Bounds on information combining for parity-check equations

Land, Ingmar; Hoeher, A.; Huber, Johannes

Published in:
2004 International Seminar on Communications

DOI (link to publication from Publisher):
10.1109/IZS.2004.1287390

Publication date:
2004

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

Bounds on Information Combining for Parity-Check Equations

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

Johannes Huber
Chair of Information Transmission
University Erlangen-Nürnberg, Germany
huber@lnt.de
www.lnt.de/lit

***Abstract*— When several independent channels are coupled by a parity check constraint on their inputs, the mutual information between the input of one channel and the outputs of all other channels can be expressed as a combination of the mutual information between the input and the output of each individual channel. This concept is denoted as information combining. For binary-input symmetric discrete memoryless channels, we present bounds on the combined information which are only based on the mutual information of the channels. Furthermore, we show that these bounds cannot be further improved. Exact expressions are provided for the case that all channels are binary symmetric channels and for the case that all channels are binary erasure channels.**

I. Introduction

Consider coded transmission over one discrete memoryless channel or over multiple parallel discrete memoryless channels. Each noisy observation of a code bit contains information on this code bit. Furthermore, each observation also contains information on other code bits and information bits (or for short, info bits) due to the code constraints. Accordingly, the overall information on a certain code bit or info bit is a combination of the information on code bits. (In this paper, we will use the term "info bit" instead of "information bit" to avoid confusion with "mutual information".)

This combining of information under code constraints is used in every decoder, but the way code constraints are taken into account may differ. Whereas an APP decoder, e.g. [1], considers all constraints at once, iterative decoding algorithms consider only a subset of all code constraints in each decoding step. An iterative decoder for parallel and serially concatenated codes (turbo codes), e.g. [2], [3], [4], takes into account only the constraints of the constituent codes. An iterative decoder for low-density parity-check codes, e.g. [5], [6], or in general, any iterative decoder operating on graphs, e.g. [7], [8], [9], takes into account only local code constraints. Such a basic local constraint between code bits is given by a parity check equation.

If the information on the code bits is represented by probabilities, log-likelihood ratios, or similar measures, then the combined information on one code bit can easily be computed by an appropriate operation as given, e.g., in [9].

In this paper, "information combining" is used in a very strict sense, namely only for *combining of mutual information*. This notion of information combining was introduced in [10], [11] for design and analysis of parallel concatenated coding schemes. Since not all statistical properties of the channels are taken into account, but only their mutual information, only bounds on the combined information can be given.

In [12], bounds on information combining were presented for the case that a binary symbol is transmitted over two independent channels. It was shown that these bounds cannot be further improved, and that the lower bound corresponds to the case that both channels are binary symmetric channels (BSCs), and that the upper bound corresponds to the case that both channels are binary erasure channels (BECs).

Two independent channels having the same input can be interpreted as two independent channels with a parity-check constraint on their inputs. In this paper, this concept is generalized to the case of an arbitrary number of independent channels with their inputs fulfilling a parity-check equation; the inputs can be regarded as the code bits of a parity check code. This scenario can also be regarded as decoding on a local parity check constraint in a graph, as mentioned above. The channels under consideration are binary-input symmetric discrete memoryless channels (BISDMC). Bounds are presented for the mutual information between a certain code bit and the observations of the other code bits; this information is denoted as extrinsic information on this code bit.

Here is an outline of the paper: In Section II, some definitions and properties for BISDMCs are given. Section III addresses information combining "across" the parity check equation. Bounds for the general case and exact expressions for the only-BSC case and for the only-BEC case are given. Finally, conclusions are drawn in Section IV.

