Information Combining

Land, Ingmar; Huber, Johannes

Published in:
Foundations and Trends in Communication and Information Theory

DOI (link to publication from Publisher):
10.1561/0100000013

Publication date:
2006

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

Foundations and Trends® in
Communications and Information Theory
Vol. 3, No 3 (November 2006) 227–330

DOI: 10.1561/0100000013

now
the essence of knowledge

Information Combining

Ingmar Land[1] and Johannes Huber[2]

[1] *Aalborg University, Denmark*
[2] *Erlangen University, Germany*

Abstract

Consider coded transmission over a binary-input symmetric memoryless channel. The channel decoder uses the noisy observations of the code symbols to reproduce the transmitted code symbols. Thus, it combines the information about individual code symbols to obtain an overall information about each code symbol, which may be the reproduced code symbol or its a-posteriori probability. This tutorial addresses the problem of "information combining" from an information-theory point of view: the decoder combines the mutual information between channel input symbols and channel output symbols (observations) to the mutual information between one transmitted symbol and all channel output symbols. The actual value of the combined information depends on the statistical structure of the channels. However, it can be upper and lower bounded for the assumed class of channels. This book first introduces the concept of mutual information profiles and revisits the well-known Jensen's inequality. Using these tools, the bounds on information combining are derived for single parity-check codes and for repetition codes. The application of the bounds is illustrated in four examples: information processing characteristics of coding schemes, including extrinsic information transfer (EXIT) functions; design of multiple turbo codes; bounds for the decoding threshold of low-density parity-check codes; EXIT function of the accumulator.

1

Introduction

In digital communications, the transmitter adds redundancy to the data to be transmitted, and the receiver exploits this redundancy to perform error correction. In this book, we restrict ourselves to binary linear channel codes and transmission over memoryless communication channels. The transmitter can thus be identified with the channel encoder and the receiver with the channel decoder. Because of the assumed channel model, the receiver obtains one noisy observation for each code symbol.

Each of these observations carries information about the corresponding code symbol at the channel input, of course. In addition to that, due to the code constraints that couple the code symbols, each observation also carries information about other code symbols. To exploit the redundancy in the code, the decoder combines all available information to estimate the value of each code symbol. In this chapter, the focus will be on optimal combining, i.e. combining such that all information about individual code symbols is retained.

This process of information combining can also be seen from an information theory point of view when the asymptotic case of codes of

infinite length[1] is considered. For each code symbol, there is a mutual information between the code symbol and the noisy observation. These values of mutual information are "combined" to obtain a value of the mutual information between a code symbol (or an information symbol) and *all* observations. The decoder is thus interpreted as a processor for mutual information. This is done in the information processing characteristic (IPC) method [1–3].

Some classes of channel codes, e.g., low-density parity-check (LDPC) codes [4, 5], are iteratively decoded: two constituent decoders exchange extrinsic values, called messages, until they agree on a certain estimated codeword, the maximum number of iterations is reached, or another stopping criterion is fulfilled. (The term "extrinsic" will be introduced later.) These constituent decoders can also be interpreted as processors for mutual information, in this case of extrinsic mutual information. This is done in the extrinsic information transfer (EXIT) chart method [6, 7]

The mutual information resulting from the combining operation can be computed exactly if exact models of the channels between the code symbols and the observations (or messages) are assumed to be known, as in the IPC method and the EXIT chart method. Thus, the combined mutual information depends on the "input" mutual information and the channel models. These models (e.g. the Gaussian noise model), however, do not apply exactly.

This chapter addresses a generalization of these ideas. The channels are only assumed to be symmetric and memoryless. Thus, the exact value of the combined mutual information cannot be determined, but an upper and a lower bound can be given. This is referred to as *bounds on information combining* [8,9]. These bounds depend then only on the values of the "input" mutual information but not on the specific channel model. This basic problem is interesting from a pure information-theory point of view. The results can, however, also be used to analyze coding schemes and iterative decoders; they can even be used to design codes for the whole class of memoryless symmetric channels [10–15]. A closer

[1] To be precise, ensembles of codes are considered and the code length tends to infinity.

look at these references as well as at references to similar or extended combining concepts are provided at the end of this chapter.

This book gives an introduction to the principles of information combining. The concept is described, the bounds for repetition codes and for single parity-check codes are proved, and some applications are provided. As we focus on the basic principles, we consider a binary symmetric source, binary linear channel codes, and binary-input symmetric memoryless channels.

Throughout this book, we use the following notation. Upper-case letters denote random variables, and lower-case letters denote realizations. Vectors and matrices are both written in boldface. The meaning of boldface upper-case letters becomes clear from the context.

1.1 Combining of Probabilities

To achieve very closely the information-theoretic performance bounds of digital communication systems, joint processing of information over long blocks of symbols is necessary. Within such blocks, information has to be combined in some sense, e.g., parity symbols are generated in a channel encoder by forming check sums over distinct subsets of the information symbols, which are fed into the encoder. For a linear block code $\mathcal{C}$ with length N of symbols taken from the binary field $\mathbb{F}_2 = \{0,1\}$, these check sums are specified by the rows of a $(N-K) \times N$ parity check matrix $\boldsymbol{H}$, where K denotes the number of dimensions of the linear subspace in $\mathbb{F}_2^N$ forming the code.

Consider a binary codeword $\boldsymbol{X} = (X_0, X_1, \ldots, X_{N-1})$ of length N that is generated from K binary information symbols; the information symbols are assumed to be independent and uniformly distributed. Each code symbol $X_i \in \{0,1\}$ is transmitted over a binary-input communication channel, which we assume to be symmetric, time-invariant, memoryless, and without feedback throughout this book. This binary input symmetric memoryless channel (BISMC) maps the input symbols X_i into output symbols Y_i taken from an M-ary set $\mathcal{Y} = \{0,1,\ldots,M-1\}$ in a random way according to the transition probabilities $\Pr(Y = j|X = x)$, see Fig. 1.1. A channel is said to be

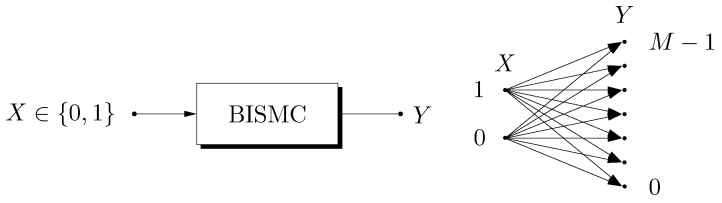

Fig. 1.1 Characterization of a BISMC by means of transition probabilities.

symmetric if it can be decomposed into strongly symmetric subchannels [16]; this is addressed in detail in Section 2.2.

If the a-priori probability $\Pr(X_i = 0)$ and the channel transition probabilities $\Pr(Y_i = y_i | X_i = x_i)$ are known, a-posteriori probabilities

$$\begin{aligned} p_i &:= \Pr(X_i = 0 | Y_i = y_i) \\ &= \frac{\Pr(X_i = 0)\Pr(Y_i = y_i | X_i = 0)}{\Pr(X_i = 0)\Pr(Y_i = y_i | X_i = 0) + (1 - \Pr(X_i = 0))\Pr(Y_i = y_i | X_i = 1)} \end{aligned} \tag{1.1}$$

are available after observing $Y_i = y_i$ for each individual code symbol. Usually, the vector $\boldsymbol{p} = (p_0, \ldots, p_{N-1})$ of these probabilities after transmission, but before decoding, is referred to as the *intrinsic probabilities for the code symbols* obtained from the communication channel [17].

Without any restriction of generality, we specify a probability on a binary variable $X \in \{0,1\}$ with respect to the value 0, i.e., $\Pr(X = 0|.)$ throughout the book. Of course, probability ratios $\Pr(X = 0|.)/(1 - \Pr(X = 0|.))$ or their logarithms, the so-called L-value $\ln\big(\Pr(X = 0|.)/(1 - \Pr(X = 0|.))\big)$ are synonymous to this notation, but in contrast to the mainstream in technical literature in the field of communications, we think that for theoretical derivations pure probabilities are more convenient than other types of probability specifications: A lot of nonlinear functions can be avoided, some equations are much more evident and easier to handle, and many readers may be more familiar with the language of basic probability theory than with specialized notation popular only in the coding and communications communities. Of course, for implementation of a decoder in hard- or software, probability ratios or, more pronounced, L-value notation may offer a lot of

advantages. But the intentions of this tutorial book are quite different; here, the development and understanding of the basic theory is the main focus.

Seen from a general point of view, values of information for individual symbols have to be combined in some way for exploiting the constraints within a sequence of symbols. *Information combining* happens in source encoding for extraction of redundancy from a source sequence or in channel decoding for improvement of data reliability. But there are many further fields where data processing essentially is some sort of information combining. To illustrate what we mean by information combining, we use the example of decoding a linear block code. Without loss of generality, the processing for code symbol X_0 will be further addressed in this example.

In a linear code, each parity check equation (e.g., Qth row of the parity check matrix $\boldsymbol{H}$) that includes X_0 provides further information on the code symbol X_0 by means of the other symbols X_{i_l} due to the check constraint

$$X_0 = X_{i_1} \oplus X_{i_2} \oplus X_{i_3} \oplus \cdots \oplus X_{i_L}. \tag{1.2}$$

Based on the intrinsic probabilities $p_i = \Pr(X_i = 0|y_i)$ of the residual symbols in a check sum, the *extrinsic probability* of code symbol X_0,

$$P_{\mathsf{ext},0} = \Pr(X_0 = 0|y_{i_1}, y_{i_2}, \ldots, y_{i_L}), \tag{1.3}$$

is computed. This probability on a code symbol is called *extrinsic* because it is calculated using only the channel outputs corresponding to the other code symbols but not the channel output corresponding to the symbol itself (see e.g. [18]).

In the case of a memoryless channel, the extrinsic probability $P_{\mathsf{ext},0}$ results in

$$P_{\mathsf{ext},0} = \frac{1}{2}\prod_{l=1}^{L}(2p_{i_l} - 1) + \frac{1}{2}. \tag{1.4}$$

(Remember that the codewords are assumed to be equiprobable.) This famous equation [4] can easily be derived from the case where only three

symbols are involved ($X_0 = 0$ if both symbols X_1 and X_2 are 0 or 1)

$$\begin{aligned} P_{\mathsf{ext}} &= \Pr(X_1 \oplus X_2 = 0 | y_1, y_2) \\ &= p_1 p_2 + (1 - p_1)(1 - p_2) \\ &= \frac{1}{2}(2p_1 - 1)(2p_2 - 1) + \frac{1}{2} \end{aligned} \tag{1.5}$$

and by induction from $L - 1$ to L. Notice that (1.4) also corresponds to the probability of observation of symbol 0 at the output of a chain (*series*) of L binary symmetric channels (BSCs) with crossover probabilities $\epsilon_i = 1 - p_i$ when symbol 0 is fed to its input, see Fig. 1.2. Therefore, we refer to (1.4) as the basic formula for *serial combining* of information.

Intrinsic and several extrinsic probabilities on a certain code symbol X are independent as long as the exploited check equations do not contain further code symbols in common and the channel is memoryless, as a memoryless channel acts independently on each of the code symbols. The task, to merge intrinsic and extrinsic probabilities on one symbol into a combined information is equivalent to the situation when a binary code symbol is transmitted over L parallel and independent channels or to the application of a repetition code of rate $1/L$ and transmission of the code symbols over a memoryless channel, see Fig. 1.3.

Thus, the second basic operation of information combining in channel decoding is to merge different, independent messages on individual code symbols and referring to Fig. 1.3, we denominate this operation as *parallel information combining*. Without loss of generality, a uniform a-priori distribution of X can be assumed because one of those "channels" may also be used to specify an a-priori probability on the variable X: a-priori knowledge is nothing else but a further independent source of extrinsic information. Basic probability calculation yields for two

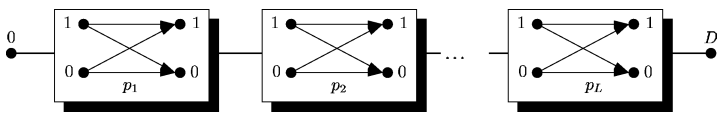

Fig. 1.2 Interpretation of Eq. (1.4) by a chain of BSCs with crossover probabilities $\epsilon_i = 1 - p_i$: serial information combining.

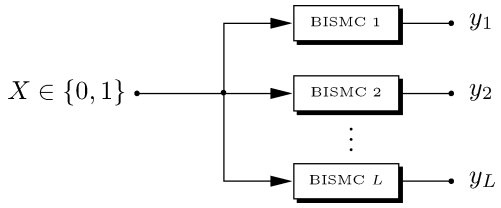

Fig. 1.3 Parallel information combining.

parallel channels (uniform a-priori distribution, cf. Fig. 1.3, too)

$$\begin{aligned} \Pr(X=0|y_1,y_2) &= \frac{p_1 p_2}{p_1 p_2 + (1-p_1)(1-p_2)} \\ &=: p_1 \otimes p_2 \,. \end{aligned} \tag{1.6}$$

In the same way, the corresponding result for L parallel channels is obtained:

$$\begin{aligned} \Pr(X=0|y_1,y_2,\ldots,y_L) &= p_1 \otimes p_2 \otimes \cdots \otimes p_L \,, \\ &= \frac{\prod_{l=1}^{L} p_l}{\prod_{l=1}^{L} p_l + \prod_{l=1}^{L}(1-p_l)} . \end{aligned} \tag{1.7}$$

Equation (1.6) is one of the reasons why probability ratios or L-values are very popular in this context: Combining independent a-posteriori probabilities on a binary symbol corresponds to the product of probability ratios or the sum of L-values, respectively. The binary operation "$\otimes$" induces an Abelian group $\mathbb{G} = \{\otimes, [0,1]\}$ onto the set $[0,1]$ of probabilities and by calculating the L-values, i.e., by the function $L : [0,1] \to \mathbb{R} : \ln(x/(1-x))$, an isomorphic mapping of the group $\mathbb{G}$ to $\{+, \mathbb{R}\}$ is established [19]. (Notice that for the basic combining equation (1.4) for check equations (serial information combining), such a nice accordance to L-values does not exist. Unfortunately, the corresponding formulas are rather involved when L-values are used, see (4.3).)

1.2 Combining of Mutual Information

The parallel and serial combination of probabilities on binary variables, i.e., Equations (1.4) and (1.6), are the basic operations for (iterative)

soft-decision decoding of linear binary codes. They also form the two key operations for iterative decoding of LDPC codes (details for LDPC codes are provided in Section 6.3). Therefore, we intend to analyze these basic information combining operations in a more general context, looking rather on averages than on individual channel actions and observations as it is usually done in information theory.

One of the key concepts in iterative decoding is the use of extrinsic probabilities (or extrinsic L-values). Correspondingly, the basic problem that we will address in the following sections is to find tight bounds on the mutual information $I(X_0;Y_1,\ldots,Y_{L-1})$ for the serial and parallel combination of information solely based on the mutual information $I(X_i;Y_i)$ provided by the channels for transmission of the individual symbols. Notice that this is an *extrinsic mutual information* (e.g., [6]) with respect to X_0 as it is the mutual information between the code symbol X_0 and the observations of only other code symbols; the direct observation of X_0 is omitted.

An introductory example is serial or parallel information combining for binary erasure channels (BECs) with erasure probabilities γ_i and capacities $I_i = 1 - \gamma_i$, cf. Fig. 1.4, which really is the simplest one.

The combination of L received symbols in a check equation leads to an erasure if at least one of the transmitted symbols is erased; otherwise, we get a surely correct extrinsic information. Therefore, the erasure probability of the combined channel reads $\gamma = 1 - \prod_{i=1}^{L}(1-\gamma_i)$, which is equivalent to the formula

$$I = \prod_{i=1}^{L} I_i \tag{1.8}$$

for serial information combining.

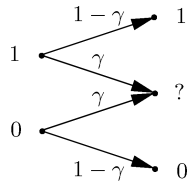

Fig. 1.4 Binary erasure channel (BEC) with erasure probability γ. The erasure is denoted by "?".

Transmission of binary symbol over L parallel BECs yields perfect knowledge at the receiver side if at least one of these channels does not deliver an erasure. Thus, the erasure probability of L parallel BECs is $\gamma = \prod_{i=1}^{L} \gamma_i$, and the overall mutual information (or capacity) reads

$$I = 1 - \prod_{i=1}^{L} (1 - I_i) \,. \tag{1.9}$$

Unfortunately, such explicit solutions do not exist in general, but we are able to derive rather tight bounds on information combining, if the individual binary input symmetric channels are only specified by their mutual information (or capacity).

1.3 Outline and Related Work

The bounds on information combining will enable us to analyze various properties of coding schemes and iterative decoding procedures in a very general way. "Mutual information" has proven to be a very useful and relevant measure to characterize a channel by a single parameter only. Correspondingly, applying it leads to easy tools to derive fairly tight performance bounds or to optimize coding schemes (e.g., the design of multiple turbo codes, see Section 6.2).

For that purpose, we will recapitulate the basic properties of BISMCs in Chapter 2 and define a new tool to fully specify channels of that type, called the mutual information profile (MIP) of a BISMC. In Chapter 3, Jensen's well-known formula is revisited and extended to a pair of inequalities, i.e., to a lower and an upper bound on the expectation of a real random variable after processing by a convex function; we will identify the probability density functions (pdfs) for real random variables for which those bounds are tight, irrespective of the actual convex function.

Equipped with these prearrangements, the central theorems of this book, i.e., bounds on mutual information for serial and parallel combination of information on binary variables, are derived in a straightforward way in Chapters 4 and 5. Chapter 6 is dedicated to examples and applications of information combining: information processing characteristic of coding schemes, design of multiple turbo codes, and bounds

on EXIT functions and bounds on thresholds for convergence of iterative decoding of LDPC codes, and EXIT functions for RA codes.

The problem of information combining for parallel channels has been addressed in [2, 20] for the first time. Here, the so-called information processing characteristic (IPC) for a coding scheme has been introduced, too, cf. Section 6.1. In [21] an example has been given on how to use an IPC and information combining for a coarse estimation of bit error probability (BEP) and BEP-curves for concatenated coding schemes. In [22, 23] the analysis and optimization of multiple turbo codes by means of information combining was proposed. Surprisingly, tight bounds on the combined extrinsic information from several constituent codes in a so-called extended serial setup decoder leads to an analysis of the iterative decoding process, which is as simple as EXIT charts for the concatenation of only two constituent codes.

A more rigorous mathematical background to information combining has been introduced in [8] by finding the proof that there are simple tight bounds on parallel information combining for the case of two channels. Initiated by that, the results were generalized and applied to code design by two groups. In [9,10,12,13], the proofs are explicitly based on the decomposition of symmetric channels into binary symmetric subchannels and the concept of mutual information profiles, which may give a more intuitive access to this subject. Furthermore, these authors address only the basic case of binary symmetric sources and channels without memory, and the optimization with respect to all channels involved. In [14, 15], the proofs are based on [24], which is a generalization of Mrs Gerber's lemma [25], and thus have a more abstract character. These authors also address the question of a uniform source with memory and the role of the symmetry of the source. Furthermore, they show that the optimization can also be done with respect to the individual channels involved.

The present book is mainly based on [13] and the slides to [26] where the material was presented in a way that emphasizes the tutorial aspect. This is the main focus of this book as well. Therefore, we will follow the approaches of the first research group mentioned above.

Even though the present book focuses on pure combining of mutual information, references to similar or extended concepts should be given in the following. Mutual information is probably one the most successfully applied parameter of a memoryless channel. However, such a channel can also be characterized by other parameters, of course, like the expectation of the conditional bit probabilities (expected "soft-bit"), the Bhattacharyya noise parameter, the mean-square error (MSE), see [27–31]. Instead of using only one parameter to describe a channel, two such parameters may be used, as considered in [28, 32]. Since more parameters may characterize a channel more precisely than a single parameter, the resulting bounds may be tighter.

2

Binary-Input Symmetric Memoryless Channel

The channel is often defined as the part of the communication system that cannot be changed or should not be changed. In this chapter, a certain class of channels is discussed, namely, channels that have a binary input, and that are symmetric and memoryless; in addition to that, it is assumed that the channel is time-invariant, and that there is no feedback from the output of the channel to its input. These channels can be decomposed into binary symmetric subchannels, and they can be characterized by their mutual information profile. These two concepts are discussed in the following. The basics about symmetric channels are shortly revised within the text; for more details, we refer the reader to standard textbooks, e.g. [16].

2.1 Binary-Input Memoryless Channels

A binary-input, time-invariant, memoryless channel (BISMC) is fully characterized by the input alphabet $\mathcal{X} = \{0, 1\}$, the output alphabet[1]

[1] The output alphabet may also be continuous-valued. In this case, the transition probabilities have to be replaced by transition probability density functions. The concepts of this chapter apply in the same way as for discrete output alphabets by considering $M \to \infty$.

$\mathcal{Y} = \{0, 1, \ldots, M-1\}$, and the conditional probabilities

$$\begin{aligned} k_{0,j} &= \Pr(Y = j | X = 0), \\ k_{1,j} &= \Pr(Y = j | X = 1), \end{aligned} \tag{2.1}$$

$j = 0, 1, \ldots, M-1$, for observing the output variable $Y \in \mathcal{Y}$ for given binary input variable $X \in \mathcal{X}$. These conditional probabilities are called the channel transition probabilities.

Usually, these conditional probabilities are collected into the $(2 \times M)$ channel transition matrix $\boldsymbol{K}$

$$\boldsymbol{K} = \begin{pmatrix} k_{0,0} & \ldots & k_{0,M-1} \\ k_{1,0} & \ldots & k_{1,M-1} \end{pmatrix}.$$

This matrix is a stochastic matrix with row sum 1. For simplicity, the subsequent derivations are restricted to time-invariant and memoryless channels; the generalization to other channels is possible.

2.2 Decomposition into Binary Symmetric Channels

A channel is called symmetric [16] if the columns of the channel matrix $\boldsymbol{K}$ can be reordered in a way such that the resulting channel matrix $\boldsymbol{K'}$ (after reordering) can be partitioned into N_a submatrices $\boldsymbol{K}_a$ of dimension $(2 \times M_a)$, $\sum_{a=1}^{N_a} M_a = M$,

$$\boldsymbol{K'} = \left(\underbrace{\boldsymbol{K}_1}_{\text{submatrix } 1} \middle| \underbrace{\boldsymbol{K}_2}_{\text{submatrix } 2} \middle| \cdots \middle| \underbrace{\boldsymbol{K}_{N_a}}_{\text{submatrix } N_a} \right), \tag{2.2}$$

where any submatrix $\boldsymbol{K}_a$ is uniform in rows and columns (double uniformity), i.e., any row (column) of $\boldsymbol{K}_a$ is found by permutation of the elements of the first row (column).

If all elements of the submatrix $\boldsymbol{K}_a$ are divided by the sum w_a of elements of an arbitrary row (all sums are equal because of uniformity!), this submatrix is a stochastic matrix again and corresponds to a so-called *strongly symmetric subchannel* [16]. Thus by definition, a symmetric channel is partitionable into N_a strongly symmetric subchannels $\boldsymbol{K}_a, a = 1, 2, \ldots, N_a$. The row sums w_a with

$$\sum_{a=1}^{N_a} w_a = 1 \tag{2.3}$$

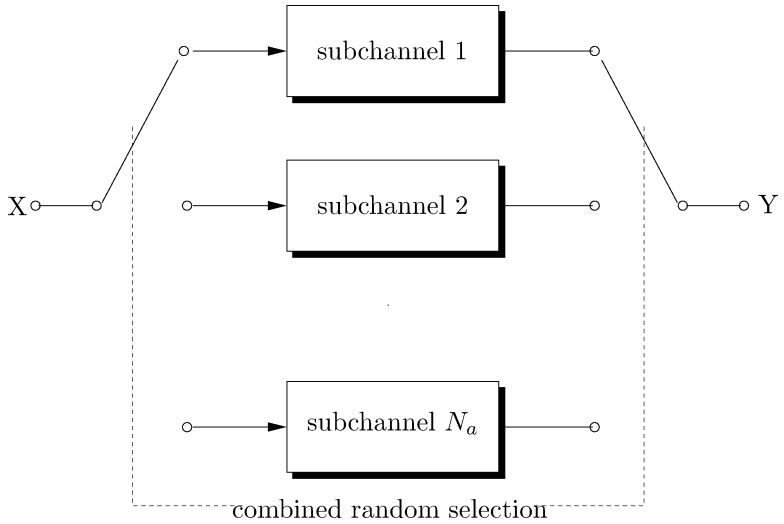

Fig. 2.1 Decomposition of a symmetric channel into strongly symmetric subchannels.

describe the probabilities for the random selection of the ath subchannel for mapping the input variable X into the output variable Y. Notice that the selection of the actually applied strongly symmetric subchannel is independent of the actual input symbol. The partitioning of a symmetric channel into strongly symmetric subchannels is illustrated in Fig. 2.1.

