) + a) \leq (1 - h(a))x + h(a)$$

for $x \in [0, 1]$ and $a \in [0, \frac{1}{2}]$. For $x = 0$ and $x = 1$, the left-hand side is equal to the right-hand side. Regarding this and the fact that $g(x)$ is convex- \cup , the right-hand side represents the secant

of $g(x)$ for $x \in [0, 1]$, and thus the inequality holds. Since these considerations hold for all a , statement b) holds for all x_2 . \square

The results of this lemma will now be used to give bounds on the information defect $I(Y_1; Y_2)$. The lower bound corresponds to the case where both constituent channels are BECs, and the upper bound corresponds to the case where both constituent channels are BSCs.

Lemma 2 (Bounds on the Information Defect): The information defect $I(Y_1; Y_2)$ is bounded as

$$\underbrace{I_1 \cdot I_2}_{\text{two BECs}} \leq I(Y_1; Y_2) \leq \underbrace{f(I_1, I_2)}_{\text{two BSCs}}.$$

Proof: As given in (5), the information defect can be written as

$$I(Y_1; Y_2) = \mathbb{E}_{j_1 \in \mathcal{J}_1} \left\{ \mathbb{E}_{j_2 \in \mathcal{J}_2} \{f(I_1(j_1), I_2(j_2))\} \right\}.$$

Since the function $f(x_1, x_2)$ is convex- \cap in each dimension according to Lemma 1, part a), Jensen's inequality [21] can be applied in the above expression first with respect to (w.r.t.) j_2 and then w.r.t. j_1 . Thus, we obtain

$$\begin{aligned} & \mathbb{E}_{j_1 \in \mathcal{J}_1} \left\{ \mathbb{E}_{j_2 \in \mathcal{J}_2} \{f(I_1(j_1), I_2(j_2))\} \right\} \\ & \leq f \left(\mathbb{E}_{j_1 \in \mathcal{J}_1} \{I_1(j_1)\}, \mathbb{E}_{j_2 \in \mathcal{J}_2} \{I_2(j_2)\} \right) = f(I_1, I_2) \end{aligned}$$

where (2) was applied in the last equation. Notice that this proof is based on the separation of BISMCS into BSCs and the convexity of the IDF. (A similar technique was implicitly used in [23] to extend Mrs. Gerber's Lemma [22].)

On the other hand, since the function $f(x_1, x_2)$ can be lower-bounded according to Lemma 1, part b), the information defect can be lower-bounded as

$$\begin{aligned} & \mathbb{E}_{j_1 \in \mathcal{J}_1} \left\{ \mathbb{E}_{j_2 \in \mathcal{J}_2} \{f(I_1(j_1), I_2(j_2))\} \right\} \\ & \geq \mathbb{E}_{j_1 \in \mathcal{J}_1} \left\{ \mathbb{E}_{j_2 \in \mathcal{J}_2} \{I_1(j_1) \cdot I_2(j_2)\} \right\} = I_1 \cdot I_2. \quad \square \end{aligned}$$

The proof of Theorem 1 follows immediately from (4) and Lemma 2.

IV. EXTRINSIC INFORMATION

In this section, the binary repetition code and the binary single parity-check code, both of length three, will be considered. These simple examples represent two basic code constraints and provide insights concerning bounds on extrinsic information.

Let X_1 , X_2 , and X_3 denote three code bits, which are coupled by the respective constraints, and let the codewords $[X_1, X_2, X_3]$ be equiprobable. The code bits are transmitted over independent BISMCS $X_i \rightarrow Y_i$, $i = 1, 2, 3$ (cf. Footnote 3).

Due to the code constraints, information about a code bit is provided not only by its direct observation, but also by the

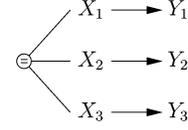


Fig. 4. Three parallel channels coupled by an equality constraint on the inputs.

observations of the other code bits. This *extrinsic information* plays an important role in the context of iterative decoding (see, e.g., [5]). With information in the sense of mutual information (cf.[15], [16]), the formal definition is as follows.

Definition 2: The extrinsic information I_{e1} on code bit X_1 is defined as

$$I_{e1} := I(X_1; Y_2 Y_3)$$

i.e., all channel outputs except the observation Y_1 are taken into account. The extrinsic informations on X_2 and X_3 are defined in an analogous way.

In the following, bounds on the extrinsic information will be derived.

A. Equality Constraint

Consider the case that the three code bits are coupled by the equality constraint

$$X_1 = X_2 = X_3$$

as depicted in Fig. 4.