II. Binary-Input Discrete Memoryless Channels: Definitions and Properties

Let $X_i \rightarrow Y_i$ denote a binary-input symmetric discrete memoryless channel (BISDMC) with inputs $X_i \in \mathbb{X} := \{-1, +1\}$ and outputs $Y_i \in \mathbb{Y}_i \subset \mathbb{R}$, where $\mathbb{X}$ and $\mathbb{Y}_i$ denote the input and the output alphabet of the channel, respectively. The transition probabilities are given by $p_{Y_i|X_i}(y|x)$, 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

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

for all $x \in \mathbb{X}$ and $y \in \mathbb{Y}_i$ without significant loss of generality. The mutual information (MI) of the channel is defined as

$$I_i := I(X_i; Y_i). \tag{1}$$

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 partitioned into the subsets

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

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 $X_i \to Y_i$ into strongly symmetric sub-channels $X_i \to Y_i | J_i = j$; it is therefore denoted as sub-channel indicator. The sub-channels are binary symmetric channels. Their conditional crossover probabilities $\epsilon_i(j)$ are defined as

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

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

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

(Note that for $J_i = 0$, the BEC with zero MI is transformed into an equivalent BSC with zero MI, so that all sub-channels are BSCs.)

Using the above definitions, the MI of the 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 a BISDMC into sub-channels which are BSCs is exploited in the following.

III. Extrinsic Information for Parity Check Equations

Let $C_1, C_2, \ldots, C_N \in \{0,1\} =: \mathbb{F}_2$ denote N binary code bits, which are uniformly distributed and independent except for the parity-check constraint

$$C_1 \oplus C_2 \oplus \cdots \oplus C_N = 0. \tag{4}$$

The code bits are mapped to the channel inputs X_i according to the one-to-one mapping $0 \mapsto +1$, $1 \mapsto -1$. (For the sake of convenience, also X_i will be denoted as code bit when possible without causing ambiguity.) The code bits X_i are transmitted over independent BISDMCs $X_i \to Y_i$ having MI $I_i := I(X_i; Y_i)$, $i = 1, 2, \ldots, N$. Note that this includes also the special case that all code bits are transmitted over the same channel, as we consider memoryless channels. The relations are depicted in Figure 1.

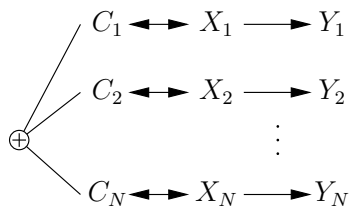

Fig. 1. N independent channels concatenated by a parity-check constraint on the inputs.

The notion of *extrinsic probability* plays an important role in the context of iterative decoding, see e.g. [2], [13]. The extrinsic probability of a bit is defined as the probability of this bit given all observations except for its direct observation. Similarly, an *extrinsic information* on a bit can be defined (as also done in [14]). In the following, the short hand notation $a_i^j := [a_i, a_{i+1}, \ldots, a_{j-1}, a_j]$ is used for subsequences.

Definition 1 *The extrinsic information I_{e1} on bit X_1 is defined as*

$$I_{e1} := I(X_1; Y_2^N),$$

i.e., all channel outputs except the direct observation Y_1 are taken into account. The extrinsic information for X_i, $i = 2, 3, \ldots, N$, is defined analogous.

The following class of functions will show to be useful.

Definition 2 *For $x_1, x_2, \ldots, x_n \in [0,1]$, $n \geq 2$, the function $f_2(x_1, x_2)$ is defined as*

$$f_2(x_1, x_2) := 1 - h\Big([1 - h^{-1}(1 - x_1)] \cdot h^{-1}(1 - x_2) + h^{-1}(1 - x_1) \cdot [1 - h^{-1}(1 - x_2)]\Big),$$

and the function $f_n(x_1, x_2, \ldots, x_n)$, $n > 2$, is recursively defined as

$$f_n(x_1, x_2, \ldots, x_n) := f_2\big(x_1, f_{n-1}(x_2, x_3, \ldots, x_n)\big).$$

An interpretation of these functions is as follows. Consider n BSCs $S_i \to R_i$, $S_i, R_i \in \{-1, +1\}$, having MI $I_i := I(S_i; R_i)$, $i = 1, 2, \ldots, n$. These BSCs are serially concatenated such that $R_i = S_{i+1}$ for $i = 1, 2, \ldots, n-1$. Then, the end-to-end MI is given as $I(S_1; R_n) = f_n(I_1, I_2, \ldots, I_n)$.