To identify the strongly symmetric subchannels of the channel transition matrix $\boldsymbol{K}$, the set $\mathcal{Y}$ of M output symbols is partitioned into disjoint subsets $\mathcal{Y}_a$, each subset corresponding to a specific subchannel $\boldsymbol{K}_a$. This index can be considered as a random variable A with realizations a, and it is referred to as the subchannel indicator [13].

Definition 2.1 The random variable A is called a *subchannel indicator* of a binary-input symmetric memoryless channel $\boldsymbol{K}$ if there is a one-to-one correspondence between the values a of A and the strongly symmetric subchannels $\boldsymbol{K}_a$ of $\boldsymbol{K}$ that cannot be further decomposed.[2]

---

[2] See also Lemma 2.2.

Resulting from this, A is a function of the channel output and it is statistically independent of the channel input.

According to the definition, we have the properties

$$I(X;A) = 0, \tag{2.4}$$

$$H(A|Y) = 0. \tag{2.5}$$

Thus, we obtain the relation

$$\begin{aligned} I(X;Y) &= I(X;Y,A) = I(X;A) + I(X;Y|A) \\ &= I(X;Y|A), \end{aligned} \tag{2.6}$$

where we first use that A is a function of Y, according to Definition 2.1, and then the chain rule for mutual information [33]. This relation shows that the mutual information between channel input and channel output does not change if it is conditioned on the subchannel indicator. This fact will often be exploited throughout this book.

In the following, we refer to a subchannel simply by the value of its subchannel indicator. Each subchannel $A = a$ has a certain value of mutual information, namely $I(X;Y|A = a) = I(X;Y_a)$, with $Y_a \in \mathcal{Y}_a \subset \mathcal{Y}$. The capacity of the channel is, of course, achieved by a uniform input distribution, as is the capacity of all of the subchannels. Thus, the capacity can be expressed by the average of the capacities C_a of the strongly symmetric subchannels $\boldsymbol{K}_a$:

$$C = \sum_{a=1}^{N_a} w_a C_a, \tag{2.7}$$

$$C_a = I(X;Y_a) \quad \text{for } \Pr(X = 0) = 1/2. \tag{2.8}$$

Notice that this is the same approach to compute the capacity of a symmetric channel as proposed by Gallager [16].

Even though the mutual information of a symmetric channel with uniformly distributed input is equal to its capacity, we will preferably use the term "mutual information," as the focus in this book is on "information combining"; accordingly, we will usually write "I" instead of "C." In the following, we will always assume uniformly distributed input symbols when not stated otherwise.

In the case of a BISMC, the situation is extremely simple: all subchannels are binary symmetric channels (BSCs). This is proved in the following theorem. Let the binary entropy function be defined as

$$h(x) := -x \log_2(x) - (1-x) \log_2(1-x), \tag{2.9}$$

$x \in (0,1)$, and $h(0) = h(1) := 0$. Let further $x = h^{-1}(y)$ denote the inverse of $y = h(x)$ for $0 \le x \le 0.5$. Since for a BSC, there is a one-to-one relation between the error probability and the mutual information for uniformly distributed input (or capacity), there is an error probability ϵ_a and a mutual information

$$I_a = 1 - h(\epsilon_a) \tag{2.10}$$

associated to each subchannel.

Theorem 2.1 Any BISMC can be decomposed into binary symmetric channels.

For the proof of Theorem 2.1, Lemma 2.2 is useful:

Lemma 2.2 A strongly symmetric binary-input channel, which is not further decomposable into strongly symmetric subchannels, at most has two output symbols.

Proof. In a double uniform stochastic $2 \times M$ matrix $\boldsymbol{S}$, there exist only two different types of columns, namely type I: $\binom{e}{f}$, and type II: $\binom{f}{e}$, where $e, f \in [0,1]$ denote probabilities. For example,

$$\boldsymbol{S} = \begin{pmatrix} e & f & f & e & f & f & \cdots \\ f & e & e & f & e & e & \cdots \end{pmatrix} \tag{2.11}$$

Type I II II I II II

In both rows, $n_{\mathrm{I}} e + n_{\mathrm{II}} f = n_{\mathrm{I}} f + n_{\mathrm{II}} e = 1$ holds, where n_{I} denotes the number of type I columns and n_{II} the number of type II columns. Therefore, we have

$$(n_{\mathrm{I}} - n_{\mathrm{II}}) \cdot (e - f) \equiv 0, \tag{2.12}$$

and thus one of the following two cases holds.

Case A: $e = f$. $\boldsymbol{S}$ is decomposable into $M(2 \times 1)$ (double uniform) matrices/vectors

$$\begin{pmatrix} e & e & e & \cdots \\ e & e & e & \cdots \end{pmatrix} = \left(\begin{pmatrix} e \\ e \end{pmatrix} \begin{pmatrix} e \\ e \end{pmatrix} \begin{pmatrix} e \\ e \end{pmatrix} \cdots \right). \tag{2.13}$$

For each subchannel, both input symbols are mapped to a single output symbol, which corresponds to a binary erasure channel (BEC) with erasure probability 1 and capacity 0. Notice that such a strict erasure channel is fully equivalent to a binary symmetric channel (BSC) with bit error probability $\epsilon = 0.5$.

Case B: $n_{\mathrm{I}} = n_{\mathrm{II}}$. By rearranging columns, we obtain

$$\begin{pmatrix} e & f & f & f & e & e & \cdots \\ f & e & e & e & f & f & \cdots \end{pmatrix} = \left(\begin{pmatrix} e & f \\ f & e \end{pmatrix} \begin{pmatrix} e & f \\ f & e \end{pmatrix} \begin{pmatrix} e & f \\ f & e \end{pmatrix} \cdots \right). \tag{2.14}$$

The strongly symmetric channel can be decomposed into $M/2$ identical BSCs with crossover probability $f/(e+f)$ and capacity $1 - h(f/(e+f))$. □

Proof of Theorem 2.1. This directly follows from Lemma 2.2, because columns with identical elements correspond to strict erasure channels, which are replaced by equivalent BSCs with crossover probability 0.5 by the introduction of further dummy output symbols. Columns with different elements can always be grouped in pairs such that each pair forms a BSC. □

For the discussions above, we focused on channels with a discrete output alphabet. However, these concepts are valid for channels with a continuous output alphabet in the same way. As an example, consider an AWGN channel with input $Z \in \{-1, +1\}$ and output $Y \in \mathbb{R}$ with BPSK mapping $X \mapsto Z$ at the input, $X \in \{0, 1\}$ as usual. The channel from X to Y is then a symmetric channel in the sense defined above. As Y is continuous, we have an infinite number of strongly symmetric subchannels with channel matrices

$$\boldsymbol{K}_a = \begin{pmatrix} p_{Y|X}(+a|0) & p_{Y|X}(-a|0) \\ p_{Y|X}(+a|1) & p_{Y|X}(-a|1) \end{pmatrix},$$

$a \in \mathbb{R}$, $a \geq 0$. The value $A = a$ is the subchannel indicator, which is continuous in this case. Notice that the subchannels are BSCs, as claimed in the theorem above. This binary-input AWGN channel will be further discussed in Section 2.4.3.

2.3 Mutual Information Profile

Each subchannel $A = a$ has a certain value of mutual information, namely $I(X;Y|A = a)$, where a uniform input distribution is assumed (cf. remark below (2.7)). Similar to the subchannel indicator, this value also can be regarded as a random variable. It is called the subchannel mutual information indicator and formally defined as follows [13]:

Definition 2.2 Consider a BISMC with input X, output Y, and the subchannel indicator A. The input symbols are assumed to be uniformly distributed. Using the mapping

$$f_J(a) := I(X;Y|A = a)$$

from a subchannel to its mutual information, the random variable

$$J := f_J(A)$$

is defined as the *mutual information indicator* of this channel. The probability distribution $p_J(j)$ of J is called the *mutual information profile (MIP)* of the channel.

For channels with a finite discrete output alphabet, the MIP reads

$$p_J(j) = \sum_{a=1}^{N_a} w_a \cdot \delta\big(j - f_J(a)\big).$$

The function $\delta(\cdot)$ denotes the Dirac δ-function.[3]

Using the mutual information indicator J and its distribution $p_J(j)$, the mutual information of a BISMC can be written as

$$I(X;Y) = \mathrm{E}\{J\} = \int p_J(j) \cdot j \; dj. \tag{2.15}$$

[3] The Dirac δ-function is defined via $\int_R \delta(z)dz = 1$ if the interval R includes the point 0, and $\int_R \delta(z)dz = 0$ otherwise.

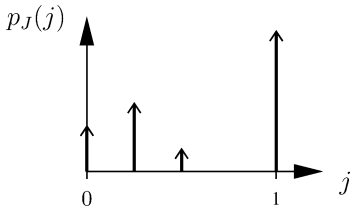

Fig. 2.2 Mutual information profile of a BISMC.

Thus, the mutual information is the first moment of the mutual information profile. An example for an MIP is depicted in Fig. 2.2. Notice that (2.15) corresponds to (2.7).

If the BISMC delivers a continuously distributed real output variable $Y \in \mathbb{R}$, e.g., a binary bipolar input additive white Gaussian noise channel (BIAWGN channel), an infinite number of subchannels may appear, and thus also an infinite number of values of the subchannel mutual information. Also in this case, the above definitions and the above theorem apply.

As a BISMC can be decomposed into BSCs according to Theorem 2.1, and as for a BSC, there is a one-to-one correspondence between the crossover probability and the mutual information[4], we have the following important result: *The mutual information profile uniquely characterizes a BISMC.* Of course, an MIP does not specify the alphabet, but this is also not relevant for the probabilistic characterization.

For $p_J(j) \neq 0$, j corresponds to a BSC with crossover probability $\epsilon = h^{-1}(1-j)$ and $p_J(j)dj$ to the differential probability of its selection. (The function $h^{-1}(\cdot)$ is defined in (2.9).)

2.4 Examples of Mutual Information Profiles

The concept of mutual information profiles (MIPs) is illustrated with some examples in the following. The mutual information of the channel is denoted by $I = I(X;Y)$. Remember that J is the random variable denoting the subchannel mutual information, and $I = \mathrm{E}\{J\}$ is its first moment (mean value). Thus, the mutual information profile $p_J(j)$

[4] As mentioned above, uniformly distributed input symbols are assumed.

indicates the distribution of the overall mutual information I over the values $J \in [0,1]$ of the subchannel mutual information.

2.4.1 Binary Symmetric Channel

The binary symmetric channel (BSC) with crossover probability ϵ consists only of one subchannel, of course; namely the one with mutual information

$$j = 1 - h(\epsilon) = I.$$

Accordingly, its mutual information profile is

$$p_J(j) = 1 \cdot \delta(1 - I). \tag{2.16}$$

This is depicted in Fig. 2.3. Obviously, for a given mutual information I, the BSC has the MIP with the smallest variance possible, as the mutual information is concentrated in a single point.

2.4.2 Binary Erasure Channel

The binary erasure channel (BEC) with erasure probability γ comprises two subchannels. The first one has subchannel mutual information $j_1 = 0$ and probability γ; it is a complete erasure channel. The second one has subchannel mutual information $j_2 = 1$ and probability $1 - \gamma$; it is an error-free channel. The mutual information profile is

$$p_J(j) = \gamma \cdot \delta(j) + (1 - \gamma) \cdot \delta(1 - j). \tag{2.17}$$

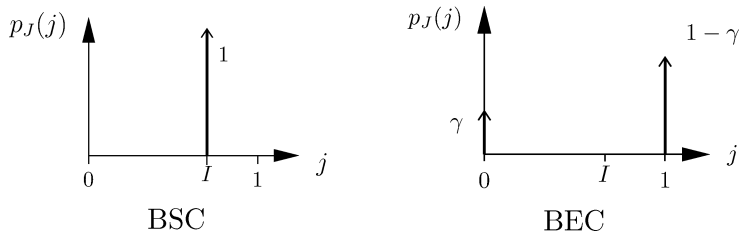

Fig. 2.3 Mutual information profiles (MIPs) of the binary symmetric channel (BSC) and the binary erasure channel (BEC).

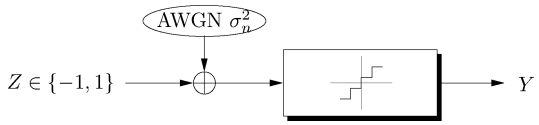

Fig. 2.4 BIAWGN channel with quantized output.

This is depicted in Fig. 2.3. Obviously, for a given mutual information I, the BEC has the MIP with the largest variance possible because the distribution is concentrated at the two ends of the interval $[0, 1]$ for j.

The interesting fact that the BSC and the BEC represent the extremes with respect to the variance of J will be of importance when looking for bounds in Chapters 4 and 5.

2.4.3 Binary-Input Additive White Gaussian Noise Channel

Consider now the BIAWGN channel with quantized output, see Fig. 2.4. The symbols $X \in \{0, 1\}$ are first BPSK-mapped to the symbols $Z \in \{-1, +1\}$. These are corrupted by white Gaussian noise N with variance σ_n^2. The output Y_c (label "c" for continuous) is then quantized to

$$Y = \mathrm{quantization}(Y_\mathrm{c}).$$

The signal-to-noise ratio (SNR) is specified by the usual quotient signal energy per channel symbol E_s to one-sided power spectral density N_0 of noise: $E_\mathrm{s}/N_0 = 1/(2\sigma_n^2)$.

In the following, different numbers of quantization levels are considered.

Binary Quantization

Assume the binary quantization

$$Y = \begin{cases} +1 & \text{for } Y_\mathrm{c} \geq 0, \\ -1 & \text{for } Y_\mathrm{c} < 0. \end{cases}$$

This transforms the channel into a BSC with crossover probability[5]

$$\epsilon = Q(1/\sigma_n),$$

cf. Section 2.4.1.

Ternary Quantization

Assume the ternary quantization

$$Y = \begin{cases} 0 & \text{for } r \leq Y_{\mathrm{c}}, \\ ? & \text{for } -r \leq Y_{\mathrm{c}} < r, \\ 1 & \text{for } Y_{\mathrm{c}} \leq -r. \end{cases}$$

for some nonnegative $r \in \mathbb{R}^+$. This transforms the channel into a binary symmetric error and erasure channel (BSEC) with erasure probability γ and crossover probability ϵ:

$$\gamma = Q((1-r)/\sigma_n) - Q((1+r)/\sigma_n), \tag{2.18}$$

$$\epsilon = Q((1+r)/\sigma_n). \tag{2.19}$$

This channel can be decomposed into two subchannels. The first subchannel is a complete erasure channel with mutual information $j_1 = 0$. The second subchannel is a BSC with error probability

$$\epsilon_2 = \frac{\epsilon}{1-\gamma}$$

and mutual information

$$j_2 = 1 - h\left(\frac{\epsilon}{1-\gamma}\right).$$

The mutual information profile is thus

$$p_J(j) = \gamma \cdot \delta(j) + (1-\gamma) \cdot \delta(j - j_2).$$

This decomposition and the mutual information profile are depicted in Fig. 2.5.

[5] The complementary Gaussian error integral is defined as $Q(x) := 1/\sqrt{2\pi} \cdot \int_x^\infty e^{-t^2/2} dt$.

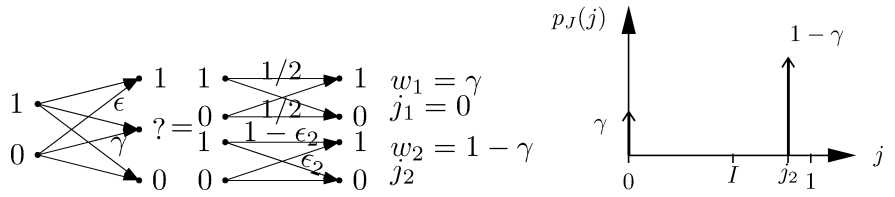

Fig. 2.5 Decomposition of a BIAWGN-channel with ternary quantization of the output variable, which is equivalent to a BSEC, into BSCs and the mutual information profile.

Quaternary Quantization

Assume the quaternary quantization

$$Y = \begin{cases} + & \text{for } r \le Y_c, \\ + & \text{for } 0 \le Y_c < r, \\ - & \text{for } -r \le Y_c < 0, \\ - & \text{for } Y_c \le -r. \end{cases}$$

for some positive $r \in \mathbb{R}$. This corresponds to a weak decision for $|Y| \le r$ and a strong decision for $|Y| > r$. The resulting channel can be decomposed into two BSCs, as depicted in Fig. 2.6. The probability of the first subchannel (weak decision) is

$$w_1 = Q\left(\frac{1-r}{\sigma_n}\right) - Q\left(\frac{1+r}{\sigma_n}\right),$$

and the corresponding probability for weak errors is

$$\epsilon_1 = Q(1/\sigma) - \epsilon_2.$$

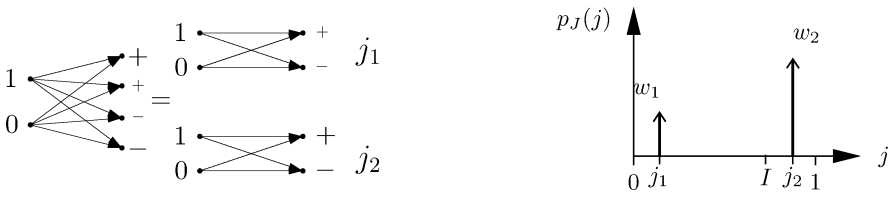

Fig. 2.6 Decomposition of a BIAWGN-channel with quaternary quantization of the output variable and the mutual information profile.

The probability of the second subchannel (strong decision) is $w_2 = 1 - w_1$, and the corresponding probability for strong errors is

$$\epsilon_2 = Q\Big(\frac{1+r}{\sigma_n}\Big).$$

The mutual information profile is thus

$$\begin{aligned} p_J(j) = w_1 \cdot \delta\left(j - \left(1 - h\left(\frac{\epsilon_1}{w_1}\right)\right)\right) \\ + (1-w_1) \cdot \delta\left(j - \left(1 - h\left(\frac{\epsilon_2}{1-w_1}\right)\right)\right), \end{aligned} \tag{2.20}$$

as shown in Fig. 2.6.

No Quantization

Consider now the case without quantization of Y_c, i.e., $Y \equiv Y_\mathrm{c}$. We first define the subchannel indicator

$$A := |Y|.$$

Thus the channel outputs $Y = a$ and $Y = -a$, $a \geq 0$, form the pair of binary output symbols for the binary symmetric subchannel for $A = a$.

Obviously, the probability density function of the continuous, real, and nonnegative subchannel indicator A is given by

$$p_A(a) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_n}\left(e^{-\frac{(a+1)^2}{2\sigma_n^2}} + e^{-\frac{(a-1)^2}{2\sigma_n^2}}\right) & \text{for } a \geq 0, \\ 0 & \text{for } a < 0. \end{cases} \tag{2.21}$$

Together with the crossover probability

$$\epsilon(a) = 1/(1 + e^{2a/\sigma_n^2}) \tag{2.22}$$

of the subchannel for $A = a$, MIP of the BIAWGN with equiprobable input variables is given by

$$p_J(j) = \frac{1}{\ln(2)} \frac{4\,a\,e^{2\,a/\sigma_n^2}}{\sigma_n^2\,(1 + e^{2\,a/\sigma_n^2})} \tag{2.23}$$

with

$$a = \frac{\sigma_n^2}{2} \ln(1/h^{-1}(1-j) - 1). \tag{2.24}$$

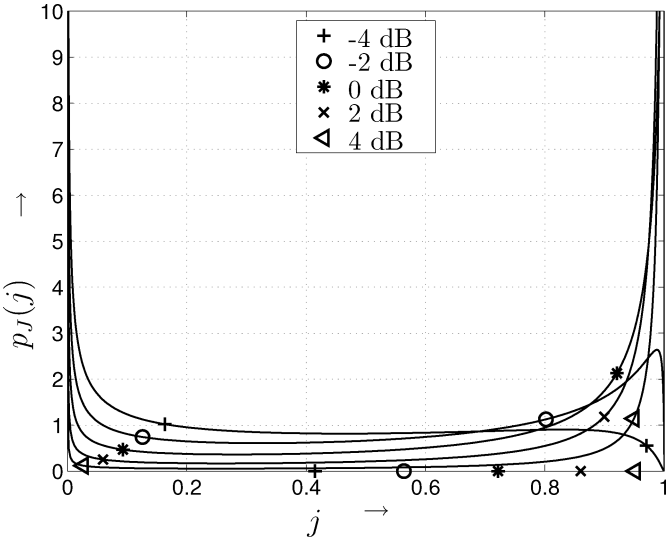

Fig. 2.7 Mutual information profiles of the BIAWGN (without quantization) at various SNRs.

Figure 2.7 shows the MIPs for the BIAWGN channel at different SNRs. Remember that the first moments of the MIP curves are the mutual information for equiprobable input symbols (or the channel capacity C). These values are marked by the corresponding symbols.

Notice that for SNR ≥ 1 (0 dB), the MIP starts to exhibit two poles, at $j = 0$ and $j = 1$, while the probability density for values in between decreases. Thus, for a moderately high SNR, i.e., $10\log_{10}(E_\mathrm{s}/N_0) \geq 4$ dB or $I \geq 0.95$, the MIP of the BIAWGN channel may be well approximated by two Dirac δ-functions at 0 and 1, i.e., by the MIP of a BEC of the same mutual information (capacity). This approximation may facilitate a lot the analysis of high-rate coding schemes operating at moderately high SNRs. For SNR $\to \infty$, a single peak at $j = 1$ remains, $p_J(j) = \delta(j - 1)$, and the BIAWGNC, the BEC, and the BSC coincide, of course.

3

Jensen's Inequality Revisited

In this chapter we extend Jensen's well-known inequality to a simple two-sided bound on the expected value of a convex function of a random variable. Both the upper and the lower bound are tight. First the theorem is revisited, and then the application to mutual information profiles is addressed.

Theorem 3.1 Let X be a real random variable defined on some interval $[A, B]$, $A < B$, with pdf $p_X(x)$, i.e., $p_X(x) = 0$ for $x \neq [A, B]$. Let $f(x)$ be a real-valued function $f : [A, B] \to \mathbb{R}$ which is[1] convex-$\cup$ in $[A, B]$. Then we have the following properties:

(a) The expectation of $f(X)$ is bounded by

$$f\left(\mathrm{E}\{X\}\right) \leq \mathrm{E}\left\{f(X)\right\} \leq f(A) + \frac{f(B) - f(A)}{B - A}\left(\mathrm{E}\{X\} - A\right).$$

(b) For any convex-$\cup$ function $f(x)$, the left inequality holds with equality if the "random variable" X is a constant,

[1] We follow Gallager's approach and use the terms "convex-$\cup$" and "convex-$\cap$" instead of "convex" and "concave" [16]. The interested reader may look for the corresponding foonote in Gallager's textbook.

i.e., if[2]

$$p_X(x) = \delta(x - \mathrm{E}\{X\}), \tag{3.1}$$

or equivalently, if $X \equiv \mathrm{E}\{X\}$.