It is obvious that the extrinsic information $I(X_1; Y_2 Y_3)$ exactly corresponds to the case of the parallel concatenation of two channels, where two independent observations of the same bit are available. Thus, we can apply the bounds given in Theorem 1.

Theorem 2 (Extrinsic Information for Equality Constraint): Given an equality constraint on the code bits X_1 , X_2 , X_3 , and the mutual informations $I_2 := I(X_2; Y_2)$ and $I_3 := I(X_3; Y_3)$, the extrinsic information $I_{e1} := I(X_1; Y_1 Y_2)$ is bounded as

$$\underbrace{I_2 + I_3 - f(I_2, I_3)}_{\text{two BSCs}} \leq I_{e1} \leq \underbrace{I_2 + I_3 - I_2 \cdot I_3}_{\text{two BECs}}.$$

The lower bound is achieved if all channels are BSCs, and the upper bound is achieved if all channels are BECs. The bounds are analogous for code bits X_2 and X_3 .

B. Parity-Check Constraint

Consider now the case that the three code bits are coupled by a parity-check constraint, as depicted in Fig. 5.

Let $C_1, C_2, C_3 \in \mathbb{F}_2 := \{0, 1\}$ be defined as $C_i = 0$ for $X_i = +1$ and $C_i = 1$ for $X_i = -1$, $i = 1, 2, 3$. The code bits X_1 , X_2 , and X_3 are said to fulfill a parity-check constraint if and only if

$$C_1 \oplus C_2 \oplus C_3 = 0.$$

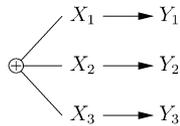


Fig. 5. Three parallel channels coupled by an parity-check constraint on the inputs.

The extrinsic information on code bit X_1 can be expressed as

$$I(X_1; Y_2 Y_3) = H(Y_2 Y_3) - H(Y_2 Y_3 | X_1). \quad (6)$$

The first term may be written as

$$H(Y_2 Y_3) = H(Y_2) + H(Y_3 | Y_2)$$

where $H(Y_3 | Y_2) = H(Y_3)$, because Y_2 and Y_3 are independent if no further information on X_1 or Y_1 is available. The second term may be written as

$$H(Y_2 Y_3 | X_1) = H(Y_2 | X_1) + H(Y_3 | X_1) - I(Y_2; Y_3 | X_1)$$

where $H(Y_2 | X_1) = H(Y_2)$ and $H(Y_3 | X_1) = H(Y_3)$, because Y_2 and X_1 are independent if no further information about X_3 or Y_3 is available, and Y_3 and X_1 are independent if no further information about X_2 or Y_2 is available. Using these equalities in (6), we obtain

$$I(X_1; Y_2 Y_3) = I(Y_2; Y_3 | X_1).$$

For given X_1 , there is a one-to-one relation between X_2 and X_3 . Thus, $I(Y_2; Y_3 | X_1)$ is equal to the information defect of the parallel concatenation of the channels $X_2 \rightarrow Y_2$ and $X_3 \rightarrow Y_3$, and we can apply the bounds given in Lemma 2.

Theorem 3 (Extrinsic Informormation for Parity-Check Constraint): Given a parity-check constraint on the code bits X_1, X_2, X_3 , and the mutual informations $I_2 := I(X_2; Y_2)$ and $I_3 := I(X_3; Y_3)$, the extrinsic information $I_{e1} := I(X_1; Y_2 Y_3)$ is bounded as

$$\underbrace{I_2 \cdot I_3}_{\text{two BECs}} \leq I_{e1} \leq \underbrace{f(I_2, I_3)}_{\text{two BSCs}}. \quad (7)$$

The lower bound is achieved if all channels are BECs, and the upper bound is achieved if all channels are BSCs. (Note that the converse holds for the equality constraint.) The bounds are analogous for code bits X_2 and X_3 .

V. APPLICATIONS

For each of the theorems given in the previous two sections, a practical example will be provided now.

A. Parallel-Concatenated Coding Scheme

Theorem 1 can be used to link the EXIT chart [15] and the IPC [14] of a parallel-concatenated coding scheme (turbo code) [5].

Consider the original turbo code according to [5]: The encoder comprises two linear binary channel encoders as constituent encoders, of which the outputs are only parity bits; the turbo codeword is formed by three parts: the systematic bits, the parity bits computed by the first encoder, and the parity bits computed by the second encoder. The turbo codeword is trans-

mitted over a symmetric memoryless channel (physical channel) that has channel capacity C . The considered turbo decoder consists of two optimum symbol-by-symbol decoders for the constituent codes. In the following, we assume LogAPP decoders [24]; equivalently, APP decoders can be employed [2], [25]. Each constituent decoder computes extrinsic LLRs based on both the extrinsic LLRs provided by the other constituent decoder and the LLRs for the systematic bits.