Using the above definitions, the main theorem of this paper can be stated as follows:

Theorem 1 (Bounds on extrinsic information) *If the channels $X_i \to Y_i$ are BISDMCs having MI $I_i := I(X_i; Y_i)$, $i = 2, 3, \ldots, N$, then the extrinsic information on code bit X_1, $I_{e1} := I(X_1; Y_2^N)$, is bounded as*

$$I_2 I_3 \cdots I_N \;\leq\; I_{e1} \;\leq\; f_{N-1}(I_2, I_3, \ldots, I_N).$$

The upper bound is achieved if all channels are BSCs, as will be shown in Theorem 2, and the lower bound is achieved if all channels are BECs, as will be shown in Theorem 3. Since we have examples achieving the bounds, these bounds cannot be further improved.

The bounds are illustrated for the case of $I_2 = I_3 = \cdots = I_N$ in Figure 2. Note that this corresponds to the case that the code bits are transmitted over channels which have the same MI but may differ in other statistical properties.

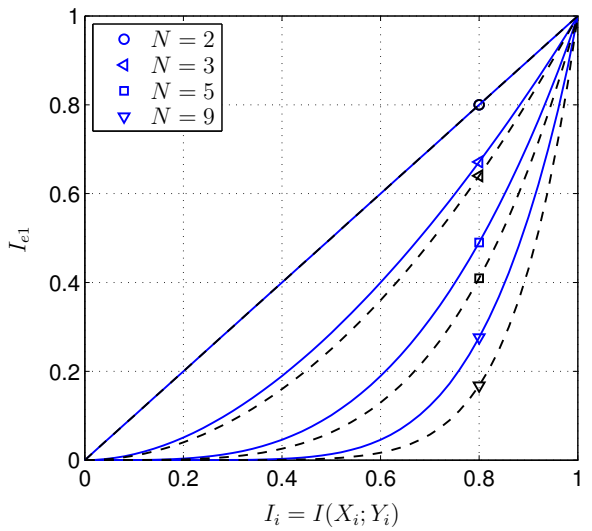

Fig. 2. Bounds on extrinsic information $I_{e1} = I(X_1; Y_2^N)$ vs $I_2 = I_3 = \cdots = I_N$. The lower bounds (dashed lines) correspond to the case of BECs, the upper bounds (solid lines) correspond to the case of BSCs.

In the following, first the extrinsic information for two special cases is computed: (a) all channels are BSCs; (b) all channels are BECs. Then, these results are used to prove the bounds for the general case, i.e., to prove Theorem 1.

A. Only Binary Symmetric Channels

If all channels $X_i \to Y_i$ are BSCs, then the extrinsic information can be expressed using function $f_n(x_1, x_2, \ldots, x_n)$ according to Definition 2.

Theorem 2 (BSC case) *If the channels $X_i \to Y_i$ are BSCs having MI $I_i := I(X_i; Y_i)$, $i = 2, 3, \ldots, N$, then the extrinsic information on code bit X_1, $I_{e1} := I(X_1; Y_2^N)$, is given as*

$$I_{e1} = f_{N-1}(I_2, I_3, \ldots, I_N).$$

Proof: Using the chain rule of mutual information, the extrinsic information can be written as

$$I(X_1; Y_2^N) = I(X_1; Y_2^{N-1}) + I(X_1; Y_N | Y_2^{N-1}).$$

Due to the independence of $X_1, X_2, \ldots, X_{N-1}$, we have $I(X_1; Y_2^{N-1}) = 0$.

Let binary random variables $Z_i \in \mathbb{F}_2$, $i = 1, 2, \ldots, N$, be defined as

$$\begin{aligned} C_1 &= Z_1, \\ Z_1 \oplus C_2 &= Z_2, \\ Z_2 \oplus C_3 &= Z_3, \\ &\cdots \\ Z_{N-2} \oplus C_{N-1} &= Z_{N-1}, \\ Z_{N-1} &= C_N, \\ Z_N &= Y_N. \end{aligned}$$