(c) For any convex-$\cup$ function $f(x)$, the right inequality holds with equality if the random variable X is only distributed to the ends A and B of the interval, i.e., if

$$p_X(x) = \varepsilon \cdot \delta(x - A) + (1 - \varepsilon) \cdot \delta(x - B), \tag{3.2}$$

$\varepsilon \in [0, 1]$. Notice that $\mathrm{E}\{X\} = \varepsilon A + (1 - \varepsilon)B$.

The lower and the upper bound are tight as they are achieved by the distributions given in (3.1) and (3.2). Of course, this theorem holds in opposite direction for functions that are convex-$\cap$ (concave) in $[A, B]$.

Proof. Let $h_{\mathrm{L}}(x)$ and $h_{\mathrm{U}}(x)$ be two straight lines with

$$h_{\mathrm{L}}(x) \leq f(x) \leq h_{\mathrm{U}}(x), \quad \text{for all } x \in [A, B], \tag{3.3}$$

cf. Fig. 3.1.

Since $x \in [A, B]$, $\mathrm{E}\{h_{\mathrm{L}}(X)\} \leq \mathrm{E}\{f(X)\} \leq \mathrm{E}\{h_{\mathrm{U}}(X)\}$ holds. But for a straight line $h(x)$, $\mathrm{E}\{h(X)\} = h(\mathrm{E}\{X\})$ is true and we have

$$\mathrm{E}\{h_{\mathrm{L}}(X)\} = h_{\mathrm{L}}(\mathrm{E}\{X\}) \leq \mathrm{E}\{f(X)\}. \tag{3.4}$$

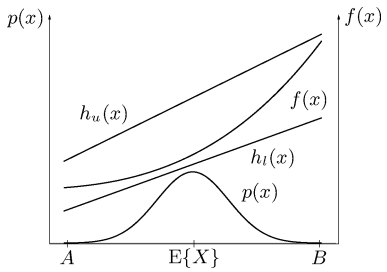

Fig. 3.1 pdf of a real random variable, a convex function, and bounding straight lines.

[2] The function $\delta(\cdot)$ denotes the Dirac δ-function, cf. Footnote 3 on p. 244.

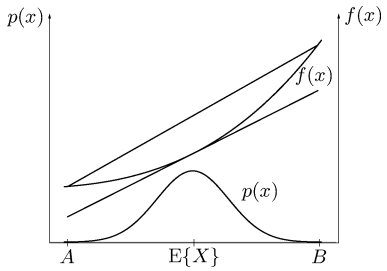

Fig. 3.2 pdf of a real random variable, a convex function, and tightest bounding straight lines.

From this inequality we find the tightest lower bound of this type by using for $h_{\mathrm{L}}(x)$ the tangent on $f(x)$ in $\mathrm{E}\{X\}$, i.e., $f(\mathrm{E}\{X\}) \leq \mathrm{E}\{f(X)\}$, which corresponds to the well-known Jensen's inequality, cf. Fig. 3.2. The tightness of this bound for concentration of $p_X(x)$ on $\mathrm{E}\{X\}$ immediately follows from Fig. 3.2, regardless of the function $f(x)$, as long as it is convex-$\cup$.

The tightest upper bound of this type follows from the lowest straight line $h_{\mathrm{U}}(x)$ which satisfies (3.3). Obviously, this is the secant through $(A, f(A))$ and $(B, f(B))$ to $f(x)$, see Fig. 3.2 and it is given by

$$h_{\mathrm{U}}(x) = f(A) + \frac{f(B) - f(A)}{B - A}\,(x - A)\,. \tag{3.5}$$

From this equation, the upper bound of Theorem 3.1 immediately follows. If the random variable X only exists in the two points A and B, where the convex function $f(x)$ and the straight line $h_{\mathrm{U}}(x)$ intersect, both expectations are equal $\mathrm{E}\{f(X)\} = \mathrm{E}\{h_{\mathrm{U}}(X)\}$, i.e., the right inequality is tight if X is distributed to the two values A and B only, irrespectively of the specific convex-$\cup$ function. □

Impact on Mutual Information Profiles

Theorem 3.1 gives rise to an interesting interpretation of mutual information profiles (MIPs) of BISMCs. The two pdfs for which these two-sided Jensen's inequalities are tight, correspond to the MIPs of the BSC (one-point distribution) and the BEC (two-point distribution at the boundaries). Thus, the BSC and the BEC may be regarded as the

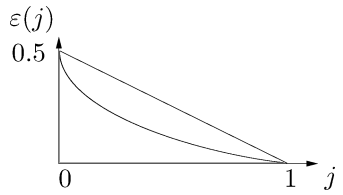

Fig. 3.3 Inverse binary entropy function and secant for upper bound.

extreme cases of a BISMC for a fixed capacity. To support these interpretation, we give a very simple example of an application of MIPs and these two-sided Jensen's inequalities: the upper bound and the lower bound on the bit error probability of a BISMC with fixed mutual information. As before, a uniformly distributed input is assumed such that this mutual information is equal to the channel capacity C.

Consider an arbitrary BISMC with capacity C. The channel output is given to a binary slicer to estimate the input variable of the channel. The BISMC is decomposable into BSCs with respect to a specific MIP $p_J(j)$, $\mathrm{E}\{J\} = C$. The bit error probability of the subchannel with mutual information $J = j$ is given by

$$\epsilon(j) = h^{-1}(1 - j).$$

Thus, the average bit error probability results as

$$\mathrm{BEP} = \mathrm{E}\Big\{h^{-1}(1 - J)\Big\} = \int_0^1 p_J(j)\, h^{-1}(1 - j)\, dj. \tag{3.6}$$

Notice that $h^{-1}(1 - j)$ is convex-$\cup$ in $j \in [0,1]$, see Fig. 3.3.

From Theorem 3.1, the following inequalities result:

$$h^{-1}(1 - C) \leq \mathrm{BEP} \leq \tfrac{1}{2}(1 - C). \tag{3.7}$$

On the left-hand side, we have the well-known Fano inequality, which is tight if the BISMC is a BSC. On the right-hand side, we have the inequality of Raviv and Hellman [34], which is tight if the BISMC is a BEC. (On average, half of the erased symbols are estimated correctly by the receiver by means of arbitrary choices.)

4

Information Combining for SPC Codes

This chapter describes the combining of mutual information for the case where the channel inputs are required to fulfill a parity-check constraint; i.e., the input symbols form a codeword of a single parity-check (SPC) code. To start with, the ideas outlined in the introduction are revised and confined to the addressed situation, and the notation is introduced. After describing the problem, first the two cases are considered where the channels are all BECs and where the channels are all BSCs. Then, this is generalized to arbitrary BISMCs. Based on this, the bounds on the combined information (the extrinsic information) are derived and discussed.

4.1 Problem Description

The problem of information combining for a single-parity-check (SPC) code is depicted in Fig. 4.1. The binary symbols $X_0, X_1, \ldots, X_{N-1} \in \mathbb{F}_2$ are required to fulfill the parity-check constraint

$$X_0 \oplus X_1 \oplus \cdots \oplus X_{N-1} = 0 \tag{4.1}$$

and thus form a codeword of length N. All codewords are assumed to have the same a-priori probability, and thus also each code symbol

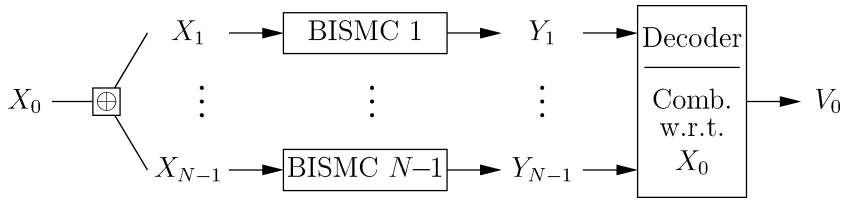

Fig. 4.1 Optimal information combining for a single-parity-check code.

is uniformly distributed. The code symbols $X_1, X_2, \ldots, X_{N-1}$ are transmitted over independent BISMCs. The channel output values $Y_1, Y_2, \ldots, Y_{N-1}$ are combined by a decoder to a value V_0. Notice that the observations $Y_1, Y_2, \ldots, Y_{N-1}$ contain information about X_0 due to the code constraints on the channel input symbols $X_0, X_1, \ldots, X_{N-1}$.

The value V_0 is an *extrinsic value* with respect to X_0 as it does not depend on a direct observation of X_0. As mentioned in Chapter 1, the motivation for looking at extrinsic values is twofold. On the one hand, extrinsic values are usually employed in iterative decoding schemes and thus are of high relevance and importance. On the other hand, the extrinsic value of a symbol (independent of the direct observation of this symbol) and the intrinsic value (dependent on only the direct observation) can be combined to obtain an a-posteriori value (depending on all observations), as also discussed in Chapter 1.

The combining operation to obtain the extrinsic value implies no loss of information if and only if

$$I(X_0; \boldsymbol{Y}_{[1,N-1]}) = I(X_0; V_0). \tag{4.2}$$

(The notation $\boldsymbol{Y}_{[1,N-1]} = [Y_1, Y_2, \ldots, Y_{N-1}]$ is used to denote partial vectors.) If this condition is fulfilled, the combining operation is called *optimal combining* with respect to X_0. In the following, we consider only optimal combining. Following the discussion from above, $I(X_0; \boldsymbol{Y}_{[1,N-1]})$ denotes an extrinsic mutual information, or for short, *extrinsic information* on code symbol X_0.

One optimal way of combining is the computation of the extrinsic probabilities according to Eq. (1.5). Notice that these extrinsic probabilities are special a-posteriori probabilities. Another optimal way of

combining is the computation of the extrinsic a-posteriori L-value

$$v_0 := \mathrm{L}(X_0|\boldsymbol{y}_{[1,N-1]}).$$

With $l_i := \mathrm{L}(X_i|y_i)$, the operation may be performed by [35]

$$v_0 = 2\tanh^{-1}\Big(\tanh\frac{l_1}{2}\cdot\tanh\frac{l_2}{2}\cdots\tanh\frac{l_{N-1}}{2}\Big). \tag{4.3}$$

Notice that v_0 is the extrinsic L-value for X_0.

Combining of Mutual Information

The decoder is now interpreted as a processor for mutual information. Each channel output Y_i conveys information about the corresponding channel input X_i,

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

$i = 1, 2, \ldots, N-1$. This information is used by the decoder to compute extrinsic information on X_0,

$$I_{\mathsf{ext},0} := I(X_0;V_0) = I(X_0;\boldsymbol{Y}_{[1,N-1]}).$$

(The last equality holds as optimal combining is assumed.) Thus, the decoder combines $I_1, I_2, \ldots, I_{N-1}$ to $I_{\mathsf{ext},0}$. This is referred to as combining of mutual information, or simply as *information combining*. The value I_i is referred to as the intrinsic information on code symbol X_i, and the value $I_{\mathsf{ext},0}$ is referred to as the extrinsic information on code symbol X_0 [17]. This nomenclature follows the one that is commonly used for probabilities and L-values.

The *question* is now as follows: For given values $I_1, I_2, \ldots, I_{N-1}$, what is the value of $I_{\mathsf{ext},0}$? The actual value of $I_{\mathsf{ext},0}$ turns out to depend on the distinct mutual information profiles of the individual channels. However, the maximum and the minimum value of $I_{\mathsf{ext},0}$ can be determined, and surprisingly, they are achieved if the individual channels are BECs or BSCs, respectively. These properties are proved using the concept of mutual information profiles (cf. Chapter 2) and the extension of Jensen's inequality (cf. Chapter 3).

4.2 Binary Erasure Channels

The case where the individual channels are all BECs was already discussed in detail in Chapter 1. For a BEC, the mutual information I and the erasure probability γ are related by $I = 1 - \gamma$. Using this and (1.8), the extrinsic information on X_0 and the values of intrinsic information are related by

$$I_{\mathsf{ext},0}^{\mathsf{BEC}} = I_1 \cdot I_2 \cdots I_{N-1}. \tag{4.4}$$

This equation represents the information combining for SPC codes when the channels are all BECs.

4.3 Binary Symmetric Channels

Consider now the case where the individual channels are all BSCs. The binary entropy function and its inverse are defined in (2.9). For a BSC, the mutual information I and the crossover probability ϵ are related by

$$I = 1 - h(\epsilon). \tag{4.5}$$

To determine analytically the mutual information on X_0, $I_{\mathsf{ext},0}$, the following approach is used [10, 13–15]: $I_{\mathsf{ext},0}$ is interpreted as the end-to-end mutual information of a serial concatenation of BSCs, where the mutual information of these BSCs are $I_1, I_2, \ldots, I_{N-1}$ (cf. Fig. 1.2). Thus, the channel $X_0 \to V_0$ is a BSC, too. Notice the similarity with the BEC-case mentioned earlier. The details of this approach are as follows.

The overall check equation can be split up into a series of small check equations, each comprising only three symbols. This is done by introducing "intermediate" symbols $Z_i \in \mathbb{F}_2$, $i = 1, \ldots, N-3$:

$$\begin{aligned}
Z_1 &:= X_0 \oplus X_1, \\
Z_2 &:= Z_1 \oplus X_2, \\
Z_3 &:= Z_2 \oplus X_3, \\
&\vdots \\
Z_{N-3} &:= Z_{N-4} \oplus X_{N-3}, \\
X_{N-1} &= Z_{N-3} \oplus X_{N-2}.
\end{aligned}$$

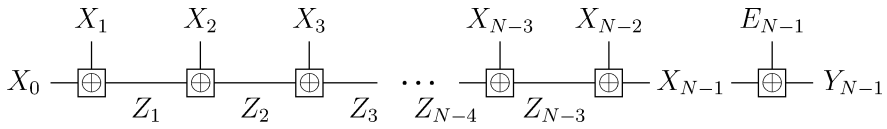

Fig. 4.2 Splitting up the parity-check equation leads to a serial concatenation of BSCs.

The last line is not a definition but "only" an equality; it results from the previous definitions and the parity-check equation (4.1). Because of their definitions, all Z_i are uniformly distributed. For the sake of a nice description, the error symbol

$$E_{N-1} := X_{N-1} \oplus Y_{N-1}$$

is introduced for the BSC $X_{N-1} \to Y_{N-1}$. The relations resulting from the equations above are depicted in Fig. 4.2.

Consider now the extrinsic information on code symbol X_0. Using the chain rule of mutual information [33], it can be written as

$$\begin{aligned} I_{\mathsf{ext},0} &= I(X_0; \boldsymbol{Y}_{[1,N-1]}) \\ &= I(X_0; \boldsymbol{Y}_{[1,N-2]}, Y_{N-1}) \\ &= \underbrace{I(X_0; \boldsymbol{Y}_{[1,N-2]})}_{=0} + I(X_0; Y_{N-1} | \boldsymbol{Y}_{[1,N-2]}). \end{aligned} \tag{4.6}$$

The first term is equal to zero, as X_0 and $\boldsymbol{Y}_{[1,N-2]}$ are independent if neither X_{N-1} nor Y_{N-1} are known.

For the time being, assume $\boldsymbol{Y}_{[1,N-2]} = \boldsymbol{y}_{[1,N-2]}$, where $\boldsymbol{y}_{[1,N-2]} \in \mathbb{F}_2^{N-2}$ denotes an arbitrary but fixed realization of $\boldsymbol{Y}_{[1,N-2]}$. Then, the random variables X_0, Z_i, $i = 1, \ldots, N-3$, X_{N-1}, Y_{N-1} form the Markov chain,

$$X_0 \to Z_1 \to Z_2 \to \cdots \to Z_{N-3} \to X_{N-1} \to Y_{N-1},$$

and each neighboring pair can be interpreted as a BSC (cf. Fig. 4.2). The mutual information of each BSC is as follows:

- $X_0 \to Z_1$: For this BSC, the code symbol X_1 represents the error symbol, and the probability of $\{X_1 = 1\}$ represents the crossover probability. For a given observation $Y_1 = y_1$,

the crossover probability of the channel $X_0 \to Z_1$ is

$$\begin{aligned}\epsilon_1 &\in \big\{p_{X_1|Y_1}(1|0), p_{X_1|Y_1}(1|1)\big\} \\ &= \big\{h^{-1}(1-I_1), 1-h^{-1}(1-I_1)\big\}\end{aligned}$$

and thus a function of the mutual information I_1 of the channel $X_1 \to Y_1$. Accordingly, the mutual information of the channel $X_0 \to Z_1$ is given by

$$I(X_0;Z_1) = 1 - h(\epsilon_1) = I_1.$$

Notice that both possible values of ϵ_1 lead to the same mutual information.

- $Z_i \to Z_{i+1}$, $i = 1, 2, \ldots, N-3$: Similar to $X_0 \to Z_1$, the mutual information is given by

$$I(Z_i;Z_{i+1}) = 1 - h(\epsilon_{i+1}) = I_{i+1}.$$

- $X_{N-1} \to Y_{N-1}$: By definition, the mutual information is given by

$$I(X_{N-1};Y_{N-1}) = I_{N-1}.$$

Notice that the mutual information of each BSC in this chain is independent of the actual value of $\boldsymbol{y}_{[1,N-2]}$.

According to these considerations, computing the mutual information

$$I(X_0;Y_{N-1}|\boldsymbol{Y}_{[1,N-2]} = \boldsymbol{y}_{[1,N-2]})$$

for some arbitrary but fixed $\boldsymbol{y}_{[1,N-2]} \in \mathbb{F}_2^{N-2}$ corresponds to computing the end-to-end mutual information of a chain of serially concatenated BSCs. This end-to-end mutual information depends only on the mutual information values of the individual channels, as shown in Appendix A. For convenience, the following function is introduced (cf. (1.4)).

Definition 4.1 Let $I_0, I_1, \ldots, I_{N-1} \in [0,1]$, $N \geq 1$. The functions

$$\begin{aligned} f_1^{\mathsf{ser}}(I_0) &:= I_0, \\ f_2^{\mathsf{ser}}(I_0, I_1) &:= 1 - h\big((1-\epsilon_0)\epsilon_1 + \epsilon_0(1-\epsilon_1)\big), \\ f_N^{\mathsf{ser}}(I_0, I_1, \ldots, I_{N-1}) &:= f_2^{\mathsf{ser}}\big(f_{N-1}^{\mathsf{ser}}(I_0, I_1, \ldots, I_{N-2}), I_{N-1}\big) \end{aligned}$$

with $\epsilon_0 := h^{-1}(1 - I_0)$ and $\epsilon_1 := h^{-1}(1 - I_1)$ are called *binary information functions for serial concatenation.*

The case $N = 1$ is included in the definition above to allow for a convenient and compact notation in the following.

The meaning of this function is as follows: Consider a serial concatenation of N BSCs, where the first BSC has a uniform input distribution. Let $I_0, I_1, \ldots, I_{N-1}$ denote the mutual information values of these BSCs. Then, the end-to-end mutual information of the serially concatenated BSC, i.e., the mutual information between the input of the first BSC and the output of the last BSC, is given by $f_N^{\mathsf{ser}}(I_0, I_1, \ldots, I_{N-1})$. (See also Appendix A.)

Using the function $f_N^{\mathsf{ser}}(\cdot)$ and the above interpretation as serially concatenated BSCs, we obtain

$$I(X_0; Y_{N-1} | \boldsymbol{Y}_{[1,N-2]} = \boldsymbol{y}_{[1,N-2]}) = f_{N-1}^{\mathsf{ser}}(I_1, I_2, \ldots, I_{N-1}).$$

This mutual information is independent of $\boldsymbol{y}_{[1,N-2]}$, as mentioned above. Thus the expected value with respect to $\boldsymbol{y}_{[1,N-2]}$ is given by

$$\begin{aligned} I(X_0; Y_{N-1} | \boldsymbol{Y}_{[1,N-2]}) &= \mathrm{E}\Big\{ f_{N-1}^{\mathsf{ser}}(I_1, I_2, \ldots, I_{N-1}) \Big\} \\ &= f_{N-1}^{\mathsf{ser}}(I_1, I_2, \ldots, I_{N-1}). \end{aligned} \tag{4.7}$$

Finally, using (4.7) in (4.6), we obtain

$$I_{\mathsf{ext},0}^{\mathsf{BSC}} = f_{N-1}^{\mathsf{ser}}(I_1, I_2, \ldots, I_{N-1}), \tag{4.8}$$

i.e., the extrinsic information on code symbol X_0 for the case where the channels are all BSCs.

4.4 Binary-Input Symmetric Memoryless Channels

Also for the case where the individual channels are all arbitrary BISMCs, an analytical expression for the combined information can be derived. To do so, the following concept is applied:

(a) the individual channels are decomposed into BSCs;
(b) the extrinsic information is computed for each set of BSCs;

(c) the average extrinsic information is obtained by averaging over the mutual information profiles of the individual channels.

Thus, the problem is solved by using the result for the all-BSC case and the concept of mutual information profiles.

All channels are assumed to be BISMCs. Therefore we may define a subchannel indicator A_i and a mutual information indicator J_i for each channel $X_i \to Y_i$ (cf. Chapter 2). Remember that the expectation of the mutual information indicator is equal to the mutual information of the channel, i.e., $\mathrm{E}\{J_i\} = I(X_i; Y_i) = I_i$.

The extrinsic information does not change if it is written conditioned on the subchannel indicators $\boldsymbol{A}_{[1,N-1]}$:

$$\begin{aligned} I_{\mathsf{ext},0} &= I(X_0; \boldsymbol{Y}_{[1,N-1]}) = I(X_0; \boldsymbol{Y}_{[1,N-1]}, \boldsymbol{A}_{[1,N-1]}) \\ &= \underbrace{I(X_0; \boldsymbol{A}_{[1,N-1]})}_{=0} + I(X_0; \boldsymbol{Y}_{[1,N-1]} | \boldsymbol{A}_{[1,N-1]}) \\ &= I(X_0; \boldsymbol{Y}_{[1,N-1]} | \boldsymbol{A}_{[1,N-1]}). \end{aligned}$$

Here we use (2.4) and the same approach as for deriving (2.6): $\boldsymbol{A}_{[1,N-1]}$ is independent of X_0 by Definition 2.1. For a given realization $\boldsymbol{a}_{[1,N-1]}$ of subchannel indicators, we have the case where the channels are all BSCs, and we can apply the function $f^{\mathsf{ser}}_{N-1}(\cdot)$ according to Definition 4.1, as done in the BSC case:

$$I(X_0; \boldsymbol{Y}_{[1,N-1]} | \boldsymbol{A}_{[1,N-1]} = \boldsymbol{a}_{[1,N-1]}) = f^{\mathsf{ser}}_{N-1}(j_1, j_2, \ldots, j_{N-1}).$$

Notice that $j_i = I(X_i; Y_i | A_i = a_i)$ is the mutual information corresponding to the subchannel $A_i = a_i$ of channel $X_i \to Y_i$.

Taking the expectation, we obtain the desired closed form expression:

$$\begin{aligned} I_{\mathsf{ext},0} &= I(X_0; \boldsymbol{Y}_{[1,N-1]}) \\ &= \mathrm{E}\Big\{ f^{\mathsf{ser}}_{N-1}(J_1, J_2, \ldots, J_{N-1}) \Big\}. \end{aligned} \tag{4.9}$$

For continuous-valued mutual information profiles $p_{J_i}(j_i)$, this expectation is evaluated as

$$I_{\mathsf{ext},0} = \int_{j_1} \cdots \int_{j_{N-1}} \underbrace{p_{J_1}(j_1) \cdots p_{J_{N-1}}(j_{N-1})}_{\text{mutual information profiles}} \cdot \cdot \underbrace{f^{\mathsf{ser}}_{N-1}(j_1, \ldots, j_{N-1})}_{\text{information combining for BSCs}} \, dj_1 \cdots dj_{N-1}. \tag{4.10}$$

Thus, the exact value of the combined mutual information can be computed via the mutual information profiles of the individual channels, i.e., the mutual information profiles are sufficient to characterize the decoding operation from an information-theory point of view.