If we assume long interleavers and take into account that we are interested only in the average symbol-wise mutual information (as done in [14], [15]), we have four virtual channels which are BISMCS, each having a certain mutual information.

- 1) The channel between encoder inputs and channel LLRs of the systematic bits; the mutual information is denoted as systematic information I_s .
- 2) The channel between encoder inputs and extrinsic LLRs computed by the first constituent decoder; the mutual information is denoted as extrinsic information I_{e1} .
- 3) The channel between encoder inputs and extrinsic LLRs computed by the second constituent decoder; the mutual information is denoted as extrinsic information I_{e2} .
- 4) The channel between encoder inputs and post-decoding LLRs of the iterative decoder; the mutual information is denoted as overall information I .

EXIT charts and IPCs were proposed as tools for analysis and design of parallel-concatenated coding schemes. Although both methods are based on average symbol-wise mutual information, they address different aspects.

EXIT charts model the decoding behavior of the constituent decoders during iterative decoding, and they describe it by means of transfer characteristics. For the first constituent decoder, the transfer characteristic is the function mapping the average mutual information at the input I_{e2} (the extrinsic information provided by the other constituent decoder), to the average mutual information at the output I_{e1} (the extrinsic information computed by this constituent decoder), under the assumption⁵ that the input is distributed according to some fixed (commonly Gaussian) distribution. The capacity C of the physical channel is regarded as parameter. The transfer characteristic of the second constituent decoder is defined in an analogous way, where I_{e1} and I_{e2} are exchanged. The transfer characteristics for the Berrou/Glavieux turbo code [5] are depicted in Fig. 6.

Whereas EXIT charts describe properties of the constituent decoders, the IPC describes properties of the overall coding scheme. The IPC is defined as the function mapping the capacity C of the physical channel to the overall mutual information I between encoder input and symbol-by-symbol decoder output, i.e., in our case, the turbo decoder output.

The overall virtual channel is obviously a parallel concatenation of the “systematic” virtual channel and the two “extrinsic” virtual channels. Accordingly, the overall information I can be

⁵This assumption represents the modeling.

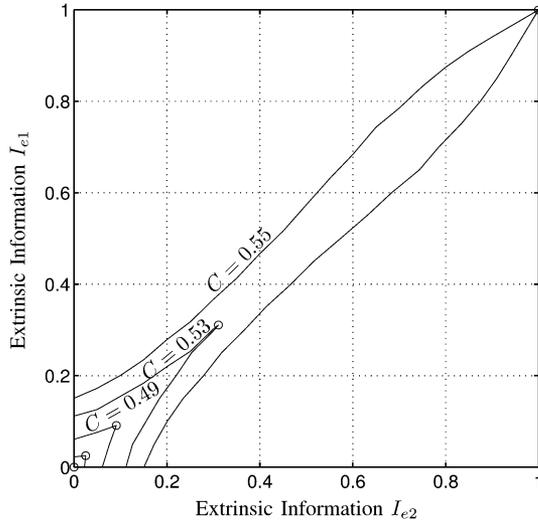


Fig. 6. Extrinsic information transfer (EXIT) chart for the constituent codes of the Berrou/Glavieux turbo code [5]. The extrinsic information I_{e1} is plotted versus the extrinsic information I_{e2} with the channel capacity C as parameter, and *vice versa*. Mutual information is given in bits per use.

computed by first combining I_{e1} and I_{e2} to the (overall) extrinsic information I_e , and then combining I_e and I_s to the overall information I . When computing the bounds according to Theorem 1 in each combining step, we get an upper and a lower bound on the IPC.

For the EXIT chart in Fig. 6, the resulting bounds on the IPC are depicted in Fig. 7. Although due to the recursive combining, the bounds may not be tight any more, they are very close to each other. Thus, it is not necessary to compute the exact value of the overall information, for which time-consuming simulation of the iterative decoder would be necessary. It is rather sufficient to compute the bounds based on the EXIT chart, for which simulation can be carried out very quickly.

Further information on this and similar methods can be found in [14], [26].

