Aalborg Universitet

AALBORG UNIVERSITY

DENMARK

Bounds on information combining

Land, Ingmar Rüdiger

Published in:
Proc. Int. Symp. on Turbo Codes & Rel. Topics, Brest, France, Sept., 2003.

Publication date:
2003

Document Version
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Bounds on Information Combining

Ingmar Land, Peter A. Hoeher

Information and Coding Theory Lab
University of Kiel, Germany
{il,ph}@tf.uni-kiel.de
www.tf.uni-kiel.de/ict

Simon Huettinger, Johannes Huber

Chair of Information Transmission
University Erlangen-Nürnberg, Germany
{huettinger,huber}@lnt.de
www.lnt.de/lit

Abstract: *When the same data sequence is transmitted over two independent channels, the overall mutual information represents a combination of the mutual information of the two channels — this concept is denoted as information combining. In this paper we present an upper and a lower bound on information combining and prove that these bounds cannot be further improved. Furthermore, we show how the concept and the bounds on information combining can be employed to link extrinsic information transfer (EXIT) charts and the information processing characteristic (IPC) of concatenated codes.*

Keywords: Mutual information, information combining, parallel concatenated codes, extrinsic information transfer (EXIT) chart, information processing characteristic (IPC).

1. Introduction

Extrinsic information transfer (EXIT) charts [1] and information processing characteristics (IPC) [2] have been proposed as tools for analysis and design of parallel concatenated codes (called turbo codes[1] [3]). Although both methods are based on average symbol-by-symbol mutual information, they address very different aspects.

EXIT charts describe the iterative decoding process of turbo codes by means of the transfer characteristic of the constituent decoders. For each constituent decoder, the transfer characteristic is defined as the function mapping the a-priori information to the extrinsic information, where the capacity C of the underlying channel is regarded as a parameter. As opposed to this, the information processing characteristic (IPC) [2] describes a property of the overall encoding/decoding scheme. The IPC is defined as the function mapping the capacity C of the underlying channel to the overall post-decoding mutual information between encoder input and symbol-by-symbol decoder output.

The overall information is obviously a combination of a-priori information and extrinsic information. This motivates to introduce the concept of information combining to link EXIT charts and IPCs. For doing so, we will first abstract the given system to get a simple, but sufficient model for analysis.

Since for long interleavers, a-priori information and extrinsic information can be assumed to be independent, they can be interpreted as the mutual information of two independent virtual channels, having the same input sequence. As these two channels are models for the end-to-end channel of fully interleaved transmission with linear binary constituent codes over symmetric physical channels, they can be assumed to be binary-input symmetric discrete memoryless channels (BISDMC). Examples for such channels are the binary symmetric channel (BSC), the binary erasure channel (BEC), and the binary-input additive white Gaussian noise channel.

Accordingly, we can use the following model: Consider the transmission of the same data sequence of independent and uniformly distributed binary symbols over two independent BISDMCs $X \rightarrow Y_1$ and $X \rightarrow Y_2$, denoted as constituent channels (CCh1 and CCh2). The overall channel, $X \rightarrow [Y_1, Y_2]$, is formed by the parallel concatenation of the constituent channels and thus denoted as parallel concatenated channel[2] (PCCh). It is easily seen that the PCCh is also a BISDMC.

Then the problem can be stated as follows: Given the mutual information (MI) of the two constituent channels, $I_1 := I(X; Y_1)$ and $I_2 := I(X; Y_2)$, what is the MI of the PCCh, $I := I(X; Y_1 Y_2)$. This "combining" of information I_1 and I_2 to the overall information I will be denoted as *information combining*[3].

In this paper, a tight upper and a tight lower bound on information combining will be stated and proven. Furthermore, it will be shown that the lower bound is achieved if the two constituent channels are BSCs, and that the upper bound is achieved if the two constituent channels are BECs. Since we can give examples achieving these bounds, they represent the fundamental limits of information combining.

This paper is structured as follows: In Section 2, some definitions for the single constituent channels will be given. In Section 3, the parallel concatenated channel is addressed, and fundamental bounds on information combining will be stated and proven. Finally, some applications will be outlined in Section 4.

---

[1] Linear binary constituent codes are assumed throughout this paper.

[2] The term "parallel concatenated channel" follows the term "parallel concatenated code".