4.5 Information Bounds

The previous three sections discussed the cases where the individual channels are known, i.e., where the complete mutual information profiles of the channels are known. In this section, it is only assumed that the individual channels are BISMCs and that their mutual information values are known, i.e., that the mean values of their mutual information profiles are known. Based on this knowledge only, tight bounds on the combined information can be given. These bounds are achieved by the cases where the individual channels are all BECs or all BSCs [10,12–15]. Notice that these cases correspond to the minimum-variance and to the maximum-variance mutual information profiles, cf. Sections 2.4.1 and 2.4.2 [26, 36].

Theorem 4.1 Let $X_0, X_1, \ldots, X_{N-1} \in \mathbb{F}_2$ denote the code symbols of a single parity check code of length N. Let $X_i \to Y_i$, $i = 1, 2, \ldots, N-1$, denote $N-1$ independent BISMCs having mutual information $I_i := I(X_i; Y_i)$. Let the extrinsic information on code symbol X_0 be defined by $I_{\mathsf{ext},0} := I(X_0; \boldsymbol{Y}_{[1,N-1]})$. Then, the following tight bounds hold:

$$I_{\mathsf{ext},0} \geq I_1 I_2 \cdots I_{N-1},$$
$$I_{\mathsf{ext},0} \leq f^{\mathsf{ser}}_{N-1}(I_1, I_2, \ldots, I_{N-1}).$$

The lower bound is achieved if the channels are all BECs, and the upper bound is achieved if the channels are all BSCs.

To prove this theorem, we make use of the following property:

Lemma 4.2 The binary information functions for serial concatenation, $f_N^{\mathsf{ser}}(I_0, I_1, \ldots, I_{N-1})$ (see Definition 4.1) are convex-$\cap$ in each argument I_i, $i = 0, 1, \ldots, N-1$.

Proof. The case $N = 1$ is trivial. The proof for $N = 2$, i.e., for $f_2^{\mathsf{ser}}(\cdot)$, is provided in Appendix B. The generalization for $N > 2$ follows from the recursive definition of $f_N^{\mathsf{ser}}(\cdot)$ in Definition 4.1. Since

$$f_N^{\mathsf{ser}}(I_0, I_1, \ldots, I_{N-1}) = f_2^{\mathsf{ser}}\big(f_{N-1}^{\mathsf{ser}}(I_0, I_1, \ldots, I_{N-2}), I_{N-1}\big)$$

and $f_2^{\mathsf{ser}}(\cdot)$ is convex-$\cap$ in its second argument, $f_N^{\mathsf{ser}}(\cdot)$ is convex-$\cap$ in its last argument. As $f_N^{\mathsf{ser}}(\cdot)$ is symmetric in all its arguments[1], this reasoning holds for all arguments, and we have the proof. □

Lemma 4.2 is now used to prove Theorem 4.1:

Proof. According to (4.9), the combined information can be written as

$$I_{\mathsf{ext},0} = \mathrm{E}\Big\{ f_{N-1}^{\mathsf{ser}}(J_1, J_2, \ldots, J_{N-1}) \Big\}.$$

The function $f_{N-1}^{\mathsf{ser}}(\cdot)$ is convex-$\cap$, and therefore, the extension of Jensen's inequality, Theorem 3.1, can be applied. Notice that Theorem 3.1 is formulated for convex-$\cup$ functions, whereas the given function is convex-$\cap$, such that the lower and the upper bound from Theorem 3.1 have to be swapped. Furthermore, Theorem 3.1 addresses only a function with one argument, whereas the given function has multiple arguments. Therefore, we have to apply this theorem successively with respect to each argument. Notice that the mutual information indicators J_i are confined to the interval $[0, 1]$.

Consider first the expectation with respect to J_1 and let the other mutual information indicators be fixed, $J_i = j_i$, $i = 2, \ldots, N-1$. Then, the upper bound results as

$$\begin{aligned} \mathrm{E}\Big\{ f_{N-1}^{\mathsf{ser}}(J_1, j_2, \ldots, j_{N-1}) \Big\} &\leq f_{N-1}^{\mathsf{ser}}\Big(\mathrm{E}\{J_1\}, j_2, \ldots, j_{N-1})\Big) \\ &= f_{N-1}^{\mathsf{ser}}(I_1, j_2, \ldots, j_{N-1}). \end{aligned}$$

[1] The symmetry of $f_N^{\mathsf{ser}}(\cdot)$ in its arguments can immediately be seen when looking at its interpretation.

The lower bound results as

$$\begin{aligned}
&\mathrm{E}\Big\{ f_{N-1}^{\mathsf{ser}}(J_1, j_2, \ldots, j_{N-1}) \Big\} \\
&\quad \geq f_{N-1}^{\mathsf{ser}}(0, j_2, \ldots, j_{N-1}) \\
&\qquad + \frac{f_{N-1}^{\mathsf{ser}}(1, j_2, \ldots, j_{N-1}) - f_{N-1}^{\mathsf{ser}}(0, j_2, \ldots, j_{N-1})}{1-0} \cdot \big(\mathrm{E}\{J_1\} - 0\big) \\
&\quad = f_{N-2}^{\mathsf{ser}}(j_2, \ldots, j_{N-1}) \cdot I_1.
\end{aligned}$$

It was used that

$$\begin{aligned}
f_{N-1}^{\mathsf{ser}}(0, j_2, \ldots, j_{N-1}) &= 0, \\
f_{N-1}^{\mathsf{ser}}(1, j_2, \ldots, j_{N-1}) &= f_{N-2}^{\mathsf{ser}}(j_2, \ldots, j_{N-1})
\end{aligned}$$

(cf. interpretation of this function), and $\mathrm{E}\{J_1\} = I_1$. Continuing in this way for $J_2, \ldots, J_{N-1}$ proves the bounds.

According to Theorem 3.1, the lower bound is achieved if the random variable is a constant. Therefore, the upper bound for the given function is achieved if $J_i = \mathrm{E}\{J_i\} = I_i$, i.e., if the channel $X_i \to Y_i$ is a BSC.

On the other hand, the upper bound in Theorem 3.1 is achieved if the probability of the random variable is concentrated at the ends of the interval. Therefore, the lower bound of the given function is achieved if $p_{J_i}(j_i)$ is different from zero only for $j_i = 0$ and for $j_i = 1$, i.e., if the channel $X_i \to Y_i$ is a BEC.

As there are actually channels achieving the bounds, the bounds are tight. This completes the proof. □

From the above proof, another interesting property becomes obvious, that was first addressed in [14,15]: Also if only one of the channels is varied, the combined mutual information becomes maximal if this channel is a BSC, and it becomes minimal if this channel is a BEC.

For illustrating examples of this theorem, we refer the reader to the applications in Chapter 6.

5

Information Combining for Repetition Codes

This section addresses the combining of mutual information for the case where the channel inputs are required to have the same values, i.e., the input symbols form a codeword of a repetition code. This is done in a similar way as for single parity-check codes in the previous chapter. First the ideas from the introduction are recapitulated and confined to the situation addressed here, the problem is defined and the notation is introduced. Then the two cases are considered where the channels are all BECs and where the channels are all BSCs. Subsequently, this is generalized to arbitrary BISMCs. Based on this, the bounds on the combined information are derived and discussed. As opposed to the case with the SPC codes, not the extrinsic information but the complete information is addressed.

5.1 Problem Description

The Problem of information combining for a repetition code is depicted in Fig. 5.1. The binary symbols $X_0, X_1, \ldots, X_{N-1} \in \mathbb{F}_2$ are required to fulfill the equality constraint

$$X_0 = X_1 = \cdots = X_{N-1} \tag{5.1}$$

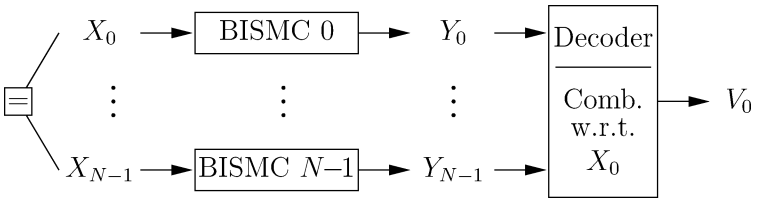

Fig. 5.1 Optimal information combining for a repetition code (parallel information combining).

and thus form a codeword of length N. All codewords are assumed to have the same a-priori probability, and thus also each code symbol is uniformly distributed. The code symbols $X_0, X_1, \ldots, X_{N-1}$ (i.e., all code symbols) are transmitted over independent parallel BISMCs. The channel output values $Y_0, Y_1, \ldots, Y_{N-1}$ are combined by a decoder to the value V_0. As before, optimal combining with respect to X_0 is assumed, such that[1]

$$I(X_0; \boldsymbol{Y}_{[0,N-1]}) = I(X_0; V_0). \tag{5.2}$$

As opposed to the case of SPC codes in the previous chapter, the value V_0 is not extrinsic with respect to X_0, as the direct observation of X_0 is also used for its computation. (Cf. the corresponding discussion in Chapter 4.) Correspondingly, this chapter deals with the complete information $I(X_0; \boldsymbol{Y}_{[0,N-1]})$ (not extrinsic). The motivation for doing so is: for repetition codes, the bounds for the extrinsic information are just a special case of the bounds for the complete information, as will be shown. (This is not the case for SPC codes.) Therefore, this approach gives us more general results.

One optimal way of combining is the computation of a-posteriori probabilities according to Eq. (1.7). Another optimal way of combining is computing the a-posteriori L-value

$$v_0 := \mathrm{L}(X_0 | \boldsymbol{y}_{[0,N-1]}).$$

With $l_i := \mathrm{L}(X_i | y_i)$, the operation may be performed by [35]

$$v_0 = l_0 + l_1 + \cdots + l_{N-1}.$$

[1] As before, the notation $\boldsymbol{Y}_{[0,N-1]} := [Y_0, Y_1, \ldots, Y_{N-1}]$ is employed for vectors of symbols.

Combining of Mutual Information

As for the SPC code, the decoder is interpreted as a processor of mutual information. Each channel output Y_i conveys information about the corresponding channel input X_i,

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

$i = 0, 1, \ldots, N-1$. This information is used by the decoder to compute the (complete) information about X_0,

$$I_{\mathsf{cmp},0} := I(X_0;V_0) = I(X_0;\boldsymbol{Y}_{[0,N-1]}).$$

(The last equality holds as optimal combining is assumed.) Thus, the decoder combines $I_0, I_1, \ldots, I_{N-1}$ to $I_{\mathsf{cmp},0}$. This is referred to as combining of mutual information, or simply as *information combining*. The value I_i is referred to as the intrinsic information on code symbol X_i, and the value $I_{\mathsf{cmp},0}$ is referred to as the complete information on code symbol X_0 [10, 13, 17].

The *question* is now as follows: For given values $I_0, I_1, \ldots, I_{N-1}$, what is the value of $I_{\mathsf{cmp},0}$? The actual value of $I_{\mathsf{cmp},0}$ depends on the distinct mutual information profiles of the individual channels; the maximum and the minimum value, however, can be determined, and they are achieved if the individual channels are BECs or BSCs. As opposed the the SPC code, the maximum value is obtained with BECs, and the minimum value with BSCs. Again, these properties are proved using the concept of mutual information profiles (cf. Chapter 2) and the extension of Jensen's inequality (cf. Chapter 3).

5.2 Binary Erasure Channels

The case where the individual channels are all BECs was already discussed in detail in Chapter 1. For a BEC, the mutual information I and the erasure probability γ are related by $I = 1 - \gamma$. Using this and (1.9), the complete information on X_0 and the values of intrinsic information are related by

$$I_{\mathsf{cmp},0}^{\mathsf{BEC}} = 1 - (1 - I_0)(1 - I_1)\cdots(1 - I_{N-1}). \tag{5.3}$$

This equation represents the information combining for repetition codes when the channels are all BECs.

5.3 Binary Symmetric Channels

Consider now the case where the individual channels are all BSCs. Remember the one-to-one relation between the mutual information I and the crossover probability ϵ for BSCs: $I = 1 - h(\epsilon)$. The complete mutual information can be computed by a combinatorial approach.

Starting with

$$I(X_0; \boldsymbol{Y}_{[0,N-1]}) = H(\boldsymbol{Y}_{[0,N-1]}) - H(\boldsymbol{Y}_{[0,N-1]}|X_0),$$

the first term requires the distribution of the vector of channel outputs using the crossover probabilities of the individual channels. The second term can be written as

$$H(\boldsymbol{Y}_{[0,N-1]}|X_0) = \sum_{i=0}^{N-1} H(Y_i|X_0),$$

using the conditional independence of the channel outputs. (Remember that the channels are assumed to be independent and memoryless.) For convenience, the following function is introduced.

Definition 5.1 Let $I_0, I_1, \ldots, I_{N-1} \in [0,1]$, $N \geq 1$. The functions

$$f_N^{\mathsf{par}}(I_0, I_1, \ldots, I_{N-1}) := -\sum_{\boldsymbol{b} \in \mathbb{F}_2^N} \psi(\boldsymbol{b}) \operatorname{ld} \psi(\boldsymbol{b}) - \sum_{i=0}^{N-1} (1 - I_i)$$

are called *binary information functions for parallel concatenation* with

$$\psi(\boldsymbol{b}) := \frac{1}{2}\Big(\prod_{i=0}^{N-1} \varphi_i(b_i) + \prod_{i=0}^{N-1} \big(1 - \varphi_i(b_i)\big) \Big)$$

and

$$\varphi_i(b_i) := \begin{cases} 1 - \epsilon_i & \text{for } b_i = 0, \\ \epsilon_i & \text{for } b_i = 1, \end{cases}$$

where $\epsilon_i := h^{-1}(1 - I_i)$ for $i = 0, 1, \ldots, N-1$.

Similar to Definition 4.1, the case $N = 1$ is included, so that the following formulas can be written in a more compact form. Appendix A provides some further details.

Using this function, the complete information on code symbol X_0 can be written as

$$I_{\mathsf{cmp},0}^{\mathsf{BSC}} = f_N^{\mathsf{par}}(I_0, I_1, \ldots, I_{N-1}). \tag{5.4}$$

5.4 Binary-Input Symmetric Memoryless Channels

Consider now the case where the individual channels are only required to be BISMCs. To compute the complete mutual information, the same method as for SPC codes is used:

(a) the individual channels are decomposed into BSCs;
(b) the complete information is computed for each set of BSCs;
(c) the average complete information is obtained by averaging over the mutual information profiles of the individual channels.

Thus, the problem is solved by using the result for the all-BSC case and the concept of mutual information profiles.

All channels are assumed to be BISMCs. Therefore, we may define a subchannel indicator A_i and a mutual information indicator J_i for each channel $X_i \to Y_i$ (cf. Chapter 2). Remember that the expectation of the mutual information indicator is equal to the mutual information of the channel, i.e., $\mathrm{E}\{J_i\} = I(X_i;Y_i) = I_i$.

The mutual information does not change if it is written conditioned on the subchannel indicators $\boldsymbol{A}_{[0,N-1]}$:

$$\begin{aligned} I_{\mathsf{cmp},0} &= I(X_0; \boldsymbol{Y}_{[0,N-1]}) = I(X_0; \boldsymbol{Y}_{[0,N-1]}, \boldsymbol{A}_{[0,N-1]}) \\ &= \underbrace{I(X_0; \boldsymbol{A}_{[0,N-1]})}_{=0} + I(X_0; \boldsymbol{Y}_{[0,N-1]} | \boldsymbol{A}_{[0,N-1]}) \\ &= I(X_0; \boldsymbol{Y}_{[0,N-1]} | \boldsymbol{A}_{[0,N-1]}). \end{aligned}$$

This holds because $\boldsymbol{A}_{[0,N-1]}$ is independent of X_0 by Definition 2.1. For a given realization $\boldsymbol{a}_{[0,N-1]}$ of subchannel indicators, we have the case where the channels are all BSCs, and we can apply the function $f_{N-1}^{\mathsf{par}}(\cdot)$ according to Definition 5.1, as done in the all-BSC case:

$$I(X_0; \boldsymbol{Y}_{[0,N-1]} | \boldsymbol{A}_{[0,N-1]} = \boldsymbol{a}_{[0,N-1]}) = f_{N-1}^{\mathsf{par}}(j_0, j_1, \ldots, j_{N-1}).$$

Notice that $j_i = I(X_i;Y_i|A_i = a_i)$ is the mutual information corresponding to the subchannel $A_i = a_i$ of channel $X_i \to Y_i$.

Taking the expectation, we obtain the desired closed form expression:

$$\begin{aligned} I_{\mathsf{cmp},0} &= I(X_0;\boldsymbol{Y}_{[0,N-1]}) \\ &= \mathrm{E}\Big\{ f_{N-1}^{\mathsf{par}}(J_0, J_1, \ldots, J_{N-1}) \Big\}. \end{aligned} \tag{5.5}$$

For continuous-valued mutual information profiles $p_{J_i}(j_i)$, this expectation is evaluated as

$$\begin{aligned} I_{\mathsf{cmp},0} = \int_{j_0} \cdots \int_{j_{N-1}} & \underbrace{p_{J_0}(j_0) \cdots p_{J_{N-1}}(j_{N-1})}_{\text{mutual information profiles}} \cdot \\ & \cdot \quad \underbrace{f_N^{\mathsf{ser}}(j_0, \ldots, j_{N-1})}_{\text{information combining for BSCs}} \quad dj_0 \cdots dj_{N-1}. \end{aligned} \tag{5.6}$$

As for the SPC codes, the exact value of the combined mutual information can be computed via the mutual information profiles of the individual channels, i.e., the mutual information profiles are sufficient to characterize the decoding operation from an information-theory point of view.

5.5 Information Bounds

The previous three sections discussed the cases where the individual channels are known, i.e., where the complete mutual information profiles of the channels are known. In this section, it is assumed that the individual channels are BISMCs and that only their mutual information values are known, i.e., that the mean values of their mutual information profiles are known. Based on this knowledge only, tight bounds on the combined information can be given. These bounds are achieved by the cases where the individual channels are all BECs or all BSCs [10,13–15]. Notice that these cases correspond to the minimum-variance and to the maximum-variance mutual information profiles [26, 36].

Theorem 5.1 Let $X_0, X_1, \ldots, X_{N-1} \in \mathbb{F}_2$ denote the code symbols of a repetition code of length N. Let $X_i \to Y_i$, $i = 0, 1, \ldots, N-1$, denote N independent BISMCs having mutual information $I(X_i;Y_i)$. Let the intrinsic information on code symbol X_i be defined by $I_i := I(X_i;Y_i)$, $i = 0, 1, \ldots, N-1$, and let the complete information on code symbol X_0 be defined by $I_{\mathsf{cmp},0} := I(X_0; \boldsymbol{Y}_{[0,N-1]})$. Then, the following tight bounds hold:

$$I_{\mathsf{cmp},0} \geq f_N^{\mathsf{par}}(I_0, I_1, \ldots, I_{N-1}),$$
$$I_{\mathsf{cmp},0} \leq 1 - (1 - I_0)(1 - I_1) \cdots (1 - I_{N-1}).$$

The lower bound is achieved if the channels are all BSCs, and the upper bound is achieved if the channels are all BECs.

Note that BSCs achieve the lower bound for repetition codes, but the upper bound for single parity-check codes; the reverse holds for BECs.

To prove this theorem, we use the following lemma.

Lemma 5.2 The binary information functions for parallel concatenation, $f_N^{\mathsf{par}}(I_0, I_1, \ldots, I_{N-1})$ (see Definition 5.1) are convex-$\cup$ in each argument I_i, $i = 0, 1, \ldots, N-1$.

Proof. For the proof, the interpretation of the function $f_N^{\mathsf{par}}(\cdot)$ is used, namely that it gives the mutual information of N parallel BSCs $X \to Y_i$ that have mutual information I_i, respectively, and that have the same equiprobable input X, $i = 0, 1, \ldots, N-1$.

Consider first the case $N = 2$. The function $f_2^{\mathsf{par}}(\cdot)$ and the function $f_2^{\mathsf{ser}}(\cdot)$ are related as

$$f_2^{\mathsf{par}}(I_0, I_1) = I_0 + I_1 - f_2^{\mathsf{ser}}(I_0, I_1). \tag{5.7}$$

The proof can be found in [8,9]. As this is very important at this point, we repeat it here.

Consider the situation depicted in Fig. 5.2: the two BSCs $X \to Y_0$ and $X \to Y_1$, having the mutual information $I_0 = I(X;Y_0)$ and $I_1 = I(X;Y_1)$, and having the same input X, which is assumed to be equiprobable. Then the relation

$$I(X; Y_0, Y_1) = I(X;Y_0) + I(X;Y_1) - I(Y_0;Y_1) \tag{5.8}$$

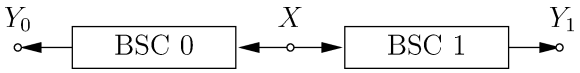

Fig. 5.2 Illustration of the channels from (5.8).

holds. By definition, we have $I(X;Y_0,Y_1) = f_2^{\mathsf{par}}(I_0,I_1)$. According to the given situation, the three random variables form the Markov chain $Y_0 \leftrightarrow X \leftrightarrow Y_1$. Thus the channel $Y_0 \to Y_1$ is the serial concatenation of the two channels $Y_0 \to X$ (which is equivalent to the channel $X \to Y_0$) and $X \to Y_1$. Therefore, its mutual information can be expressed as $I(Y_0;Y_1) = f_2^{\mathsf{ser}}(I_0,I_1)$. This proves (5.8).

Notice that this relation holds *only* for $N = 2$. Equation (5.8) together with Lemma 4.2 prove the convexity for $N = 2$.

Consider now the case $N > 2$. The vector $\boldsymbol{Y}_{[0,N-1]}$ is partitioned[2] into Y_0 and $Y' := \boldsymbol{Y}_{[1,N-1]}$. The superchannel $X_0 \to Y'$ is also a BISMC as the individual channels $X_i \to Y_i,\ i = 1,\ldots,N-1$, are BISMCs. Therefore, there is a subchannel indicator A' and a mutual information indicator J' for this channel.[3] Furthermore, each subchannel of the superchannel $X_0 \to Y'$ is a BSC. Notice that $\mathrm{E}\{J'\} = I(X;\boldsymbol{Y}_{[1,N-1]})$.

Using this superchannel, the mutual information of the N parallel BSCs and thus function $f_N^{\mathsf{par}}(\cdot)$ can be written as

$$I(X;\boldsymbol{Y}_{[0,N-1]}) = f_N^{\mathsf{par}}(I_0,I_1,\ldots,I_{N-1}) = \mathrm{E}\Big\{ f_2^{\mathsf{par}}(I_0,J') \Big\}. \tag{5.9}$$

The function $f_2^{\mathsf{par}}(I_0,J')$ is convex-$\cup$ in I_0 for any $J' = j'$, as shown above. Furthermore, $f_N^{\mathsf{par}}(I_0,I_0,\ldots,I_{N-1})$ is a weighted sum of such functions according to (5.9). Therefore, $f_N^{\mathsf{par}}(I_0,I_1,\ldots,I_{N-1})$ is convex-$\cup$ in I_1.

The same reasoning can be applied for $I_1,I_2,\ldots,I_{N-1}$, which completes the proof. □

Lemma 5.2 is now used to prove Theorem 5.1:

Proof. According to (5.5), the combined information can be written as

$$I_{\mathsf{cmp},0} = \mathrm{E}\Big\{ f_N^{\mathsf{par}}(J_0,J_1,\ldots,J_{N-1}) \Big\}.$$

[2] The choice of this particular partioning is arbitrary and without loss of generality.

[3] Notice that the BSCs into which the superchannel $X_0 \to Y'$ is decomposed are quite different from the $N-1$ BSCs which form this superchannel.