B. EXIT Charts for the Repetition Code and the Single Parity-Check Code

The EXIT chart method has not only shown to be a useful tool for analysis and design of parallel-concatenated codes, but also for serially concatenated codes [16]. The EXIT chart of an outer code of a serially concatenated coding scheme describes the input–output behavior of the outer decoder. The input is characterized by the symbol-wise *a priori* information⁶ on the code bits I_a , and the output is characterized by the average symbol-wise extrinsic information on the code bits

$$I_e = \frac{1}{N} \sum_{i=1}^N I_{ei}$$

where N denotes the code length. The EXIT characteristic is the function mapping I_a to I_e .

Using the theorems given in Section IV, the EXIT characteristics for the simple codes considered above can be computed in a straightforward manner. Due to symmetry, the extrinsic infor-

⁶This information is the same for all code bits, as the inputs of the decoder are assumed to be distributed according to some fixed distribution, cf. Section V-A.

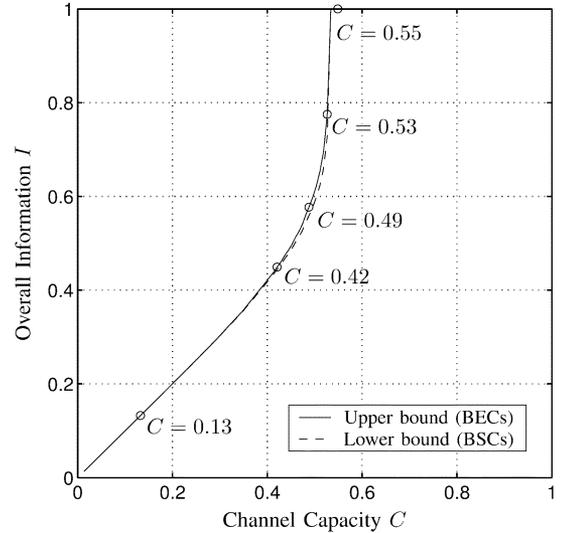


Fig. 7. Bounds on the IPC of the Berrou/Glavieux turbo code [5] based on the EXIT charts of the constituent codes. The overall information I is plotted versus the channel capacity C . Mutual information is given in bits per use.

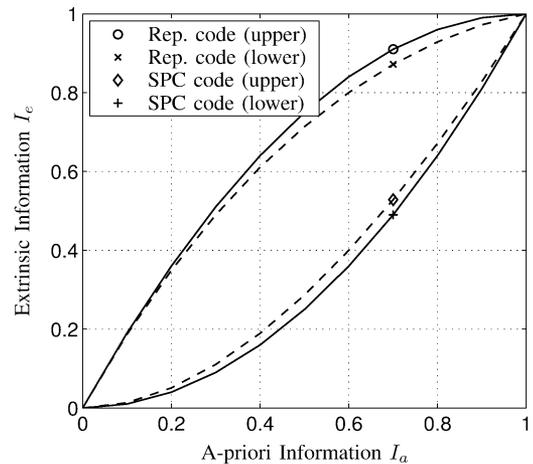


Fig. 8. Upper and lower bounds on the EXIT characteristics for the repetition (Rep.) code and the single parity-check (SPC) code of length three as outer codes in a serially concatenated coding scheme. The extrinsic information I_e is plotted versus the *a priori* information I_a . (Dashed lines: all channels are BSCs; solid lines: all channels are BECs.) Mutual information is given in bits per use.

mation is the same for all code bits X_1 , X_2 , and X_3 . The mutual informations of the channels are given by the *a priori* information

$$I_1 = I_2 = I_3 = I_a.$$

For the repetition code of length three, the extrinsic information can be bounded using Theorem 2

$$2I_a - f(I_a, I_a) \leq I_e \leq 2I_a - I_a^2. \quad (8)$$

For the single parity-check code of length three, the extrinsic information can be bounded using Theorem 3

$$I_a^2 \leq I_e \leq f(I_a, I_a). \quad (9)$$

These bounds on the EXIT characteristics are depicted in Fig. 8.

Since the repetition code and the single parity-check code are dual codes, their EXIT characteristics are symmetric with respect to the point $[I_a = 0.5, I_e = 0.5]$ if the channels are

BECs [27]. This corresponds to the upper bound for the repetition code and the lower bound for the single parity-check code in Fig. 8. This symmetry property does not hold for the bounds corresponding to BSCs, as can easily be seen by comparing the curves.

Although the codes are very simple, they illustrate how the concept of information combining may be used to obtain bounds on the extrinsic information, based only on the *a priori* information without further knowledge about the distribution of the *a priori* soft values.

VI. CONCLUSION

We presented tight upper and lower bounds on combined information which are based only on the mutual informations of the individual channels. Furthermore, we showed that these bounds are achieved by either the case that all channels are BSCs or the case that all channels are BECs.