[3] Even though the mutual information of a BISDMC for uniformly distributed input is equal to its capacity, we employed the term "mutual information" to emphasize that our focus is on "information" processing.

2. Constituent Channels

In this section, the constituent channels will be considered separately. Since both constituent channels are BISDMCs and thus have the same properties, we will label the variables with the generic subindex i, where $i = 1$ corresponds to CCh1 and $i = 2$ corresponds to CCh2.

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

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

for all $x \in \mathbb{X}$ and $y \in \mathbb{Y}_i$ without loss of generality.

The MI of a constituent channel is defined as

$$I_i := I(X; Y_i), \tag{1}$$

as already stated above.

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

$$J_i := |Y_i|.$$

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

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

(The extra treating 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 $X \rightarrow Y_i$ into strongly symmetric sub-channels $X \rightarrow Y_i|J_i = j$. These sub-channels are binary symmetric channels and occur with probability

$$q_i(j) := p_{J_i}(j),$$

$j \in \mathbb{J}_i$. Their conditional crossover probabilities $\epsilon_i(j)$ are defined as

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

Let $h(x) := -x \operatorname{ld} x - (1-x) \operatorname{ld}(1-x)$, $x \in [0,1]$, 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 MI of sub-channel j is given as

$$I_i(j) := I(X; Y|J_i = j) = 1 - h(\epsilon_i(j)). \tag{2}$$

As mentioned above, $j = 0$ was treated as a special case in the definitions. The actual sub-channel,

a BEC with erasure probability 1, was transformed into a BSC with crossover probability $\frac{1}{2}$. This transformation does not change the MI ($I_i(0) = 0$), but will simplify the following derivations, since thus all sub-channels are BSCs without exceptions.

Using the above definitions, the MI of a constituent channel can be written as the expected value of the MI of its sub-channels:

$$I_i = \mathop{\mathrm{E}}_{j \in \mathbb{J}_i} \{I_i(j)\}. \tag{3}$$

The separation of the constituent channels into binary symmetric sub-channels will be exploited in the following section.

3. Parallel Concatenated Channel

Consider now the parallel concatenated channel (PCCh) $X \rightarrow [Y_1 Y_2]$ composed of the two constituent channels $X \rightarrow Y_1$ (CCh1) and $X \rightarrow Y_2$ (CCh2), which are assumed to be BISDMCs. It can easily be seen that the PCCh is also a BISDMC.

In the following, it will be investigated how the MI of the PCCh,

$$I := I(X; Y_1 Y_2), \tag{4}$$

is related to the MI $I_1 = I(X; Y_1)$ and $I_2 = I(X; Y_2)$ of CCh1 and CCh2, respectively. First, I can be written as

$$\begin{aligned} I &= I(X; Y_1 Y_2) \\ &= I(X; Y_1) + I(X; Y_2) - I(Y_1; Y_2) \\ &= I_1 + I_2 - I(Y_1; Y_2). \end{aligned} \tag{5}$$

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 information I_1 and I_2.

Let consider two simple examples [2]. (These two cases will prove to be *the* fundamental limits of information combining.) If the two constituent channels are BECs, then

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

If the two constituent channels are BSCs, then

$$I(Y_1; Y_2) = 1 - h\Big([1 - \epsilon_1] \epsilon_2 + \epsilon_1 [1 - \epsilon_2] \Big),$$

where $\epsilon_1 = h^{-1}(1 - I_1)$ and $\epsilon_2 = h^{-1}(1 - I_2)$ are the crossover probabilities of CCh1 and CCh2, respectively. In both cases, the information defect is expressed solely by the MI of the constituent channels.

For the following discussion, it will be useful to introduce a function for the latter case.

---

[4]The term "information defect" follows the term "mass defect" used in nuclear physics.

Definition 1 (Information defect function)
The information defect function (IDF) $f(x_1, x_2)$ is defined as

$$f(x_1, x_2) := 1 - h\Big(\big[1 - h^{-1}(1-x_1)\big] \cdot h^{-1}(1-x_2)$$
$$+ h^{-1}(1-x_1) \cdot \big[1 - h^{-1}(1-x_2)\big] \Big)$$

for $x_1, x_2 \in [0,1]$.

Thus, for the second example (two BSCs) the information defect can 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 in a compact form.