The function $f_N^{\mathsf{par}}(\cdot)$ is convex-$\cup$, and therefore, the extension of Jensen's inequality, Theorem 3.1, can be applied. As Theorem 3.1 addresses only a function with one argument, whereas the given function has multiple arguments, the theorem has to be applied successively with respect to each argument. Notice that the mutual information indicators J_i are confined to the interval $[0,1]$.

Consider first the lower bound. For the time being, let only J_0 be a random variable and let the other mutual information indicators be fixed, $J_i = j_i$, $i = 1,\ldots,N-1$. Then, the lower bound results as

$$\begin{aligned}
&\mathrm{E}\Big\{ f_N^{\mathsf{par}}(J_0, j_1, \ldots, j_{N-1}) \Big\} \\
&\qquad \geq f_N^{\mathsf{par}}\Big(\mathrm{E}\{J_0\}, j_1, \ldots, j_{N-1}) \Big) \\
&\qquad = f_N^{\mathsf{par}}(I_0, j_1, \ldots, j_{N-1}).
\end{aligned}$$

Evaluating the expected values with respect to $J_1,\ldots,J_{N-1}$ in an analog way, the overall lower bound is obtained.

Consider now the upper bound. For the time being, let only J_0 be a random variable and let the other mutual information indicators be fixed, $J_i = j_i$, $i = 1,\ldots,N-1$. Then, according to Theorem 3.1, the upper bound results as

$$\begin{aligned}
&\mathrm{E}\Big\{ f_N^{\mathsf{par}}(J_0, j_1, \ldots, j_{N-1}) \Big\} \\
&\quad \leq f_N^{\mathsf{par}}(0, j_1, \ldots, j_{N-1}) \\
&\qquad + \frac{f_N^{\mathsf{par}}(1, j_1, \ldots, j_{N-1}) - f_N^{\mathsf{par}}(0, j_1, \ldots, j_{N-1})}{1-0} \cdot \big(\mathrm{E}\{J_0\} - 0 \big) \\
&\quad = f_{N-1}^{\mathsf{par}}(j_1, \ldots, j_{N-1}) + I_0 - f_{N-1}^{\mathsf{par}}(j_1, \ldots, j_{N-1}) \cdot I_0 \\
&\quad = 1 - (1 - I_0)\Big(1 - f_{N-1}^{\mathsf{par}}(j_1, \ldots, j_{N-1}) \Big).
\end{aligned}$$

It was used that

$$\begin{aligned}
f_N^{\mathsf{par}}(0, j_1, \ldots, j_{N-1}) &= f_{N-1}^{\mathsf{par}}(j_1, \ldots, j_{N-1}), \\
f_N^{\mathsf{par}}(1, j_1, \ldots, j_{N-1}) &= 1
\end{aligned}$$

(cf. interpretation of this function) and $\mathrm{E}\{J_0\} = I_0$. Notice that $f_1^{\mathsf{par}}(j) = j$. Evaluating the expected values with respect to $J_1,\ldots,J_{N-1}$ in an analog way, the overall upper bound is obtained.

As there are actually channels achieving the bounds, the bounds are tight. This completes the proof. □

If only one of the channels is varied, the combined mutual information becomes maximal if this channel is a BEC, and it becomes minimal if this channel is a BSC, as first addressed in [14, 15]; this is similar to the case of SPC codes. On the other hand, the above proof shows that the upper bound is achieved only if *all* channels are BECs, as opposed to the case of SPC codes.

If not the complete information but the extrinsic information on a code symbol is of interest, Theorem 5.1 can directly be applied; the only difference is that the direct observation of this code symbol has to be omitted. To make it easier to refer to these bounds on the extrinsic information for repetition codes, the following theorem is added.

Theorem 5.3 Let $X_0, X_1, \ldots, X_{N-1} \in \mathbb{F}_2$ denote the code symbols of a repetition code of length N. Let $X_i \to Y_i$, $i = 1, 2, \ldots, N-1$, denote $N-1$ independent BISMCs having mutual information $I(X_i;Y_i)$. Let the intrinsic information on code symbol X_i be defined by $I_i := I(X_i;Y_i)$, $i = 1, 2, \ldots, N-1$, and let the extrinsic information on code symbol X_0 be defined by $I_{\mathsf{ext},0} := I(X_0; \boldsymbol{Y}_{[1,N-1]})$. Then, the following tight bounds hold:

$$I_{\mathsf{ext},0} \geq f_{N-1}^{\mathsf{par}}(I_1, I_2, \ldots, I_{N-1}),$$
$$I_{\mathsf{ext},0} \leq 1 - (1 - I_1)(1 - I_2) \cdots (1 - I_{N-1}).$$

The lower bound is achieved if the channels are all BSCs, and the upper bound is achieved if the channels are all BECs.

6

Applications and Examples

The earlier chapters addressed the information-theoretic bounds on information combining, in particular for single parity-check codes and for repetition codes. These bounds are now applied in several ways. First, the information processing characteristic (IPC) and the extrinsic information transfer (EXIT) functions are bounded. Then it is shown how these functions can be used to design multiple turbo codes. The bounds on the EXIT functions are applied to derive bounds on the decoding thresholds of LDPC codes. Finally, the bounds on information combining are used to bound the EXIT function of the accumulator.

6.1 Information Processing Characteristic

The performance of a channel coding scheme is usually visualized using plots of the bit or frame error probability curves versus one or more parameter(s) specifying the communication channel. The channel coding scheme comprises a code, an encoding rule and a decoding method; the communication channel is the model used for the channel between the encoder and the decoder.

The quality of the decisions in favor of the estimated codeword or of individual information symbols, however, is not adequately expressible by means of this method. It is well known that for concatenation of data processors (e.g. source and channel coding or concatenated channel coding), the propagation of reliability estimates together with data estimates (soft output decoding) for frames or individual symbols from a constituent decoder to a subsequent data processor yields high gains over hard output decoding (hard decision decoding).

Especially in the area of unreliable decisions, e.g., when the code rate exceeds the channel capacity, the decoder output sequence is characterized rather insufficiently by the error rates of the hard decisions. But exactly this situation usually is present for the constituent codes in iteratively decodable schemes as long as no or only little extrinsic information is available from further data processors.

To provide a very general framework of the specification of the quality of a coding scheme, the concept of information processing characteristics (IPC) was suggested in [20]. Consider the typical transmission system depicted in Fig. 6.1, consisting of a binary linear encoder, a binary-input symmetric memoryless channel (BISMC), and a decoder. Without restriction of generality, we assume that the input symbols U_i of the encoder are binary, and independent and uniformly distributed, and thus have entropy $H(U) = 1$ bit/symbol. As the focus in this book is restricted to binary-input channels, the code symbols are assumed to be binary. The encoder of rate $R = K/N$ encodes K information symbols U into N code symbols X and forwards them to the communication channel. Notice that the individual code symbols are equiprobable, and that R is also the average entropy per code symbol. The channel is assumed to have channel capacity C. The channel output symbols

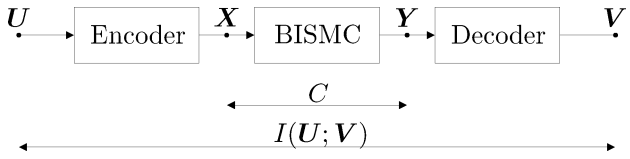

Fig. 6.1 Transmission system with binary linear encoder, binary-input symmetric memoryless channel (BISMC), and decoder.

are given to a decoder, which processes them to the values V. These outputs of the channel decoder may be hard decisions, soft-decisions, or any other values.

In [20], it is suggested to visualize the mutual information $I(\mathbf{U};\mathbf{V})$ between the encoder input sequence $\mathbf{U}$ and the decoder output sequence $\mathbf{V}$ versus the capacity C of the communication channel; to be more precise, the mutual information per information symbol is used. As already mentioned in Chapter 1, the mutual information corresponds to the channel capacity, as only symmetric channels and uniformly distributed channel inputs are considered. Depending on what is more appropriate in the context, we may use the term "information" or the term "capacity."

We call the curves $I(\mathbf{U};\mathbf{V})$ versus C the *IPC* of the coding scheme, IPC(C), because it expresses concisely how the coding scheme with rate $R = K/N$ is able to exploit N uses of the communication channel for recovery of K binary information symbols by means of the decoder output sequence $\mathbf{V}$.

6.1.1 Types of IPCs

We distinguish four types of IPCs, cf. [20], according to different decoding processes.

IPC for Sequence Decoding

The IPC for sequence decoding is defined as

$$\mathrm{IPC}(C) := \frac{1}{K} I(\mathbf{U};\mathbf{V}), \tag{6.1}$$

where K denotes the length of encoder input sequence $\mathbf{U}$ per codeword (frame), and N denotes the length of encoder output sequence $\mathbf{X}$ per codeword (frame). In the function IPC(C), the memory within the output sequence $\mathbf{V}$ due to the code constraints and decoding procedure is taken into account.

An *optimum* sequence decoder is a lossless data processor on a channel output sequence $\mathbf{Y}$ of length N. An example for such an optimal sequence decoder is one that calculates and delivers a complete list of

the a-posteriori probabilities of all 2^K codewords $\boldsymbol{x}_l$; i.e., the output vector $\boldsymbol{v}$ is the vector with the elements $v_l = \Pr(\boldsymbol{x}_l \mid \mathbf{Y})$, $l = 1, 2, \ldots, 2^K$. Since a-posteriori probabilities are sufficient statistics, the inequality (data processing theorem) $I(\mathbf{U};\mathbf{Y}) \geq I(\mathbf{U};\mathbf{V})$ holds with equality. A detailed proof can be found in [20].

IPC for Symbol-by-Symbol Decoding

In symbol-by-symbol decoding, the individual a-posteriori probabilities $\Pr(U_i \mid \mathbf{Y})$ for the information symbols are computed (e.g., using the BCJR-algorithm [37] for codes which can be represented in a trellis with a manageable number of states). Any memory within the decoder output sequence is disregarded.

The IPC for symbol-by-symbol decoding is defined as

$$\mathrm{IPC_I}(C) := \frac{1}{K} \sum_{i=1}^{K} I(U_i;V_i). \tag{6.2}$$

This situation corresponds to the application of a sufficiently large interleaver (subscript "I") in front of the encoder, an appertaining deinterleaver after the decoder, and a these devices for further data processing. If a-posteriori a-posteriori probabilities are forwarded to the user, i.e., $V_i := \Pr(U_i \mid \mathbf{Y})$, we denote a symbol-by-symbol decoder to be optimum. In this case, the inequality $I(U_i;\mathbf{Y}) \geq I(U_i;V_i)$ holds with equality.

IPC for Application of a Specific Decoder

The previous two cases were for optimal decoding, in the sense described above. But in the same, the IPC for any specific decoder can be defined:

$$\mathrm{IPC_{DEC}}(C) := \frac{1}{K} I(\mathbf{U};\mathbf{V}). \tag{6.3}$$

Examples for suboptimal decoders are those employing the MaxLogAPP algorithm [38,39] or sequential search algorithms [40–42].

IPC for Extrinsic-Output Decoding

In the case of systematic encoding, the information symbols U_i are also transmitted over the channel. Therefore, the end-to-end mutual information $I(U_i, \mathbf{Y})$ may be decomposed into an intrinsic part $I(U_i, Y_i) = C$ and an extrinsic part

$$\mathrm{IPC}_{\mathrm{EXT}}(C) := \frac{1}{K} \sum_{i=1}^{K} I(U_i, \mathbf{Y}_{\setminus i}), \tag{6.4}$$

where $\mathbf{Y}_{\setminus i} = [Y_0, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_{N-1}]$ denotes the vector of received values excluding Y_i. The extrinsic part is identical to the famous extrinsic information transfer (EXIT) characteristic of a coding scheme, see e.g., [6, 43, 44].

6.1.2 Properties and Applications of IPCs

Information processing characteristics have several interesting properties and they may be applied in several ways. Some properties and applications are discussed in the following.

Sequence Decoding Versus Symbol-by-Symbol Decoding

Application of the chain rule for mutual information yields the following inequality:

$$\begin{aligned} I(\mathbf{U}; \mathbf{V}) &= I(U_1; \mathbf{Y}) + I(U_2; \mathbf{Y} \mid U_1) + \cdots + I(U_K; \mathbf{Y} \mid U_1 \cdots U_{K-1}) \\ &\geq I(U_1; \mathbf{Y}) + I(U_2; \mathbf{Y}) + \cdots + I(U_K; \mathbf{Y}), \end{aligned} \tag{6.5}$$

because of $I(U_i; \mathbf{Y} \mid \mathbf{Z}) \geq I(U; \mathbf{Y})$ for any $\boldsymbol{Z}$ that is a partial vector of U and does not contain U_i, see e.g. [33]. Thus, for optimum decoding, the inequality $\mathrm{IPC}(C) \geq \mathrm{IPC}_{\mathrm{I}}(C)$ holds because in $\mathrm{IPC}(C)$, the memory in the decoder output sequence is taken into account whereas it is not for $\mathrm{IPC}_{\mathrm{I}}(C)$ (memory increases capacity).

IPC of an Ideal Coding Scheme

We define a coding scheme with rate R to be *ideal* if the end-to-end bit error probability BEP does not exceed the minimum achievable value

BEP_T (index "T" for "tolerated") that has to be tolerated at all, i.e., for a scheme which achieves the rate-distortion bound [33] with equality

$$\mathrm{BEP} = \mathrm{BEP}_\mathrm{T} = \begin{cases} h^{-1}(1 - C/R) & \text{for } C < R, \\ 0 & \text{for } C \geq R. \end{cases} \tag{6.6}$$

cf. Fano's inequality (3.7). In [20], we proved that for such an ideal coding scheme, the functions $\mathrm{IPC}(C)$ and $\mathrm{IPC}_\mathrm{I}(C)$ are identical, and that they are represented by

$$\mathrm{IPC}(C) = \mathrm{IPC}_\mathrm{I}(C) = \min(1, C/R). \tag{6.7}$$

Notice that this equation corresponds to two straight lines.

Together with (6.5), the following inequalities hold for any coding scheme:

$$\mathrm{IPC}_\mathrm{I}(C) \leq \mathrm{IPC}(C) \leq \min(1, C/R). \tag{6.8}$$

For an ideal coding scheme, the end-to-end channel corresponds to a BSC with $\mathrm{BEP} = \mathrm{BEP}_\mathrm{T}$, i.e., all decoder output symbols show identical reliabilities and there is neither difference between hard and soft output decoding nor in sequence and symbol-by-symbol decoding. The IPC of an ideal coding scheme is depicted in Fig. 6.2, right-hand side.

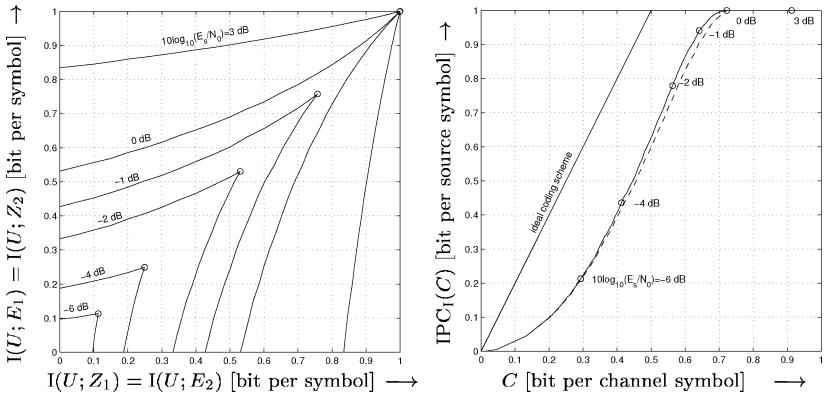

Fig. 6.2 EXIT chart for a rate-1/2 repeat-accumulate code (parallel representation) transmitted over the BIAWGN channel, and bounds on the information processing characteristic $\mathrm{IPC}_\mathrm{I}(C)$.

Bound on Code-Symbol Information

No subset of K binary code symbols X_{l_i} out of the N code symbols provides more information on the transmitted codeword than the K binary information symbols [45]:

$$\frac{1}{K}\sum_{i=1}^{K} I(X_{l_i};\mathbf{Y}) \leq \frac{1}{K} I(\mathbf{U};\mathbf{Y}) \leq \min(1, C/R). \tag{6.9}$$

Taking the average over all possible subsets of K out of N code symbols yields

$$\frac{1}{N}\sum_{i=1}^{N} I(X_i;\mathbf{Y}) \leq \min(1, C/R), \tag{6.10}$$

see [45]. This inequality advises to define an $\mathrm{IPC_I}(C)$ for non-systematic encoding together with optimum symbol-by-symbol decoding:

$$\mathrm{IPC_I}(C) := \frac{1}{N}\sum_{i=1}^{N} I(X_i;\mathbf{Y}). \tag{6.11}$$

For systematic encoding and optimum symbol-by-symbol decoding, the definition in (6.11) coincides with the definition in (6.2). Therefore, the same notation is applied.

Decomposition of the IPC

The end-to-end mutual information $\mathrm{IPC_I}(C)$ may be decomposed into intrinsic and extrinsic mutual information. The intrinsic part exactly corresponds to channel capacity. Thus, $\mathrm{IPC_I}(C)$ follows from the parallel information combining of the abscissa value C and the ordinate value $\mathrm{IPC_{EXT}}(C)$ of the EXIT-curve for serial concatenation of coding schemes. Using the results of Chapter 5, the following bounds hold:

$$\begin{aligned} f_2^{\mathsf{par}}(C, \mathrm{IPC_{EXT}}(C)) &\leq \mathrm{IPC_I}(C) \\ &\leq 1 - (1-C)\,(1 - \mathrm{IPC_{EXT}}(C)) \end{aligned} \tag{6.12}$$

or

$$\frac{\mathrm{IPC_I}(C) - C}{1 - C} \leq \mathrm{IPC_{EXT}}(C) \leq f_2^{\mathsf{par},-1}(C, \mathrm{IPC_{EXT}}(C)) \tag{6.13}$$

with the definition of $f_2^{\mathsf{par},-1}(.,.)$ by means of

$$f_2^{\mathsf{par},-1}\left(f_2^{\mathsf{par}}(\alpha,\beta),\beta\right)=\alpha, \tag{6.14}$$

$\alpha,\beta\in[0,1]$.

The inequalities (6.12) may be used to estimate $\mathrm{IPC_I}(C)$ for a limited number of iterations from the vertices of the ZIG-ZAG trajectory in the EXIT chart for a concatenated coding scheme. An example for the construction of bounds on $\mathrm{IPC_I}(C)$ from the intersection points of EXIT trajectories is given in Fig. 6.2 for the rate 1/2 repeat-accumulate (RA-) code. (In Fig. 6.2, the interpretation of an RA-code as a parallel concatenation is used because here only half the number of iterations is necessary, cf. [1].) One can observe that the bounds constructed in this way are rather tight.

On the other hand, the EXIT curves of constituent codes in concatenated schemes can be approximated within a small interval from $\mathrm{IPC_I}(C)$ using the inequalities (6.13). Together with (6.10) and (6.11), we are also able to express an upper bound on EXIT curves for constituent codes with rate R in serial concatenation, see [3, 13]:

$$\mathrm{IPC_{EXT}}(C)\leq f_2^{\mathsf{par},-1}(\min(1,C/R),C). \tag{6.15}$$

Estimation of Bit Error Probability

From (3.7), bounds on the bit error probability (inequalities of Fano and Raviv/Hellman) result from $\mathrm{IPC_I}(C)$ for symbol-by-symbol decoding, cf. [1], and via inequality (6.12) directly from vertices of the ZIG-ZAG trajectory in the EXIT chart analysis. Thus, by means of parallel information combining, we find bounds for the parametrization of the EXIT-chart with curves for equal bit error probability, which so far were known only in an approximative manner using the Gaussian approximation for pdfs of L-values [6].

Figure 6.3 shows bounds on the bit error probability for the rate-1/2 repeat-accumulate (RA) code that are derived from the bounds in $\mathrm{IPC_I}(C)$ of Fig. 6.3. (Notice that the bounds on the IPC are computed based on EXIT curves that were determined by simulation.) In addition to the bounds, actual simulation results for the bit error probability are also provided.

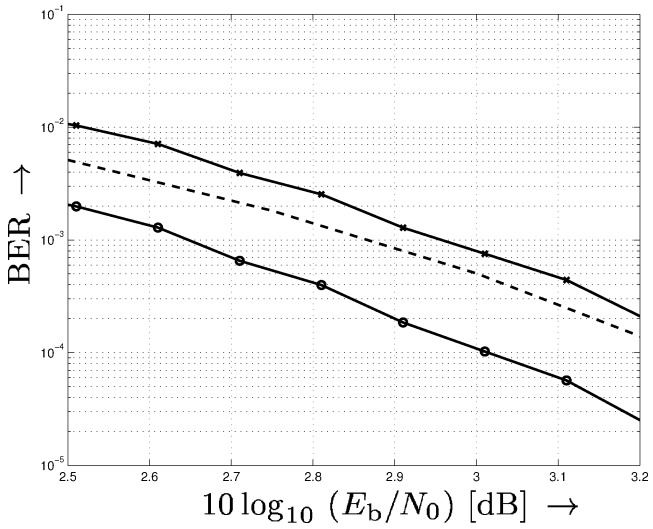

Fig. 6.3 Bounds on the bit error probability derived from the EXIT chart in Fig. 6.2 via bounds on $\mathrm{IPC_I}(C)$ for the rate-1/2 RA code (BIAWGN channel). Dashed line: Simulation result for codeword length 2×10^5, 8 iterations.

6.1.3 Bounds on IPC for Repetition Code

For the $(N,1)$ binary repetition code, the information processing characteristics $\mathrm{IPC}(C)$ and $\mathrm{IPC_I}(C)$ are identical, of course, and they are identical to the curves for parallel information combining, cf. Chapter 5. Notice that both bounds corresponding to BSCs (lower bound) and BECs (upper bound) are surprisingly close together, see Fig. 6.4. The usually applied Gaussian approximation is close to the BEC curve for $0.8 < C < 1$ because the MIP of the AWGN channel is well approximated by the BEC model in this case (cf. Fig. 2.7).

In Fig. 6.4, these bounds are compared with the IPC of an ideal coding scheme with the same rate, $R = 1/N$, but infinite codeword length. The plot shows that such simple repetition codes operate surprisingly close at the optimum in the area $C < R$. The high performance of repeat accumulate codes is based on this effect.

Notice that the extrinsic information $\mathrm{IPC_{EXT}}(C)$ for the $(N,1)$ repetition code corresponds to the $IPC(C)$ for the $(N-1,1)$ repetition code.

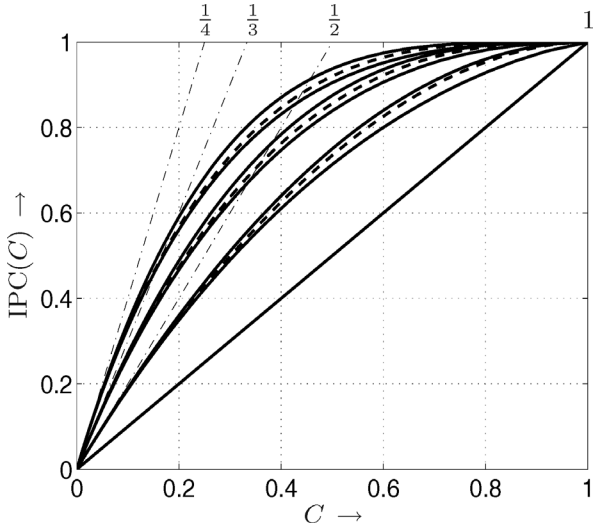

Fig. 6.4 Comparison of bounds on IPCs and Gaussian approximation (dashed line) of repetition codes with rates 1/4, 1/3, 1/2, 1 (from left to right) with IPCs of ideal coding schemes (dash-dotted lines).