It follows from the definitions that all Z_i are uniformly distributed and that

$$I(X_1; Y_N | Y_2^{N-1}) = I(Z_1; Z_N | Y_2^{N-1}). \tag{5}$$

For the time being, assume $Y_2^{N-1} = y_2^{N-1}$, where y_2^{N-1} denotes an arbitrary but fixed realization of Y_2^{N-1}. Then, the random variables Z_i form a chain of BSCs,

$$Z_1 \to Z_2 \to Z_3 \to \cdots \to Z_{N-2} \to Z_{N-1} \to Z_N.$$

Consider now the MI of each BSC:
$Z_1 \to Z_2$: Since C_2 represents the error bit, the crossover probability is given by

$$\epsilon_2 := \min_{y_2'} \{ p_{C_2|Y_2}(1|y_2') \} = h^{-1}(1 - I_2).$$

Thus, the MI for this channel is simply given by $1 - h(\epsilon_2) = I_2$, which is independent from y_2.
$Z_i \to Z_{i+1}$, $i = 2, 3, \ldots, N-2$: Similarly to $Z_1 \to Z_2$, the MI is given by I_i, respectively.
$Z_{N-1} \to Z_N$: This channel is equal to the channel $X_N \to Y_N$, and thus the MI is I_N.
Using the interpretation of Definition 2, we immediately get

$$I(Z_1; Z_N | Y_2^{N-1} = y_2^{N-1}) = f_{N-1}(I_2, I_3, \ldots, I_N).$$

Due to the independence from y_2^{N-1}, we have

$$\begin{aligned} I(Z_1; Z_N | Y_2^{N-1}) &= \mathrm{E}\{ I(Z_1; Z_N | Y_2^{N-1} = y_2^{N-1}) \} \\ &= f_{N-1}(I_2, I_3, \ldots, I_N). \end{aligned}$$

Together with (5), this concludes the proof. ■

B. Only Binary Erasure Channels

In this subsection, it is assumed that all channels $X_i \to Y_i$ are BECs. In this case, the extrinsic information can be derived using a simple combinatorial approach.

Theorem 3 (BEC case) *If the channels $X_i \to Y_i$ are BECs having MI $I_i := I(X_i; Y_i)$, $i = 2, 3, \ldots, N$, then the extrinsic information on code bit X_1, $I_{e1} := I(X_1; Y_2^N)$, is given as*

$$I_{e1} = I_2 I_3 \cdots I_N.$$

Proof: Let $\delta_i = \Pr(Y_i = \Delta)$ denote the erasure probability of channel $X_i \to Y_i$. X can be recovered from Y_2^N only if no erasure occurred, due to the independence of X_i and the parity check constraint. This happens with probability $(1-\delta_2)(1-\delta_3)\cdots(1-\delta_N)$, and the corresponding MI is then equal to 1. In all other cases, the MI is equal to zero. Thus, we have

$$I(X_1; Y_2^N) = (1 - \delta_2)(1 - \delta_3) \cdots (1 - \delta_N).$$

Regarding $\delta_i = 1 - I_i$ concludes the proof. ■

C. General Symmetric Memoryless Channels

If the channels $X_i \to Y_i$ are only assumed to be BISDMCs without further restrictions, the extrinsic information cannot be computed exactly. Nevertheless, bounds can be given according to Theorem 1. In this section, the proof of this theorem is given.

First, two properties of the functions according to Definition 2 are needed.

Lemma 1 *The function $f_n(x_1, x_2, \ldots, x_n)$, $x_1, x_2, \ldots, x_n \in [0, 1]$, $n \geq 2$, has the following two properties:*
(a) $f_n(x_1, x_2, \ldots, x_n)$ is convex-$\cap$ in each x_i, $i = 1, 2, \ldots, n$;
(b) $f_n(x_1, x_2, \ldots, x_n)$ is not less than the product of its arguments:

$$f_n(x_1, x_2, \ldots, x_n) \geq x_1 x_2 \cdots x_n.$$

The proof follows immediately from Lemma 2 in [12] or Lemma 2 in [15].

Using the above lemma, Theorem 1 can be proved as follows.

Proof: The extrinsic information does not change if it is written conditional on the sub-channel indicators J_2^N as

$$\begin{aligned} I(X_1; Y_2^N) &= I(X_1; Y_2^N | J_2^N) \\ &= \mathop{\mathrm{E}}_{j_2^N} \{ I(X_1; Y_2^N | J_2^N = j_2^N) \} \\ &= \mathop{\mathrm{E}}_{j_2^N} \big\{ f_{N-1}\big(I_2(j_2), I_3(j_3), \ldots, I_N(j_N)\big) \big\}. \end{aligned} \quad (6)$$

The argument in the second line corresponds to the case, where all channels are BSCs, due to the conditions. Therefore, this expression can be replaced by the function $f_{N-1}(\ldots)$ according to Theorem 2.