Theorem 1 (Bounds on inform. combining)
Given the MI $I_1 = I(X; Y_1)$ and $I_2 = I(X; Y_2)$ of the constituent channels, the MI of the parallel concatenated channel, $I = I(X; Y_1, Y_2)$, is bounded as

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

Note that the lower bound corresponds to the case that both constituent channels are BSCs, and the upper bound corresponds to the case that both constituent channels are BECs. Thus, for both the upper and the lower bound, we have an example actually achieving this bound. Consequently, these two bounds cannot be further improved, and they represent the fundamental limits of information combining.

In Figure 1, the bounds for I are plotted versus I_1 with several values for I_2. Note that the two bounds are very close to each other. Thus, the MI of the constituent channels dominate the value of the combined information, rather than the actual structures of the constituent channels.

The above theorem will be proven in three steps. Firstly, the information defect for the general case will be written as an expected value which includes only the IDF. Secondly, two properties of the IDF will stated. Finally, these properties will be used to give upper and lower bounds for the information defect and thus of the MI of the PCCh.

Taking the sub-channel indicators J_1 and J_2 into account, the information defect can be written as

$$\begin{aligned} I(Y_1; Y_2) &= I(Y_1; Y_2 | J_1, J_2) \\ &= \mathop{\mathrm{E}}_{\substack{j_1 \in \mathbb{J}_1 \\ j_2 \in \mathbb{J}_2}} \{I(Y_1; Y_2 | J_1 = j_1, J_2 = j_2)\} \\ &= \mathop{\mathrm{E}}_{j_1 \in \mathbb{J}_1} \Big\{ \mathop{\mathrm{E}}_{j_2 \in \mathbb{J}_2} \{f(I_1(j_1), I_2(j_2))\} \Big\}. \quad (6) \end{aligned}$$

In the first line, it was used that J_1 does not contain information about Y_2, and that J_2 does not contain information about Y_1. In the last line, the statistical independence of J_1 and J_2 was exploited.

For an exact evaluation of (6), firstly the two constituent channels have to be separated into their binary symmetric sub-channels, and the corresponding
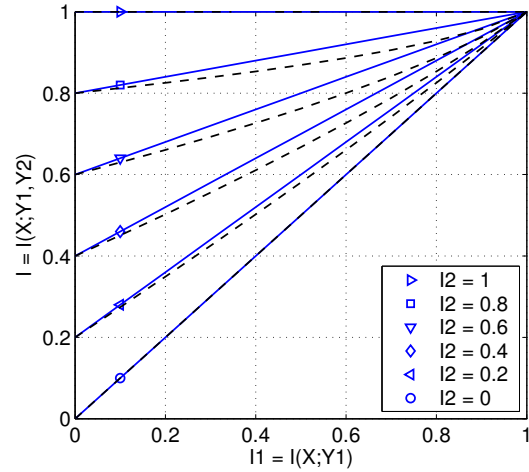

Figure 1: Bounds for combined information $I = I(X; Y_1 Y_2)$ (PCCh) vs $I_1 = I(X; Y_1)$ (CCh1) for several $I_2 = I(X; Y_2)$ (CCh2). The lower bounds (dashed lines) correspond to the case of two BSCs, the upper bounds (solid lines) correspond to the case of two BECs.

values of the incidence probabilities, $q_1(j)$ and $q_2(j)$, and the crossover probabilities, $\epsilon_1(j)$ and $\epsilon_2(j)$, have to be computed. Then, the MI of the sub-channels can be determined according to (2).

Let now consider two important properties of the IDF in

Lemma 2 (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:
(a) $f(x_1, x_2)$ is convex-$\cap$ in x_1 for constant x_2, and vice versa.
(b) $f(x_1, x_2)$ 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 define the function

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

$x \in [0,1]$, with parameter $a \in [0, \frac{1}{2}]$. 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. The function $g(x)$ is plotted vs x for several values of a in Figure 2. The formal proof of convexity is omitted due to the limited amount of space. But the plot clearly shows that $g(x)$ is convex-$\cup$ for all a.