6.1.4 Bounds on IPCs for Single Parity-Check Code

For optimum symbol-by-symbol decoding of an $(N, N-1)$ single parity-check (SPC) code, bounds on $\mathrm{IPC_I}(C)$ simply result from parallel information combining of extrinsic and intrinsic information [13]. To be precise, the following procedure is applied: The intrinsic information, C, and the upper bound on the extrinsic information, $f_{N-1}^{\mathsf{ser}}(C, C, \ldots, C)$, are inserted into the upper bound on parallel information combining; this results in an upper bound on the overall mutual information $\mathrm{IPC_I}(C)$. In a similar way, the lower bound is derived. The resulting bounds are:

$$\mathrm{IPC_I}(C) \leq I_{\mathrm{U}}[N] := 1 - (1-C)\left(1 - f_{N-1}^{\mathsf{ser}}(C, C, \ldots, C)\right), \quad (6.16)$$

$$\mathrm{IPC_I}(C) \geq I_{\mathrm{L}}[N] := f_2^{\mathrm{par}}\left(C, C^{N-1}\right). \quad (6.17)$$

These bounds are valid for any BISMC but they are not tight because both extreme channels w.r.t. information combining are used in a mixed way. For example, the upper bound on the extrinsic information for SPC codes is obtained by BECs; the upper bound for parallel combining, however, is obtained by BSCs. In spite of this effect, these bounds

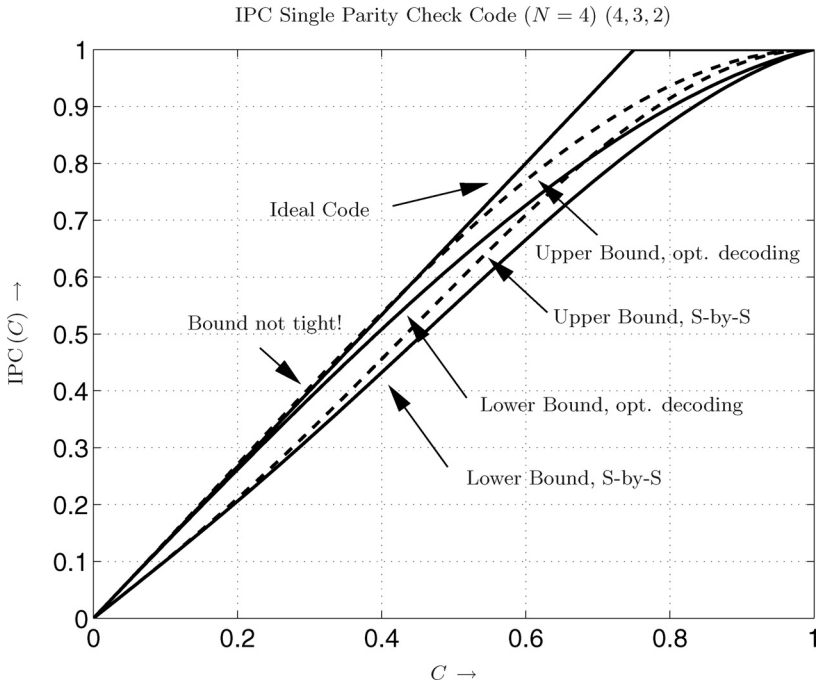

Fig. 6.5 Bounds on IPCs for rate-3/4 SPC code.

are rather close to each other, as can be seen in Fig. 6.5. At very low channel capacity, $C \ll R$, both curves follow the diagonal straight line with slope 1, i.e. no reliability is gained from decoding of extrinsic information, respectively.

Consider now bounds on the IPC for optimum sequence decoding. We first apply the chain rule of mutual information, cf. (6.5):

$$\begin{aligned} I(\mathbf{U};\mathbf{Y}) &= I(U_1;\mathbf{Y}) + I(U_2;\mathbf{Y} \mid U_1) \\ &\quad + \cdots + I(U_K;\mathbf{Y} \mid U_1,\ldots,U_{K-1}) \\ &:= I_{\mathrm{SPC}}[N] + I_{\mathrm{SPC}}[N-1] + \cdots + I_{\mathrm{SPC}}[2] \end{aligned} \tag{6.18}$$

with

$$I_{\mathrm{SPC}}[n] := \mathrm{IPC}_{\mathrm{I}}(C), \tag{6.19}$$

denoting the IPC for symbol-by-symbol decoding of the $(n, n-1)$ SPC code. The value $I(U_1;\mathbf{Y})$ corresponds to the IPC for symbol-by-symbol

decoding for a SPC code of length N, of course. Consider now $I(U_2; \mathbf{Y} \mid U_1)$. If the information symbol U_1 is known, the code is equivalent to an SPC code of length $N-1$. Therefore, the value $I(U_2; \mathbf{Y} \mid U_1)$ corresponds to the IPC for symbol-by-symbol decoding for an SPC code of length $N-1$. In the general case, the decoding of an $(N, N-1)$ SPC code, where L symbols are known, degrades to the decoding of an $(N-L, N-L-1)$ SPC code. This gives the above formulas.

Each term in (6.18) is now bounded using the expressions from (6.16) and (6.17). This gives bounds on the IPC for optimal sequence decoding:

$$\sum_{n=2}^{N} I_{\mathrm{L}}[n] \leq \mathrm{IPC}(C) \leq \sum_{n=2}^{N} I_{\mathrm{U}}[n]. \tag{6.20}$$

These bounds of IPC for sequence decoding are depicted in Fig. 6.5 too. Again, these bounds are not tight because of the mixture of extreme channels. Therefore, the upper bound slightly crosses the straight line for the ideal coding scheme. In this area, this straight line for the ideal coding scheme is obviously a tighter upper bound. In contrast to $\mathrm{IPC_I}(C)$, both bounds follow the optimum line with slope $1/R$ for $C \ll R$; i.e., a list of a-posteriori probabilities of all codewords makes available substantially more information than a-posteriori symbol probabilities because of preserving the memory in the decoder output sequence. For $C < R/2$, these simple codes perform almost as good as possible, whereas the well-known high losses have only to be taken into account if reliable communication, i.e. $\mathrm{IPC}(C) = 1$, is desired.

6.2 Analysis and Design of Multiple Turbo Codes

The extension of the original turbo codes [18, 46] to more than two branches is called multiple turbo codes (see [3] and the references therein). An example for such an encoder is shown in Fig. 6.6.

It is difficult to visualize the convergence of the iterative decoding process for multiple turbo codes by means of regular EXIT charts. In a so-called extended serial setup decoder, extrinsic information from all other constituent codes are combined to have best possible a-priori information for the decoding procedure of the current constituent code,

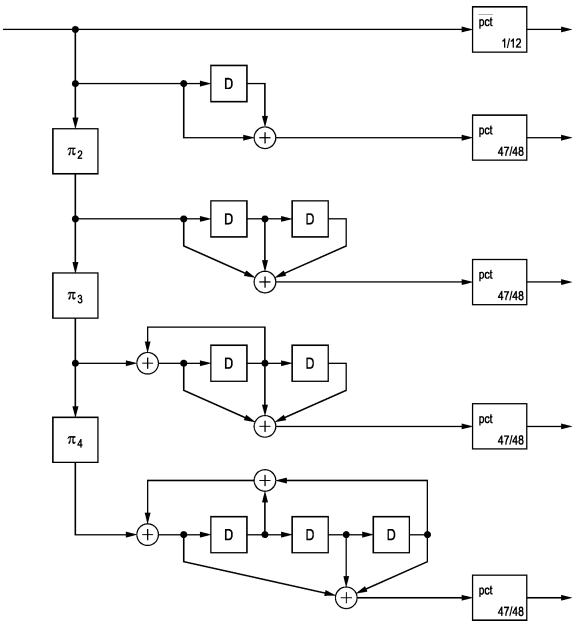

Fig. 6.6 Block diagram of the best multiple turbo code of rate 1/4 with memory ≤ 3 constituent codes designed by means of EXIT curves and parallel information combining, see [3, 22].

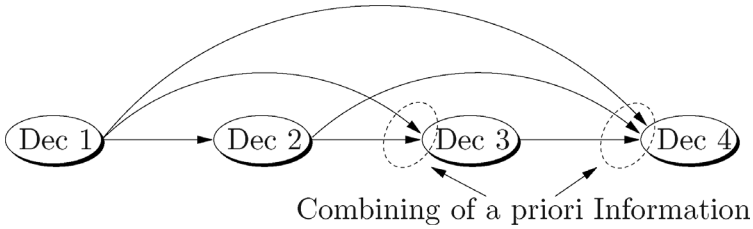

Fig. 6.7 Extended serial setup decoding for a multiple turbo code.

i.e., the sum of extrinsic L-values is applied for representing extrinsic information, see Fig. 6.7.

In terms of EXIT chart analysis (based on simulations of the constituent codes and the Gaussian approximation, as usual, [47]), the extrinsic information from the EXIT curves of all other constituent codes yields the abscissa value for the usual two-dimensional EXIT curve of the actual constituent code to be decoded. Thus, parallel

information combining makes possible to analyze the convergence behavior of an M-ary turbo code based on the usual well-known two-dimensional EXIT curves of the individual constituent codes in a circular serial way [3,22]: The abscissa value at the actual chart is formed from the actual values of the ordinates of all other charts by means of parallel information combining. Convergence is given if at least in one chart the point (1,1) is achieved very closely during the iterative process. For this analysis, the usual EXIT curves of the constituent codes are sufficient in the same way as it is for the classic parallel or serial concatenation of two encoders.

The iterative process going from vertex to vertex in a series of $M-1$ EXIT charts may be performed in a twofold way, one for the lower and one for the upper bound for parallel information combining of the extrinsic information from the other constituent codes (worst case and best case analysis). Many tests have shown that the results for both versions usually differ only slightly and results with sufficient reliability are obtained using the simple formula for parallel information combining valid for the BEC (best case analysis), see Eq. (1.9).

An optimization process for multiple turbo codes using parallel information combining, as proposed in [1], works as follows: Two dimensional EXIT curves are represented by means of simulation points and spline interpolation for the selected subset of possible constituent codes at different signal-to-noise ratios (SNRs) and puncturing rates. Using this database, the multidimensional ZIG-ZAG process is simulated for all possible combinations of constituent codes and puncturing rates yielding the desired overall code rate. The combination is identified for which convergence at the point $(1,1)$ is achieved at least for one constituent code very closely at the lowest SNR. Because the simulation of the ZIG-ZAG process is based on the very simple operations as function read out and parallel information combining for the BEC, a huge number of such runs is possible within a few seconds on a usual PC.

To give an example, Fig. 6.6 shows the structure of the best rate 1/4 4-ary turbo code for the BIAWGN channel, which was found by this process when the number of states per constituent code is restricted to 8 [3]. Simulation results for this code are given in Fig. 6.7 for $K = 10^5, N = 4 \times 10^5$. A comparison in Fig. 6.8 of this curve with the curve

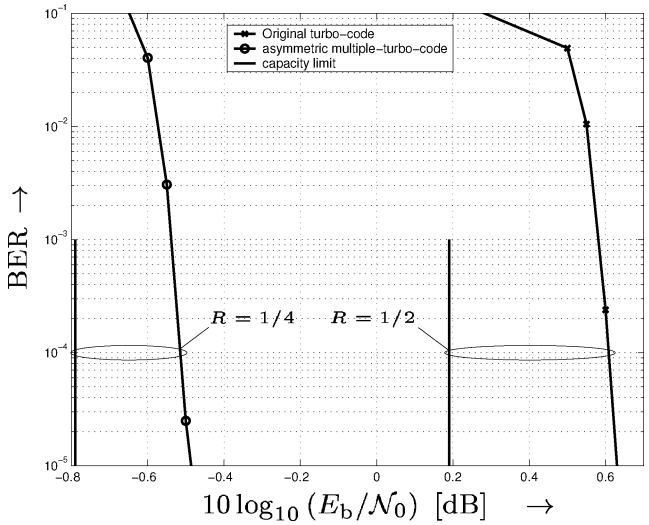

Fig. 6.8 Simulation results of bit error ratios for the 4-ary rate 1/4 turbo code of Fig. 6.7 ($K = 10^5$, $N = 4 \times 10^5$) and comparison with the original rate 1/2 turbo code ($K = 65536$) (BIAWGN-channel).

for the original turbo code [46] ($K = 65546$, $R = 1/2$) shows that this code of rate 1/4 is even closer to the corresponding capacity limit than the original turbo code. Another approach for designing multiple turbo codes, which is also based on EXIT charts, can be found in [48].

6.3 Decoding Thresholds of LDPC Codes

The bounds on information combining can be used to analyze the message-passing decoding of low-density parity-check (LDPC) codes. The assumption of Gaussian distributed messages, used in the original EXIT chart method, is not necessary; this is replaced by assuming that the virtual channels between the code symbols and the corresponding messages are BISMCs, as opposed to the stricter assumption of AWGN channels in the original EXIT chart method. For each variable node and check node decoder, the EXIT function is bounded, and resulting from that, an upper and a lower bound of the decoding threshold is obtained [15]. Thus in a way, the single value of the decoding threshold based on the Gaussian assumption is replaced by an exact range of the

decoding threshold based on information combining; or in other words: the single value that is an approximation is replaced by a region which contains the correct value for all BISMCs. These principles can also be used for code design [15].

In the following, the basics for LDPC codes will shortly be recapitulated. The concept of information combining is used to bound the EXIT functions for the variable-node and the check-node decoder. Finally, based on these bounds, the region for the decoding threshold is determined. For convenience, the discussion is restricted to regular LDPC codes, as the focus in this book is on the presentation of the concepts and principles. The generalization to irregular LDPC codes is straightforward.

6.3.1 LDPC Codes

LDPC codes are defined by their parity-check matrices. The parity-check constraints may be graphically represented in a factor graph comprising one variable node for each code symbol, one check node for each check equation, and edges between variable nodes and check nodes according to the parity-check matrix [49–52].

In the case of regular LDPC codes, each variable node is connected to d_{v} check nodes, and each check node is connected to d_{c} variable nodes. These two values d_{v} and d_{c} are called the variable node degree and the check node degree, respectively. The rate R of an LDPC code is lower-bounded by the design rate R_{d}, which is a function of the two degrees:

$$R \geq R_{\mathrm{d}} = 1 - \frac{d_{\mathsf{v}}}{d_{\mathsf{c}}}.$$

The factor graph is the basis for the iterative decoding algorithm: messages are passed between variable nodes and check nodes; each node computes new outgoing messages based on the incoming messages, i.e., it performs information combining. As usually done, the messages are assumed to be independent, i.e., the cycles in the graph are assumed to be sufficiently large for the decoding process; this turns out to be justified with high probability for long codes. For details on LDPC

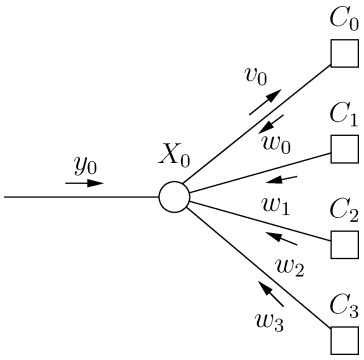

Fig. 6.9 A variable node of degree $d_{\mathsf{v}} = 4$ and messages. (The variable node is drawn as a circle and the check nodes are drawn as squares.)

codes and the message passing algorithm, the reader is referred to [4, 5, 49–54].

Consider now the decoding operation of variable nodes (cf. Fig. 6.9). The variable node representing the code symbol X_0 obtains the output y_0 of the communication channel, referred to as the message based on the direct observation, and d_{v} messages $w_0, \ldots, w_{d_{\mathsf{v}}-1}$ from the connected check nodes. All these incoming messages are assumed to be independent. The outgoing message v_0 sent to check node C_0 (the one that sent w_0) is assumed to be generated in an optimal way such that

$$I(X_0; V_0) = I(X_0; Y_0, W_1, \ldots, W_{d_{\mathsf{v}}-1}).$$

The decoding operation corresponds to decoding of repetition codes. Notice that the message is an extrinsic message as it is independent of w_0. The messages to the other check nodes are computed correspondingly.

Consider now the decoding of check nodes (cf. Fig. 6.10). Without loss of generality, assume that the variable nodes corresponding to the code symbols $X_0, X_1, \ldots, X_{d_{\mathsf{c}}-1}$ are connected to check node C_0. This check node obtains d_{c} messages from these variable nodes, denoted by $v_0, \ldots, v_{d_{\mathsf{c}}-1}$. All these incoming messages are assumed to be independent. The outgoing message w_0 sent to variable node X_0 (the one that sent v_0) is assumed to be generated in an optimal way such that

$$I(X_0; W_0) = I(X_0; V_1, V_2, \ldots, V_{d_{\mathsf{c}}-1}).$$

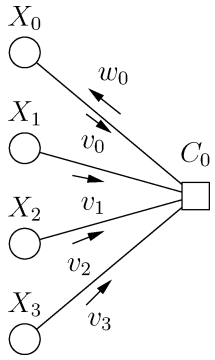

Fig. 6.10 A check node of degree $d_{\mathsf{v}} = 4$ and messages. (The variable-nodes are drawn as circles and the check node is drawn as a square.)

The decoding operation corresponds to decoding of SPC codes. Notice that the message is an extrinsic message as it is independent of v_0. The messages to the other variable nodes are computed correspondingly.

A fundamental question is: When an LDPC code with degrees d_{v} and d_{c} is used on a communication channel with a certain parameter (e.g., an AWGN channel with a certain SNR), what is the worst channel such that the communication is still error-free when the code length tends to infinity? The parameter corresponding to this worst-case channel is called the *decoding threshold*. The question can also be formulated as: What are the optimum degrees for a given communication channel, i.e., the degrees with the largest design rate such that decoding is error-free.

These problems can be solved by density evolution [51,52,55], where the evolution of the conditional probability densities of the messages during the iterations are tracked. This can be rather simplified by assuming that the incoming messages are Gaussian distributed. Then, it is sufficient to track a single parameter of the message distribution. The most successful method of this kind is the extrinsic information transfer (EXIT) chart method [6, 7, 44, 47].

6.3.2 EXIT Charts

In the EXIT chart method, the average mutual information between code symbols and messages is tracked. Assume a communication

channel with mutual information I_{ch} between channel input and channel output, i.e.,

$$I_{\mathsf{ch}} = I(X_i;Y_i)$$

for all code symbols X_i. This value is also referred to as the channel (mutual) information. Consider now a certain iteration of the iterative decoder.

For the variable node of code symbol X_0, the mutual information between this code symbol and the incoming messages,

$$I_{\mathsf{apri}}^{(\mathsf{v})} = I(X_0;W_i),$$

is called the a-priori (mutual) information. The mutual information between the code symbol and the outgoing message,

$$I_{\mathsf{ext}}^{(\mathsf{v})} = I(X_0;V_0) = I(X_0;Y_0,W_1,\ldots,W_{d_\mathsf{v}-1}),$$

is called the extrinsic (mutual) information. As the value is the same for all outgoing messages, this is also the average extrinsic information. The mapping from the a-priori information to the extrinsic information with the channel information as a parameter,

$$\mathrm{EXIT}_\mathsf{v} : I_{\mathsf{apri}}^{(\mathsf{v})} \mapsto I_{\mathsf{ext}}^{(\mathsf{v})}, \tag{6.21}$$

is called the variable-node EXIT function.

From an information theory point of view, the variable node decoder combines one value of channel information I_{ch} (via the communication channel) and $d_\mathsf{v}-1$ values of a-priori information $I_{\mathsf{apri}}^{(\mathsf{v})}$ to the extrinsic information $I_{\mathsf{ext}}^{(\mathsf{v})}$. The corresponding model is depicted in Fig. 6.11. The virtual channels are BISMCs as the communication channel is assumed to be a BISMC and the decoding operations in the variable nodes and check nodes are symmetric.

For the check node connected to the variable nodes of the code symbol $X_0,X_1,\ldots,X_{d_\mathsf{c}-1}$, the mutual information between a code symbol and the corresponding incoming messages,

$$I_{\mathsf{apri}}^{(\mathsf{c})} = I(X_i;V_i),$$

is called the a-priori (mutual) information. The mutual information between a code symbol and the corresponding outgoing message,

$$I_{\mathsf{ext}}^{(\mathsf{c})} = I(X_i;V_i) = I(X_0;V_1,V_2,\ldots,V_{d_\mathsf{c}-1}),$$

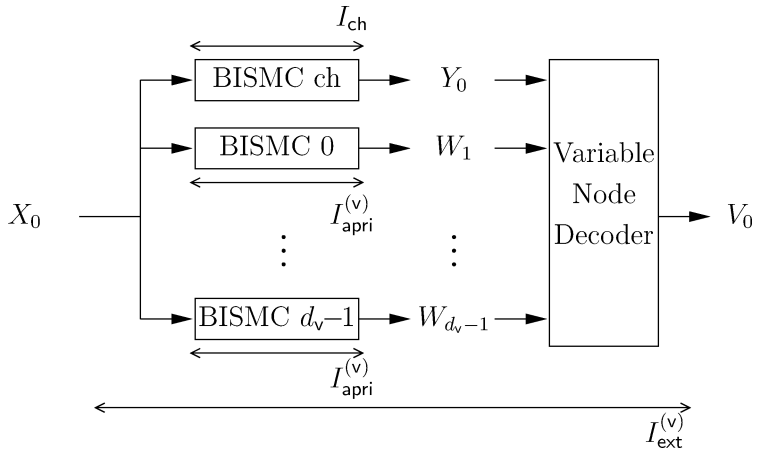

Fig. 6.11 Decoding model for the variable node decoder.

is called the extrinsic (mutual) information. As the value is the same for all outgoing messages, this is also the average extrinsic information. The mapping from the a-priori information to the extrinsic information,

$$\mathrm{EXIT}_{\mathsf{c}} : I^{(\mathsf{c})}_{\mathsf{apri}} \mapsto I^{(\mathsf{c})}_{\mathsf{ext}}, \tag{6.22}$$

is called the check node EXIT function.

From an information theory point of view, the check node decoder combines $d_{\mathsf{c}} - 1$ values of a-priori information $I^{(\mathsf{c})}_{\mathsf{apri}}$ to the extrinsic information $I^{(\mathsf{c})}_{\mathsf{ext}}$. The corresponding model is depicted in Fig. 6.12. The virtual channels are BISMCs as the communication channel is assumed to be a BISMC and the decoding operations in the variable nodes and check-nodes are symmetric.

In the original EXIT chart method, the two EXIT functions are computed independently based on the assumption that the incoming messages are Gaussian distributed. During iterative decoding, the following identities hold:

$$I^{(\mathsf{c})}_{\mathsf{apri}} = I^{(\mathsf{v})}_{\mathsf{ext}} \quad \text{and} \quad I^{(\mathsf{v})}_{\mathsf{apri}} = I^{(\mathsf{c})}_{\mathsf{ext}}. \tag{6.23}$$

The EXIT chart is a plot of both EXIT functions in a single diagram. The decoding threshold can be identified with the worst communication channel, for which there is no intersection between the two EXIT functions.

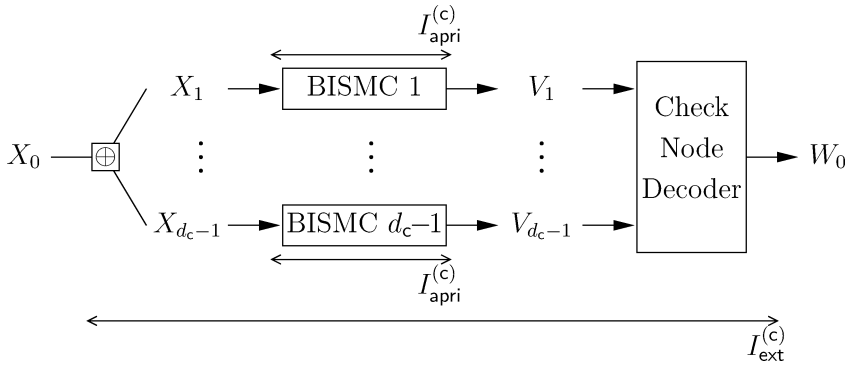

Fig. 6.12 Decoding model for the check node decoder.