In this paper, we restricted ourselves to three simple but very basic scenarios for information combining: a) the overall information for the parallel concatenation of two channels; b) the extrinsic information for the repetition code of length three; c) the extrinsic information for the single parity-check code of length three. Moreover, only BSMCs were taken into consideration.

Each scenario was illustrated by an example. This included the conversion of the EXIT chart of a parallel-concatenated code into its information-processing characteristic and the computation of the EXIT charts for (simple) outer codes of serially concatenated coding schemes.

The presented concept of bounding combined information may lead to new techniques for analyzing or proving convergence of iterative decoding schemes. This will provide new insights into processing and combining of (mutual) information in channel coding.

REFERENCES

- [1] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Trans. Info. Theory*, vol. 42, no. 2, pp. 429–445, Mar. 1996.
- [2] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 2, pp. 284–287, Mar. 1974.
- [3] C. R. Hartmann and L. D. Rudolph, "An optimum symbol-by-symbol decoding rule for linear codes," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 5, pp. 514–517, Sep. 1976.
- [4] G. Battail, M. C. Decouvelaere, and P. Godlewski, "Replication decoding," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 3, pp. 332–345, May 1979.
- [5] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1261–1271, Oct. 1996.
- [6] S. Benedetto and G. Montorsi, "Unveiling turbo codes: Some results on parallel concatenated coding schemes," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 409–428, Mar. 1996.
- [7] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenation of interleaved codes: Performance analysis, design, and iterative decoding," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 909–926, May 1998.
- [8] R. Gallager, "Low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. IT-8, no. 1, pp. 21–28, Jan. 1962.
- [9] D. J. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 399–431, Mar. 1999.
- [10] R. M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 5, pp. 533–547, Sep. 1981.
- [11] N. Wiberg, H.-A. Loeliger, and R. Koetter, "Codes and iterative decoding on general graphs," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Whistler, BC, Canada, Sep. 1995, p. 468.
- [12] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
- [13] S. Huettinger, J. Huber, R. Johannesson, and R. Fischer, "Information processing in soft-output decoding," in *Proc. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2001.
- [14] S. Huettinger, J. Huber, R. Fischer, and R. Johannesson, "Soft-output-decoding: Some aspects from information theory," in *Proc. Int. ITG Conf. Source and Channel Coding*, Berlin, Germany, Jan. 2002, pp. 81–90.
- [15] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.
- [16] —, "Code characteristic matching for iterative decoding of serially concatenated codes," *Ann. Télécommun.*, vol. 56, no. 7–8, pp. 394–408, 2001.
- [17] S. Huettinger and J. Huber, "Performance estimation for concatenated coding schemes," in *Proc. IEEE Information Theory Workshop*, Paris, France, Mar./Apr. 2003, pp. 123–126.
- [18] I. Land, S. Huettinger, P. A. Hoeher, and J. Huber, "Bounds on information combining," in *Proc. Int. Symp. Turbo Codes and Related Topics*, Brest, France, Sep. 2003, pp. 39–42.
- [19] I. Sutskov, S. Shamai (Shitz), and J. Ziv, "Extremes of information combining," in *Proc. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2003.
- [20] S. Huettinger, "Analysis and design of power-efficient coding schemes," Ph.D. dissertation, Univ. Erlangen-Nürnberg, Erlangen, Germany, 2004.
- [21] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [22] A. D. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications: Part I," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 6, pp. 769–772, Nov. 1973.
- [23] N. Chayat and S. Shamai (Shitz), "Expansion of an entropy property for binary input memoryless symmetric channels," *IEEE Trans. Inf. Theory*, vol. 35, no. 5, pp. 1077–1079, Sep. 1989.
- [24] P. Robertson, P. Hoeher, and E. Villebrun, "Optimal and sub-optimal maximum a posteriori algorithms suitable for turbo decoding," *Europ. Trans. Telecommun.*, vol. 8, no. 2, pp. 119–125, Mar./Apr. 1997.
- [25] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "A soft-input soft-output app module for iterative decoding of concatenated codes," *IEEE Commun. Lett.*, vol. 1, no. 1, pp. 22–24, Jan. 1997.
- [26] J. Huber and S. Huettinger, "Information processing and combining in channel coding," in *Proc. Int. Symp. Turbo Codes and Related Topics*, Brest, France, Sep. 2003, pp. 95–102.
- [27] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions, information functions, support weights, and duality," in *Proc. Int. Symp. Turbo Codes and Related Topics*, Brest, France, Sep. 2003, pp. 223–226.