(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 as

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

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. □
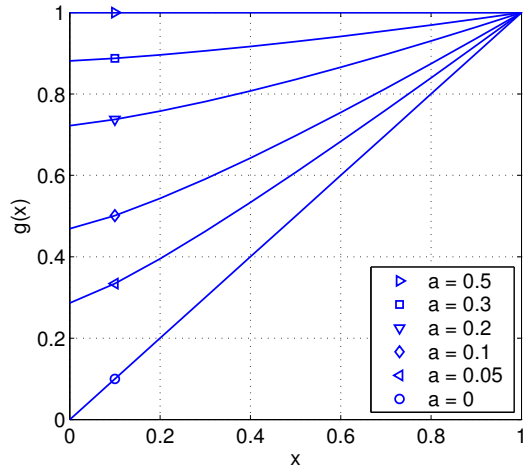

Figure 2: Function $g(x)$ vs x for several values of parameter a (cf. Lemma 2).

The results of this lemma will now be used to give bounds for 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 3 (Bounds on the information defect)
The information defect $I(Y_1; Y_2)$ is bounded as

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

Proof: As given in (6), the information defect can be written as

$$I(Y_1; Y_2) = \underset{j_1 \in \mathbb{J}_1}{\mathrm{E}} \Big\{ \underset{j_2 \in \mathbb{J}_2}{\mathrm{E}} \big\{ f\big(I_1(j_1), I_2(j_2)\big) \big\} \Big\}$$

Since the function $f(x_1, x_2)$ is concave in each dimension according to Lemma 2(a), Jensen's inequality [4] can be applied in the above expression, first w.r.t. j_2 and then w.r.t. j_1, and we have

$$\underset{j_1 \in \mathbb{J}_1}{\mathrm{E}} \Big\{ \underset{j_2 \in \mathbb{J}_2}{\mathrm{E}} \big\{ f\big(I_1(j_1), I_2(j_2)\big) \big\} \Big\} \leq$$
$$\leq f\big(\underset{j_1 \in \mathbb{J}_1}{\mathrm{E}} \{I_1(j_1)\}, \underset{j_2 \in \mathbb{J}_2}{\mathrm{E}} \{I_2(j_2)\} \big) = f(I_1, I_2),$$

where (3) was applied in the last equation. On the other hand, since the function $f(x_1, x_2)$ can be lower-bounded according to Lemma 2(b), the information

defect can be lower-bounded as

$$\underset{j_1 \in \mathbb{J}_1}{\mathrm{E}} \Big\{ \underset{j_2 \in \mathbb{J}_2}{\mathrm{E}} \big\{ f\big(I_1(j_1), I_2(j_2)\big) \big\} \Big\} \geq$$
$$\geq \underset{j_1 \in \mathbb{J}_1}{\mathrm{E}} \Big\{ \underset{j_2 \in \mathbb{J}_2}{\mathrm{E}} \big\{ I_1(j_1) \cdot I_2(j_2) \big\} \Big\} = I_1 \cdot I_2.$$

□

The proof of Theorem 1 follows immediately from (5) and Lemma 3.

4. Applications

There are several direct applications of information combining (see also [5]): (i) As outlined in the introduction, the IPC of a parallel concatenated code can be derived from its EXIT chart by combining the a-priori and the extrinsic information. (ii) The EXIT chart of the outer code of a serially concatenated code can be derived from its IPC by separating the a-priori information (intrinsic information) and the extrinsic information ("reversion of combining") [2]. (iii) Whereas the original EXIT chart method works only for the concatenation of two constituent codes, information combining allows to extend this method for the design of multiple turbo codes [6] and multiple serial concatenation via (i) and (ii).

In each of these applications, the bounds given in Theorem 1 can be applied to state both a pessimistic and an optimistic result. Thus, the accuracy can be precisely given. It turns out that the difference between the upper and the lower bound for information combining affects the results only to a minor degree. This strongly justifies these methods.

REFERENCES

[1] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.

[2] S. Huettinger, J. Huber, R. Fischer, and R. Johannesson, "Soft-output-decoding: Some aspects from information theory," in *Proc. Int. ITG Conf. on Source and Channel Coding*, Berlin, Germany, Jan. 2002, pp. 81–90.

[3] 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.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 1991.

[5] J. Huber and S. Huettinger, "Information processing and combining in channel coding," in *Proc. Int. Symp. on Turbo Codes & Rel. Topics*, Brest, France, Sept. 2003.

[6] S. Huettinger and J. Huber, "Design of "multiple–turbo–codes" with transfer characteristics of component codes," in *Proc. Conf. Inform. Sciences and Systems (CISS)*, Princeton University, Princeton, NJ, USA, Mar. 2002.