Regarding EXIT charts for LDPC codes, two remarks are worth mentioning. First, the EXIT functions are usually determined by running the constituent decoders. They can, however, also be determined analytically by applying (4.10) and (5.6). More results on analytical EXIT functions can be found in [56]. Second, the EXIT chart analysis is valid only for the asymptotic case of infinite code lengths. For further details, we refer the reader to literature on LDPC codes.

6.3.3 Bounds on EXIT Functions

EXIT charts help to predict the behavior of the iterative decoder without actually running it. The EXIT functions of the two constituent decoders are determined separately by assuming a certain distribution of the incoming messages, or equivalently, by assuming a certain channel model for the virtual channel between the code symbols and the correspond. This method was shown to predict quite accurately the decoding threshold of iterative decoders.

Nevertheless, EXIT functions depend on the model for the virtual channel and the communication channel. This has two implications:

(a) The predicted decoding threshold is accurate only if the model for the virtual channel is correct. Although the often used AWGN channel provides a good approximation of the channel that can be observed in an iterative decoder,

it is still an approximation. Thus, the predicted decoding thresholds cannot be proved to be accurate.

(b) The decoding threshold depends on the model for the communication channel.

The original EXIT chart method is now extended using the concept of information combining.

The EXIT functions for specific channel models will be replaced by bounds on EXIT functions that are valid for all BISMCs. Thus, bounds on the true decoding threshold can be obtained, and these bounds are valid for all models of communication channels that are BISMCs. Based on [8], this method was developed in [9–11,13] and [14, 15]. In the following, the bounds on information combining developed in Chapters 4 and 5 will be applied.

Consider a regular LDPC code with variable node degree d_{v} and check node degree d_{c}. The code symbols are transmitted over a communication channel that is a BISMC and has mutual information I_{ch}, referred to as channel information. The iterative decoder operates on the factor graph of the parity-check matrix of the code, and the variable-node and the check-node decoders are assumed to perform optimal decoding.

The decoding operation in a variable node is equivalent to decoding a repetition code of length $d_{\mathsf{v}} + 1$. For each outgoing extrinsic message, the decoder uses the channel information I_{ch} and $(d_{\mathsf{v}} - 1)$ values of a-priori information $I_{\mathsf{apri}}^{(\mathsf{v})}$; the resulting extrinsic information is denoted by $I_{\mathsf{ext}}^{(\mathsf{v})}$. Using Theorem 5.3, the extrinsic information can be bounded by

$$I_{\mathsf{ext}}^{(\mathsf{v})} \geq f_{d_{\mathsf{v}}}^{\mathsf{par}}(I_{\mathsf{ch}}, \underbrace{I_{\mathsf{apri}}^{(\mathsf{v})}, \ldots, I_{\mathsf{apri}}^{(\mathsf{v})}}_{(d_{\mathsf{v}}-1) \text{ arguments}}), \tag{6.24}$$

$$I_{\mathsf{ext}}^{(\mathsf{v})} \leq 1 - (1 - I_{\mathsf{ch}})(1 - I_{\mathsf{apri}}^{(\mathsf{v})})^{d_{\mathsf{v}}-1}. \tag{6.25}$$

The lower bound corresponds to BSCs, and the upper bound corresponds to BECs. This gives the bounds on the EXIT function of the variable-node decoders.

Similarly, the decoding operation in a check node is equivalent to decoding of a single parity-check code of length d_{c}. For each outgoing extrinsic message, the decoder uses $(d_{\mathsf{c}} - 1)$ values of a-priori information $I_{\mathsf{apri}}^{(\mathsf{c})}$; the resulting extrinsic information is denoted by $I_{\mathsf{ext}}^{(\mathsf{c})}$. Using Theorem 4.1, the extrinsic information can be bounded by

$$I_{\mathsf{ext}}^{(\mathsf{c})} \geq (I_{\mathsf{apri}}^{(\mathsf{c})})^{d_{\mathsf{c}}-1}, \tag{6.26}$$

$$I_{\mathsf{ext}}^{(\mathsf{c})} \leq f_{d_{\mathsf{c}}-1}^{\mathsf{ser}}(I_{\mathsf{apri}}^{(\mathsf{c})}, I_{\mathsf{apri}}^{(\mathsf{c})}, \ldots, I_{\mathsf{apri}}^{(\mathsf{c})}). \tag{6.27}$$

The lower bound corresponds to BECs, and the upper bound corresponds to BSCs. This gives the bounds on the EXIT functions for the check-node decoder.

The resulting extended EXIT chart with bounds on the EXIT functions is illustrated in the following example. Consider a regular LDPC code with variable node degree $d_v = 3$ and check node degree $d_c = 4$, having design rate $R_{\mathrm{d}} = 1/4$. This extended EXIT chart of this code is depicted in Fig. 6.13 for two values of channel information I_{ch}. Note that the EXIT function for the check node is flipped; therefore, the upper curve corresponds to the lower bound on the EXIT function, and vice versa.

6.3.4 Bounds on Decoding Threshold

The bounds on the EXIT functions for variable nodes and check nodes can be used to determine the smallest channel information that is necessary for convergence, and the smallest channel information that is sufficient for convergence. Thus, a necessary and a sufficient condition for convergence are obtained, and these conditions are given in terms of the mutual information of the communication channel. These bounds on the convergence threshold are valid for all a-priori channels and all communication channels that are BISMCs. Notice that this mutual information of the communication channel is identical to its capacity, as it is assumed that the channel is a BISMC and that the codewords are equiprobable.

To explain the way to determine the bounds on the convergence threshold, the previous example is continued (cf. Fig. 6.13).

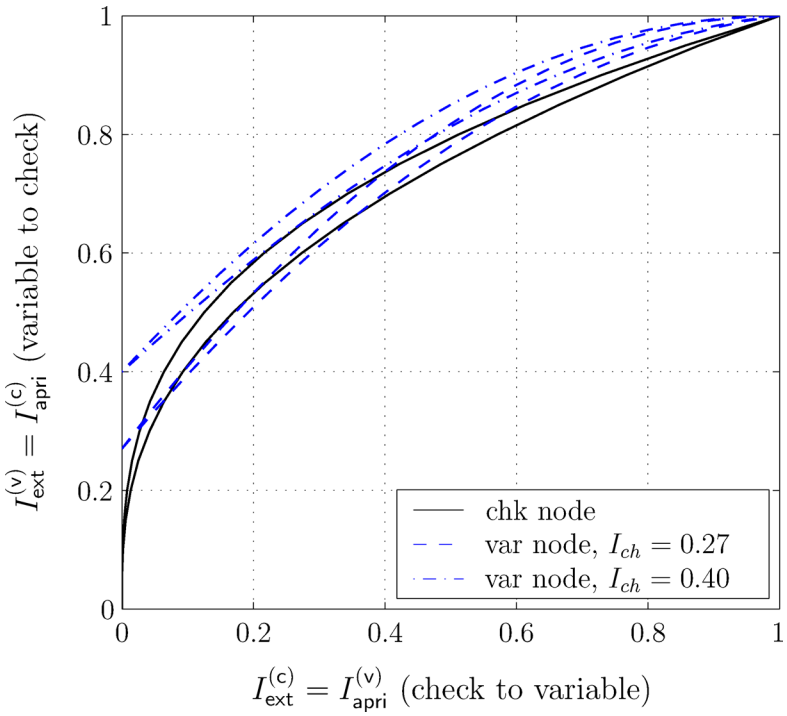

Fig. 6.13 Extended EXIT chart for an LDPC code with variable node degree $d_{\mathsf{v}} = 3$ and check node degree $d_{\mathsf{c}} = 4$ for a communication channel with mutual information $I_{\mathsf{ch}} = 0.27$ and $I_{\mathsf{ch}} = 0.40$. The bounds on the EXIT functions for variable nodes and check nodes are depicted. (Mutual information is given in bit/use.)

Lower bound on decoding threshold: For $I_{\mathsf{ch}} = 0.27$, the upper bound on the check node EXIT function and the upper bound on the variable node EXIT function intersect. Therefore, the decoder cannot converge for any model of virtual channel and any model of the communication channel if $I_{\mathsf{ch}} < 0.27$ (provided that both are BISMCs). Thus, $I_{\mathsf{ch}} = 0.27$ represents a lower bound on the convergence threshold.

Upper bound on decoding threshold: For $I_{\mathsf{ch}} = 0.40$, the lower bound on the check node EXIT function and the lower bound on the variable node EXIT function do not intersect. Accordingly, the decoder will converge for any model of the virtual channel and any model of the communication channel if $I_{\mathsf{ch}} > 0.40$. Thus, $I_{\mathsf{ch}} = 0.40$ represents an upper bound on the convergence threshold.

These two bounds are determined based on the graphical representation of the bounds of the EXIT functions. To obtain more precise results, the decoding thresholds can be determined numerically using conditions for the curves to intersect. The following theorem is based on this concept and gives a lower and an upper bound on the decoding threshold, and thus a necessary and a sufficient condition for convergence, respectively [13–15].

Theorem 6.1 Consider a regular LDPC code (of infinite length) with variable node degree d_{v} and check node degree d_{c}. The communication channel is assumed to be an arbitrary BISMC with mutual information (capacity) I_{ch}. Let $I_{\mathsf{ch},\mathrm{low}}$ be defined as the maximum value $I_{\mathsf{ch}} \in [0,1]$ such that there is an $I \in [0,1]$ fulfilling

$$I = 1 - (1 - I_{\mathsf{ch}})\big(1 - f_{d_{\mathsf{c}}-1}^{\mathsf{ser}}(I, I, \ldots, I)\big)^{d_{\mathsf{v}}-1}.$$

Let further $I_{\mathsf{ch},\mathrm{upp}}$ be defined as the maximum value $I_{\mathsf{ch}} \in [0,1]$ such that there is an $I \in [0,1]$ fulfilling

$$I = f_{d_{\mathsf{v}}}^{\mathsf{par}}(I_{\mathsf{ch}}, I^{d_{\mathsf{c}}-1}, \ldots, I^{d_{\mathsf{c}}-1}).$$

The iterative decoder can converge only if $I_{\mathsf{ch}} > I_{\mathsf{ch},\mathrm{low}}$ (necessary condition), and it converges for sure if $I_{\mathsf{ch}} > I_{\mathsf{ch},\mathrm{upp}}$ (sufficient condition).

The first equality corresponds to an intersection of the upper bounds on the EXIT functions, and the second equality corresponds to an intersection of the lower bounds on the EXIT functions. These intersections of EXIT functions are the basis for proving this theorem.

Proof. Consider first the case that the upper bounds on the EXIT functions have an intersection. The upper bounds are given by (6.25) and (6.27), and they are "coupled" by (6.23). Therefore, they may be written as

$$\begin{aligned} I_{\mathsf{ext}}^{(\mathsf{v})} &= 1 - (1 - I_{\mathsf{ch}})(1 - I_{\mathsf{ext}}^{(\mathsf{c})})^{d_{\mathsf{v}}-1}, \\ I_{\mathsf{ext}}^{(\mathsf{c})} &= f_{d_{\mathsf{c}}-1}^{\mathsf{ser}}(I_{\mathsf{ext}}^{(\mathsf{v})}, I_{\mathsf{ext}}^{(\mathsf{v})}, \ldots, I_{\mathsf{ext}}^{(\mathsf{v})}). \end{aligned}$$

An intersection of the EXIT functions corresponds to a fixed point of these two equations. Substituting the second equation into the first one

and using $I := I_{\mathsf{ext}}^{(\mathsf{v})}$, we obtain the fixed point condition

$$I = 1 - (1 - I_{\mathsf{ch}})\big(1 - f_{d_{\mathsf{c}}-1}^{\mathsf{ser}}(I, I, \ldots, I)\big)^{d_{\mathsf{v}}-1}$$

for an $I \in [0,1]$. When a fixed point exists for a given channel information I_{ch}, the upper bounds on the EXIT functions intersect, and convergence of the iterative decoder is impossible. Therefore, the channel information has to be larger than the maximum value I_{ch} for which a fixed point exists, such that convergence is possible. This proves the first part of the theorem.

Consider now the case that the lower bounds on the EXIT functions have an intersection. The lower bounds are given by (6.24) and (6.26), and they are "coupled" by (6.23). Thus, they may be written as

$$\begin{aligned} I_{\mathsf{ext}}^{(\mathsf{v})} &= f_{d_{\mathsf{v}}}^{\mathsf{par}}(I_{\mathsf{ch}}, I_{\mathsf{ext}}^{(\mathsf{c})}, \ldots, I_{\mathsf{ext}}^{(\mathsf{c})}), \\ I_{\mathsf{ext}}^{(\mathsf{c})} &= (I_{\mathsf{ext}}^{(\mathsf{v})})^{d_{\mathsf{c}}-1}. \end{aligned}$$

An intersection of the EXIT functions corresponds to a fixed point of these two equations. Substituting the second equation into the first one and using $I := I_{\mathsf{ext}}^{(\mathsf{v})}$, we obtain the fixed point condition

$$I = f_{d_{\mathsf{v}}}^{\mathsf{par}}(I_{\mathsf{ch}}, I^{d_{\mathsf{c}}-1}, \ldots, I^{d_{\mathsf{c}}-1})$$

for an $I \in [0,1]$. When a fixed point exists for a given channel information I_{ch}, the lower bounds on the EXIT functions intersect and convergence of the iterative decoder cannot be guaranteed (but may be possible). Therefore, the channel information has to be larger than the maximum value I_{ch} for which a fixed point exists, such that convergence can be guaranteed. This proves the second part of the theorem.

□

Theorem 6.1 may be used to compute the upper and the lower bound on the decoding threshold numerically. The results are usually more precise than those determined graphically using the conventional EXIT chart method. Notice that the lower bound corresponds to a communication channel that is a BEC, and that the upper bound corresponds to a communication channel that is a BSC. For the example considered before, one may compute the thresholds as $I_{\mathsf{ch},\mathrm{low}} = 0.278$ and $I_{\mathsf{ch},\mathrm{upp}} = 0.398$.

The necessary condition of convergence may be used to state whether a code can achieve capacity. If the design rate is smaller than the lower bound on the channel information that is necessary for convergence, the code cannot achieve the capacity of any communication channel that is a BISMC. Consider the example again. The design rate of the code is $R_\mathrm{d} = 1/4$, and the lower bound on the decoding threshold is $I_{\mathsf{ch},\mathrm{low}} = 0.278$. Since $R_\mathrm{d} < I_{\mathsf{ch},\mathrm{low}}$, the given code is not capacity achieving for any communication channel that is a BISMC.

The discussion was restricted to regular LDPC codes. Using similar concepts as for the analysis of LDPC codes for the binary erasure channel, the method can be extended to irregular LDPC codes in a straightforward way. This was done, for example, in [14, 15].

6.4 EXIT Function for the Accumulator

Repeat-accumulate (RA) codes were proposed in [29,57,58] as a class of iteratively decodable codes with low encoding complexity. In particular, the encoding complexity is linear in the code length. They were further developed in [59] such that they can operate very close to the channel capacity when properly designed. In fact, they can be interpreted as LDPC codes. In the following, the structure of RA codes is shortly revised, and then we focus on the accumulator. For more details about RA codes, we refer the reader to the literature.

The encoder of the original proposal [57] consists of repetition encoders as outer encoder, an interleaver, and an accumulator, see Fig. 6.14. The accumulator is the recursive convolutional encoder with rate $R = 1$ and memory length $m = 1$, defined by the generator function

$$g(D) = \frac{1}{1+D}.$$

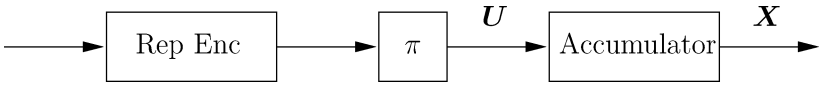

Fig. 6.14 Encoder for the RA code.

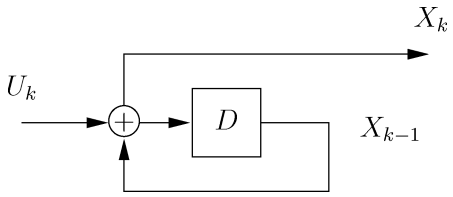

Fig. 6.15 The accumulator and the relation of the information symbols U_k and the code symbols X_k.

It is depicted in Fig. 6.15. The information word[1] of length K is mapped onto the codeword of the same length,

$$[u_0, u_1, \ldots, u_{K-1}] \mapsto [x_0, x_1, \ldots, x_{K-1}].$$

According to the generator function, the binary information symbols U_k and the code symbols X_k, $U_k, X_k \in \mathbb{F}_2$, fulfill the parity-check constraint

$$U_k \oplus X_{k-1} = X_k, \tag{6.28}$$

for $k = 0, 1, \ldots, K-1$; the initial value is defined as $X_{-1} := 0$.

According to the encoder structure, the symbols U_k are the output symbols of the repetition encoders and the input symbols of the accumulator. Correspondingly, during iterative decoding of the RA code, the decoder for the repetition codes (the outer decoder) and the decoder for the accumulator (the inner decoder) exchange extrinsic probabilities on the symbols U_k, see Fig. 6.16. Thus, the decoder for the accumulator obtains one observation y_k (from the communication channel) for each symbol X_k and one a-priori value W_k (generated by the outer decoder) for each information symbol U_k. Based on this, it computes one extrinsic probability v_k for each U_k, which is passed back to the outer decoder:

$$v_k = \Pr(U_k = 0 | \boldsymbol{w}_{\setminus k}, \boldsymbol{y}),$$

where the following notation for vectors is used:

$$\boldsymbol{y} := [y_0, y_1, \ldots, y_{K-1}],$$
$$\boldsymbol{w}_{\setminus k} := [w_0, \ldots, w_{k-1}, w_{k+1}, \ldots, w_{K-1}].$$

[1] As we focus on the accumulator and consider everything from its point of view, we refer to its input symbols as information symbols and to its input word as the information word.

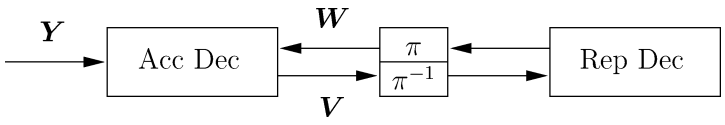

Fig. 6.16 Decoder for the RA code.

As for LDPC codes, this is optimal decoding and thus we have

$$I(U_k;V_k) = I(U_k;\boldsymbol{W}_{\setminus k},\boldsymbol{Y}).$$

Notice that $\boldsymbol{w}_{\setminus k}$ is the vector of all a-priori values excluding w_k.

In the sequel, bounds on the EXIT function of the accumulator are determined. First, the factor graph of the accumulator is introduced. Then, the bounds of information combining are applied on the factor graph in a recursive way. This method was first presented in [60]. The derivation of the bounds for the accumulator is more involved than for the single parity check codes and the repetition codes. However, the technique applied may have the potential to be extendable to other convolutional codes.

Factor Graph and Mutual Information

A factor graph is a graphical representation of code constraints (cf. [49] and the references therein). When depicting the check constraint in (6.28) for all symbols, one obtains the factor graph of the accumulator, shown in Fig. 6.17. This factor graph represents the basis for the following analysis.

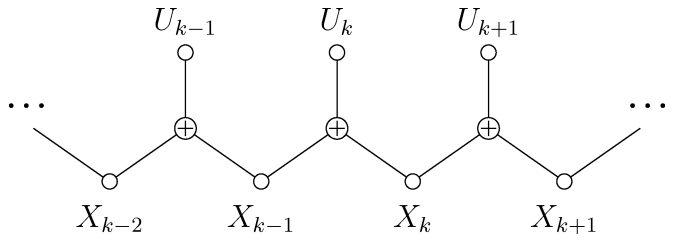

Fig. 6.17 Factor graph of the accumulator.

The lengths K of the information word and of the codeword are assumed to be infinite, i.e., $K \to \infty$, such that the concepts of mutual information can be applied. During iterative decoding, a-priori information is available for all symbols U_k and channel information is available for all symbols X_k,

$$\begin{aligned} I(U_k;W_k) &= I_{\mathsf{apri}}, \\ I(X_k;Y_k) &= I_{\mathsf{ch}}, \end{aligned}$$

for $k = 0, 1, \ldots, K-1$. The extrinsic information on symbol U_k is defined as

$$I_{\mathsf{ext},k} := I(U_k; \boldsymbol{Y}, \boldsymbol{W}_{\setminus k}).$$

Because of the regular code structure and the assumption of infinite length, the average extrinsic information I_{ext} tends to the extrinsic information on a symbol U_k in the middle of the factor graph when the codeword length approaches infinity, i.e.,

$$I_{\mathsf{ext}} \to I_{\mathsf{ext},K/2} \tag{6.29}$$

for $K \to \infty$. Therefore, the average extrinsic information can be bounded by bounding the extrinsic information on a symbol U_k in the middle of the information word, based on the a-priori information and the channel information. If not stated otherwise, $0 \ll k \ll K$ and $K \to \infty$, or for simplicity, $k \approx K/2$ and $K \to \infty$ are assumed.

Derivation of Bounds

Bounds on the extrinsic information for the accumulator that are only based on the channel information and on the a-priori information are given in the following theorem.

Theorem 6.2 Consider an accumulator, where the channel information I_{ch} is available for all code symbols and a-priori information I_{apri} is available for all information symbols. The extrinsic information I_{ext} on information symbols is bounded as

$$I_{\min} \cdot I_{\min} \le I_{\mathsf{ext}} \le f_2^{\mathsf{ser}}(I_{\max}, I_{\max}),$$

where $I_{\min}$ is the minimum value and $I_{\max}$ is the maximum value $I \in [0,1]$ fulfilling simultaneously the two inequalities

$$f_2^{\mathsf{par}}(I_{\mathsf{ch}}, I_{\mathsf{apri}} \cdot I) \quad \leq \quad I \quad \leq \quad 1 - (1 - I_{\mathsf{ch}})\big(1 - f_2^{\mathsf{ser}}(I_{\mathsf{apri}}, I)\big).$$

For proving these bounds, Theorems 4.1 and 5.3 are applied in a nested way, and a stationarity condition is employed.