In the next step, the two properties of function $f_{N-1}(\ldots)$ given in Lemma 1 are exploited. First, using the lower bound for this function in (6), we get

$$\begin{aligned} &\mathop{\mathrm{E}}_{j_2^N} \big\{ f_{N-1}\big(I_2(j_2), I_3(j_3), \ldots, I_N(j_N)\big) \big\} \geq \\ &\qquad \geq \mathop{\mathrm{E}}_{j_2^N} \big\{ I_2(j_2) I_3(j_3) \cdots I_N(j_N) \big\} = \\ &\qquad\qquad = I_2 I_3 \cdots I_N, \end{aligned}$$

where in the last line, (3) was used. Second, since the function $f_{N-1}(\ldots)$ is convex-$\cap$, Jensen's inequality can be applied in (6), and we get

$$\begin{aligned} &\mathop{\mathrm{E}}_{j_2^N} \big\{ f_{N-1}\big(I_2(j_2), I_3(j_3), \ldots, I_N(j_N)\big) \big\} \leq \\ &\quad \leq f_{N-1}\big(\mathop{\mathrm{E}}_{j_2}\{I_2(j_2)\}, \mathop{\mathrm{E}}_{j_3}\{I_3(j_3)\}, \ldots, \mathop{\mathrm{E}}_{j_N}\{I_N(j_N)\}\big) = \\ &\qquad = f_{N-1}(I_2, I_3, \ldots, I_N). \end{aligned}$$

■

IV. Conclusions

In this paper, binary-input symmetric discrete memoryless channels connected by a parity-check constraint on their inputs were considered. Bounds on the combined extrinsic information on one bit based on the mutual information of each channel were presented. For the cases that all BISDMCs are BSCs or BECs, respectively, the extrinsic information was computed exactly. Since these two cases correspond to the two bounds, the given bounds cannot be further improved.

In this paper, we focused on the case that the channels are coupled by a parity-check constraint on their inputs. But the applied techniques may be used or extended to yield similar results for other constraints. Of special interest is certainly the case that all channels have the same inputs, which may be denoted as equality constraint [12].

In [11], the concept of information combining and the corresponding bounds were applied to analysis and design of concatenated codes. Using the bounds for the parity-check constraint, presented in this paper, and the bounds for the equality constraint, mentioned above, the iterative decoder for low-density parity-check codes can be analyzed without the commonly used Gaussian approximation for the a-priori distribution.

References

[1] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, pp. 284–287, Mar. 1974.

[2] 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.

[3] S. Benedetto and G. Montorsi, "Unveiling turbo codes: Some results on parallel concatenated coding schemes," *IEEE Trans. Inform. Theory*, vol. 42, no. 2, pp. 409–428, Mar. 1996.

[4] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenation of interleaved codes: Performance analysis, design, and iterative decoding," *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 909–926, May 1998.

[5] R. Gallager, "Low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 8, no. 1, pp. 21–28, Jan. 1962.

[6] D. J. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 399–431, Mar. 1999.

[7] R. M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inform. Theory*, vol. 27, no. 5, pp. 533 –547, Sept. 1981.

[8] N. Wiberg, H.-A. Loeliger, and R. Koetter, "Codes and iterative decoding on general graphs," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Sept. 1995, p. 468.

[9] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 498 –519, Feb. 2001.

[10] S. Huettinger, J. Huber, R. Johannesson, and R. Fischer, "Information processing in soft-output decoding," in *Proc. Allerton Conf. on Communications, Control, and Computing*, Monticello, Illinois, USA, Oct. 2001.

[11] 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.

[12] I. Land, S. Huettinger, P. A. Hoeher, and J. Huber, "Bounds on information combining," in *Proc. Int. Symp. on Turbo Codes & Rel. Topics*, Brest, France, Sept. 2003, pp. 39–42.

[13] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Trans. Inform. Theory*, vol. 42, no. 2, pp. 429–445, Mar. 1996.

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

[15] A. D. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications: Part I," *IEEE Trans. Inform. Theory*, vol. 19, no. 6, pp. 769–772, Nov. 1973.