To start with, (6.28) is written as

$$U_k = X_{k-1} \oplus X_k.$$

Accordingly, the extrinsic information on information symbol U_k,

$$I_{\mathsf{ext},k} := I(U_k; \boldsymbol{Y}, \boldsymbol{W}_{\setminus k}),$$

can be expressed using

$$I_{\alpha,k-1} := I(X_{k-1}; \boldsymbol{Y}_{[0,k-1]} \boldsymbol{W}_{[0,k-1]}), \tag{6.30}$$

$$I_{\beta,k} := I(X_k; \boldsymbol{Y}_{[k,K-1]} \boldsymbol{W}_{[k+1,K-1]}), \tag{6.31}$$

which can be seen from the factor graph, Fig. 6.17. The first term corresponds to the forward recursion and the second term to the backward recursion in the BCJR algorithm [37]; therefore the labels "α" and "β" are used, as usually done in BCJR literature. To bound the extrinsic information, Theorem 4.1 is applied:

$$I_{\alpha,k-1} \cdot I_{\beta,k} \quad \leq \quad I_{\mathsf{ext},k} \quad \leq \quad f_2^{\mathsf{ser}}(I_{\alpha,k-1}, I_{\beta,k}). \tag{6.32}$$

Consider first the mutual information corresponding to the *forward recursion*, $I_{\alpha,k-1}$. This information on code symbol X_{k-1} can be separated into the intrinsic information based on a direct observation, $I(X_{k-1}; Y_{k-1})$, and the left-extrinsic information based on indirect observations,

$$I_{\alpha\mathsf{ext},k-1} := I(X_{k-1}; \boldsymbol{Y}_{[0,k-2]} \boldsymbol{W}_{[0,k-1]}).$$

This term is called left-extrinsic, because it denotes extrinsic information based only on observations that are to the left-hand side of X_{k-1} in the factor graph. Regarding that $I(X_{k-1}; Y_{k-1}) = I_{\mathsf{ch}}$ and applying Theorem 5.3, one obtains

$$I_{\alpha,k-1} \geq f_2^{\mathsf{par}}(I_{\mathsf{ch}}, I_{\alpha\mathsf{ext},k-1}) \tag{6.33}$$

$$I_{\alpha,k-1} \leq 1 - (1 - I_{\mathsf{ch}})(1 - I_{\alpha\mathsf{ext},k-1}). \tag{6.34}$$

Because of the parity-check equation

$$X_{k-1} = U_{k-1} \oplus X_{k-2}$$

following from (6.28), the term $I_{\alpha\mathsf{ext},k-1}$ can be expressed using the intrinsic information on info symbol U_{k-1}, $I(U_{k-1};W_{k-1}) = I_{\mathsf{apri}}$, and the information on code symbol X_{k-2},

$$I(X_{k-2}; \boldsymbol{Y}_{[0,k-2]}\boldsymbol{W}_{[0,k-2]}) = I_{\alpha,k-2};$$

the last identity can be deducted from (6.30). When applying Theorem 4.1, one obtains

$$I_{\mathsf{apri}} \cdot I_{\alpha,k-2} \le I_{\alpha\mathsf{ext},k-1} \le f_2^{\mathsf{ser}}(I_{\mathsf{apri}}, I_{\alpha,k-2}).$$

Substituting the lower bound into (6.33) and the upper bound into (6.34), bounds for the forward recursion are obtained:

$$I_{\alpha,k-1} \ge f_2^{\mathsf{par}}(I_{\mathsf{ch}}, I_{\mathsf{apri}} \cdot I_{\alpha,k-2}) \tag{6.35}$$

$$I_{\alpha,k-1} \le 1 - (1 - I_{\mathsf{ch}})\big(1 - f_2^{\mathsf{ser}}(I_{\mathsf{apri}}, I_{\alpha,k-2})\big). \tag{6.36}$$

The factor graph has a regular structure, and thus $I_{\alpha,k-2} \to I_{\alpha,k-1}$ for $k \to K/2$ and $K \to \infty$. (Remember that the interest is in the extrinsic information "in the middle" of the factor graph.) Let

$$I_\alpha := \lim_{K\to\infty} I_{\alpha,K/2} \tag{6.37}$$

denote the stationary value of $I_{\alpha,k}$. When assuming *stationarity* in (6.35) and (6.36), one obtains

$$I_\alpha \ge f_2^{\mathsf{par}}(I_{\mathsf{ch}}, I_{\mathsf{apri}} \cdot I_\alpha), \tag{6.38}$$

$$I_\alpha \le 1 - (1 - I_{\mathsf{ch}})\big(1 - f_2^{\mathsf{ser}}(I_{\mathsf{apri}}, I_\alpha)\big). \tag{6.39}$$

These two relations are necessary conditions for possible stationary values I_α. Thus, the following lemma for the forward recursion is proved.

Lemma 6.3 Let $I_{\alpha,k}$ be defined according to (6.30), and let I_α be defined as the stationary value of $I_{\alpha,k}$ according to (6.37). Bounds on I_α are given by the minimum and the maximum value of $I_\alpha \in [0,1]$, fulfilling simultaneously (6.38) and (6.39).

Consider now the mutual information corresponding to the *backward recursion*, $I_{\beta,k}$. Since the factor graph is symmetric with respect to k, the analysis for the backward recursion is identical to the one for the forward recursion. Let

$$I_\beta := \lim_{K\to\infty} I_{\beta,K/2} \tag{6.40}$$

denote the stationary value. The necessary conditions for stationary values I_β are the same as for I_α, i.e., they are given in (6.38) and (6.39). Thus, the following lemma for the backward recursion is proved:

Lemma 6.4 (Backward Recursion). Let $I_{\beta,k}$ be defined according to (6.31), and let I_β be defined as the stationary value of $I_{\beta,k}$ according to (6.40). Bounds on I_β are given by the minimum and the maximum value of $I_\alpha \in [0,1]$, fulfilling simultaneously (6.38) and (6.39).

The bounds on I_α and on I_β according to Lemmas 6.3 and 6.4 are now used to bound the extrinsic information $I_{\mathsf{ext},K/2}$. Using the minimum values for I_α and I_β in the lower bound in (6.32), a lower bound on the extrinsic information is obtained. Similarly, using the maximum values of I_α and I_β in the upper bound in (6.32), an upper bound on the extrinsic information is obtained. Considering that the two minimum values are equal and that the two maximum values are equal, as discussed above, and taking into account (6.29), Theorem 6.2 is proved.

Illustration

The bounds on the extrinsic information according to Theorem 6.2 depend only on the channel information and on the a-priori information. Accordingly, they represent bounds on the EXIT functions of the accumulator. These bounds are depicted in Fig. 6.18.

It can be seen that the bounds are close to each other only for small and for large a-priori information. This can be explained as follows. For the derivation of the bounds, Theorems 4.1 and 5.3 were applied in a nested way. The lower bound for repetition codes is achieved if both channels are BSCs, and the lower bound for SPC codes is achieved if

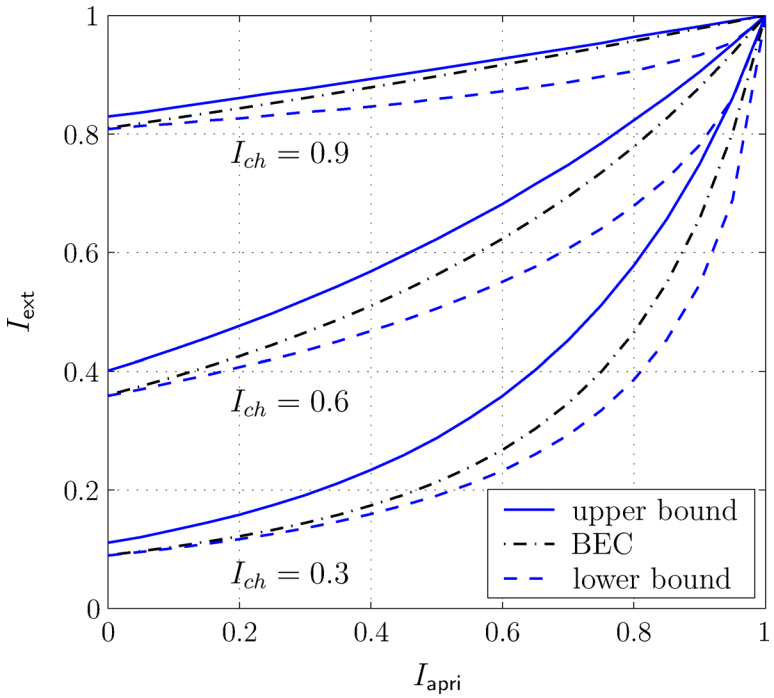

Fig. 6.18 Bounds on the EXIT functions of the accumulator and exact EXIT functions for BECs; several values of channel information I_{ch}.(Mutual information is given in bit/use.)

both channels are BECs. Thus, if the two lower bounds are applied in a nested way, the assumptions contradict each other and the resulting lower bound may be too pessimistic. Similarly, this holds for a nested upper bound, and thus the upper bound may be too optimistic. Therefore, the bounds given in Theorem 6.2 are not tight and may have a potential for improvement.

The extrinsic information can be computed exactly when both the communication channel and the a-priori channel are BECs. To do so, the same approach is used as above, but each time Theorems 4.1 and 5.3 are applied, the expression corresponding to the case of BECs are applied. This results in the stationarity condition

$$I_{\alpha}^{\mathsf{BEC}} = I_{\mathsf{ch}} + I_{\mathsf{apri}} \cdot I_{\alpha}^{\mathsf{BEC}} - I_{\mathsf{ch}} \cdot I_{\mathsf{apri}} \cdot I_{\alpha}^{\mathsf{BEC}}, \tag{6.41}$$

corresponding to (6.38) and (6.39). After solving for I_α and applying Theorem 4.1, one obtains

$$I_{\mathsf{ext}}^{\mathsf{BEC}} = \big(I_\alpha^{\mathsf{BEC}}\big)^2 = \Big(\frac{I_{\mathsf{ch}}}{1-(1-I_{\mathsf{ch}})I_{\mathsf{apri}}}\Big)^2. \tag{6.42}$$

This method corresponds to the one used in [59] (cf. also references therein).

These EXIT functions are also shown in Fig. 6.18. When the channel information I_{ch} is small, the BEC curves are close to the lower bounds for small I_{apri}, and they are close to the upper bounds for large I_{apri}. Thus, the case where channels are all BECs may not be an extreme case for the accumulator, as opposed to SPC codes and repetition codes.

7

Conclusions

This book describes the concept, the theory, and some applications of bounding combined mutual information. The scope was limited to binary-input symmetric memoryless channels (BISMCs) without feedback, and to parity-check constraints and equality constraints on the inputs of the channels.

The starting point was properties of BISMCs. Using subchannel indicators, BISMCs can be decomposed into subchannels that are all binary symmetric channels (BSCs). Each of these subchannels is associated to a value of mutual information, called the subchannel mutual information, and thus, the BISMC can be characterized by the distribution of these values of subchannel mutual information. This distribution is called the mutual information profile (MIP) of the BISMC.

The concept of decomposing the overall channel into BSCs and the concept of mutual information profiles was then used to bound combined mutual information. The two situations addressed are parallel independent BISMCs, of which the channel input symbols are either coupled by a parity-check constraint (corresponding to single parity-check codes) or by an equality constraint (corresponding to repetition codes). The combined mutual information turned out to take on the

extreme values if the channels are BSCs and if the channels are BECs. Remarkably, these two kinds of channels are also the BISMCs with the minimum-variance and the maximum-variance mutual information profiles. The bounds on information combining are proved using an extension of the well-known Jensen's inequality.

The results on information combining were applied in several ways to illustrate this concept:

- Information processing characteristics, including the well-known EXIT charts, can be upper and lower bounded.
- Multiple turbo codes can be designed by combining the extrinsic information from several constituent decoders.
- The decoding thresholds of LDPC codes can be upper and lower bounded. The Gaussian assumption, used in the original EXIT chart method, is not required any more, only the assumption of symmetric channels is required.
- The EXIT function of the accumulator can be upper and lower bounded, again without Gaussian assumption, but only assuming symmetric channels.

This book was restricted to channels with binary inputs and to the two constraints given by single parity-check codes and repetition codes. A generalization to non-binary channels or to other kinds of constraints at the channel inputs would certainly be of great interest. Furthermore, the methods applied to derive the EXIT function of the accumulator lead to bounds that are not tight. An improvement of such bounds, or even a generalization to other convolutional codes may also be a topic for future research.

There are several other directions to extend the concept of bounding combined information, as already mentioned in the introduction. In [32], the channels are characterized not only by their mutual information, but also by the bit-error probability. Thus, not only one parameter of the channel, but two parameters are used. When applying this to determine decoding thresholds for LDPC codes, the resulting bounds become tighter. In [28], again two parameters are tracked to determine bounds on the decoding threshold of LDPC codes, the expected "soft-bit" (expected conditional probability) and the Bhattacharyya

parameter; furthermore, non-binary codes are addressed. In [61], the authors consider the basic problem of information combining, namely the case of two parallel channels, and study moments of "soft-bits" for this scenario.

The concept of information combining is still young. So it is highly likely that new extensions and new applications will come up in the near future. This tutorial may stimulate research into this direction.

Acknowledgments

The authors would like to thank the anonymous reviewers. Their thorough and detailed comments helped to substantially improve this book. Furthermore we would like to thank our colleagues Peter Adam Hoeher and Simon Huettinger for the discussions about this topic and the contributions to this book.

A

Binary Information Functions

The binary information functions for serial and parallel concatenation are introduced in Definitions 4.1 and 5.1. This chapter provides some more explanation.

A.1 Serially Concatenated BSCs

Consider N binary symmetric channels (BSCs) $X_i \to Y_i$, $X_i, Y_i \in \mathbb{F}_2$, $i = 0, 1, \ldots, N-1$, which are serially concatenated such that $Y_i = X_{i+1}$ for $i = 0, 1, \ldots, N-2$:

$$X_0 \to Y_0 = X_1 \to Y_1 = X_2 \to \cdots \to Y_{N-2} = X_{N-1} \to Y_{N-1}.$$

The input of the first channel is assumed to be uniformly distributed. The mutual information of each individual channel is denoted by $I_i := I(X_i;Y_i)$, $i = 0, 1, \ldots, N-1$. The end-to-end mutual information between the input of the first and the output of the last channel is denoted by $I := I(X_0;Y_{N-1})$.

It is now shown that the end-to-end mutual information is given by the binary information function for serial concatenation according to Definition 4.1, i.e.,

$$I = f_N^{\mathsf{ser}}(I_0, I_1, \ldots, I_{N-1}). \tag{A.1}$$

Notice that the mutual information I is equal to the channel capacity, as the serially concatenated channel $X_0 \to Y_{N-1}$ is symmetric.

We start with the case $N = 2$. It can easily be seen that the serially concatenated channel $X_0 \to Y_1$ is also a BSC. Let ϵ_{01} denote its crossover probability. We have an error on this channel if an error occurs either on the first or on the second channel. For the crossover probability of the first channel, we have

$$\epsilon_0 \in \Big\{h^{-1}(1 - I_0), 1 - h^{-1}(1 - I_0)\Big\},$$

and for the crossover probability of the second channel, we have

$$\epsilon_1 \in \Big\{h^{-1}(1 - I_1), 1 - h^{-1}(1 - I_1)\Big\}.$$

Consider first the case $\epsilon_0 = h^{-1}(1 - I_0)$ and $\epsilon_1 = h^{-1}(1 - I_1)$. Then, the crossover probability of the serially concatenated channel is given by

$$\epsilon_{01} = (1 - \epsilon_0)\epsilon_1 + \epsilon_0(1 - \epsilon_1).$$

This operation is called the convolution of the two probabilities ϵ_0 and ϵ_1 in [24, 25]. Thus, its mutual information is given by

$$\begin{aligned} I(X_0; Y_1) &= 1 - h(\epsilon_{01}) \\ &= 1 - h\Big((1 - \epsilon_0)\epsilon_1 + \epsilon_0(1 - \epsilon_1)\Big). \end{aligned}$$

For $\epsilon_0 = 1 - h^{-1}(1 - I_0)$ and $\epsilon_1 = 1 - h^{-1}(1 - I_1)$, the crossover probability of the serially concatenated channel remains ϵ_{01}; for the other two cases the crossover probability of the serially concatenated channel becomes $1 - \epsilon_{01}$. Since the binary entropy function is symmetric, the same end-to-end mutual information is obtained in all cases. This completes the proof of (A.1) for $N = 2$.

The generalization for $N > 2$ follows immediately by induction.

A.2 Parallel Concatenated BSCs

Consider N binary symmetric channels (BSCs) $X \to Y_i$, $X, Y_i \in \mathbb{F}_2$, $i = 0, 1, \ldots, N - 1$, that have the same input X. Following the accepted practice for parallel concatenated codes, see [62], we call these channels parallel concatenated. The input X is assumed to be uniformly

distributed. The mutual information of each channel is denoted by $I_i := I(X;Y_i)$, $i = 0, 1, \ldots, N-1$. The vector of channel outputs is written as $\boldsymbol{Y} := [Y_0, Y_1, \ldots, Y_{N-1}]$. The overall mutual information between the input and the vector of channel outputs is denoted by $I := I(X;\boldsymbol{Y})$.

It is now shown that the overall mutual information is given by the binary information function for parallel concatenation according to Definition 5.1, i.e.,

$$I = f_N^{\mathsf{par}}(I_0, I_1, \ldots, I_{N-1}). \tag{A.2}$$

Notice that the mutual information I is equal to the channel capacity, as the parallel concatenated channel $X \to \boldsymbol{Y}$ is symmetric.

To start with, we write the overall mutual information as

$$\begin{aligned} I &= I(X;\boldsymbol{Y}) \\ &= H(\boldsymbol{Y}) - H(\boldsymbol{Y}|X). \end{aligned}$$

The first term can be computed using the joint probabilities of the channel outputs,

$$\begin{aligned} H(\boldsymbol{Y}) &= \mathrm{E}\big\{-\mathrm{ld}\, p_{\boldsymbol{Y}}(\boldsymbol{y})\big\} \\ &= -\sum_{\boldsymbol{y}\in\mathbb{F}_2^N} p_{\boldsymbol{Y}}(\boldsymbol{y}) \cdot \mathrm{ld}\, p_{\boldsymbol{Y}}(\boldsymbol{y}) \end{aligned}$$

with

$$\begin{aligned} p_{\boldsymbol{Y}}(\boldsymbol{y}) &= \sum_{x\in\mathbb{F}_2} p_{X,\boldsymbol{Y}}(x,\boldsymbol{y}) \\ &= \sum_{x\in\mathbb{F}_2} p_X(x) \cdot p_{\boldsymbol{Y}|X}(\boldsymbol{y}|x) \\ &= \sum_{x\in\mathbb{F}_2} p_X(x) \cdot \prod_{i=0}^{N-1} p_{Y_i|X}(y_i|x). \end{aligned}$$

In the last line, we have used the conditional independence of the channel outputs for a given channel input. Notice that $p_X(x) = \frac{1}{2}$ due to the uniform input distribution. The transition probabilities of each channel can be expressed using its mutual information:

$$p_{Y_i|X}(y_i|x) \in \{\epsilon_i, 1-\epsilon_i\}$$

with

$$\epsilon_i := h^{-1}(1 - I_i),$$

$i = 0, 1, \ldots, N-1$. Thus, the joint probability of a vector of channel outputs y can be obtained according to

$$p_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{1}{2}\Big(\prod_{i=0}^{N-1} \varphi_i(y_i) + \prod_{i=0}^{N-1} \big(1 - \varphi_i(y_i)\big)\Big),$$

with

$$\varphi_i(y_i) := \begin{cases} \epsilon_i & \text{for } y_i = 0, \\ 1 - \epsilon_i & \text{for } y_i = 1. \end{cases}$$

The second term in (A.3) can be written as

$$H(\boldsymbol{Y}|X) = \sum_{i=0}^{N-1} H(Y_i|X) = \sum_{i=0}^{N-1} (1 - I_i),$$

where the conditional independence of the channel outputs for a given channel input has again been used.

By substituting the above equations into (A.3), we obtain the proof of (A.2).

B

Convexity Lemma

This appendix provides the proof of Lemma 4.2, i.e., of the convexity of the binary information function for serial concatenation, $f_2^{\mathsf{ser}}(j_1, j_2)$, $j_1, j_2 \in [0,1]$, according Definition 4.1. Notice that j_1 and j_2 correspond to values of mutual information.

The function $f_2^{\mathsf{ser}}(j_1, j_2)$ is symmetric in its two arguments, and thus it is sufficient to show that it is convex in j_1 for constant j_2. To show that, the function

$$\begin{aligned} g(x) &:= 1 - f_2^{\mathsf{ser}}\big(1-x, 1-h(a)\big) \\ &= h\big([1-2a]h^{-1}(x) + a\big), \end{aligned}$$

$x \in [0,1]$ and $a \in [0, \frac{1}{2}]$, is introduced. (The range of the parameter a is chosen such that $h^{-1}(h(a)) = a$.) Notice that x corresponds to an entropy, and a corresponds to a probability. Then, $f_2^{\mathsf{ser}}(j_1, j_2)$ is convex-$\cap$ in j_1 for any constant j_2 if and only if $g(x)$ is convex-$\cup$ in x for any a.

The convexity of $g(x)$ is proved in the following Lemma [13]. Alternatively, Mrs. Gerber's lemma [25] may be used [63].

Lemma B.1 The function

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

$x \in [0,1]$, $a \in [0, \frac{1}{2}]$, is convex-$\cup$.

Proof. The function is convex-$\cup$ if the second derivative of $g(x)$ with respect to x is non-negative, i.e., if

$$\frac{d^2 g(x)}{dx^2} \geq 0 \tag{B.1}$$

for $x \in [0,1]$ and $a \in [0, \frac{1}{2}]$.

First, this function is parametrized. Let $x = h(t)$, $t \in [0, \frac{1}{2}]$, and let $y = g(x)$. Then we have

$$y = g\big(h(t)\big) = h([1-2a]t + a). \tag{B.2}$$

The derivatives of x with respect to t are given as

$$\frac{dx}{dt} = h'(t) = \mathrm{ld}\,\frac{1-t}{t} \geq 0, \tag{B.3}$$

$$\frac{d^2 x}{dt^2} = h''(t) = -\frac{\mathrm{ld}\,e}{t(1-t)} \leq 0, \tag{B.4}$$

$$\frac{d^3 x}{dt^3} = h'''(t) = \frac{1-2t}{t^2(1-t)^2} \cdot \mathrm{ld}\,e. \tag{B.5}$$

For the first and the second derivative, the co-domains are also stated. Similarly, the derivatives of y with respect to t are given as

$$\frac{dy}{dt} = [1-2a] \cdot h'([1-2a]t + a) \geq 0, \tag{B.6}$$

$$\frac{d^2 x}{dt^2} = [1-2a]^2 \cdot h''([1-2a]t + a) \leq 0. \tag{B.7}$$

Again, the co-domains are also stated.

Consider now the relations between the derivatives with respect to x and the derivatives with respect to t. The first derivative can be written as

$$\frac{dy}{dt} = \frac{dy}{dx} \cdot \frac{dx}{dt},$$

and thus it follows that

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt}. \tag{B.8}$$

The second derivative can be written as

$$\begin{aligned}\frac{d^2y}{dt^2} &= \frac{d}{dt}\Big(\frac{dy}{dx}\Big)\cdot\frac{dx}{dt} + \frac{dy}{dx}\cdot\frac{d}{dt}\Big(\frac{dx}{dt}\Big)\\ &= \frac{d^2y}{dx^2}\cdot\frac{dx}{dt}\cdot\frac{dx}{dt} + \frac{dy}{dx}\cdot\frac{d^2x}{dt^2},\end{aligned}$$

and thus it follows that

$$\frac{d^2y}{dx^2} = \frac{\frac{d^2y}{dt^2} - \frac{dy}{dx}\cdot\frac{d^2x}{dt^2}}{\big(\frac{dx}{dt}\big)^2}.$$

Since $(dx/dt)^2 \geq 0$, we have

$$\begin{aligned}\frac{d^2y}{dx^2} \geq 0 &\Leftrightarrow \frac{d^2y}{dt^2} - \frac{dy}{dx}\cdot\frac{d^2x}{dt^2} \geq 0\\ &\Leftrightarrow \frac{d^2y}{dt^2} \geq \frac{dy}{dx}\cdot\frac{d^2x}{dt^2} = \frac{dy/dt}{dx/dt}\cdot\frac{d^2x}{dt^2}\\ &\Leftrightarrow \frac{d^2y/dt^2}{dy/dt} \geq \frac{d^2x/dt^2}{dx/dt},\end{aligned} \tag{B.9}$$

where (B.8) was applied in the second line, and $dy/dt \geq 0$ was used in the third line.

Using the expressions for the derivatives, (B.3), (B.4), (B.6), (B.7), in (B.9) yields

$$\frac{[1-2a]^2\cdot h''([1-2a]t+a)}{[1-2a]\cdot h'([1-2a]t+a)} \geq \frac{h''(t)}{h'(t)}. \tag{B.10}$$

Let

$$b(t) := \frac{h''(t)}{h'(t)} \tag{B.11}$$

and

$$c(a) := [1-2a]\cdot b([1-2a]t+a).$$

Then (B.10) can be written as

$$c(a) \geq c(0)$$

for $a \in [0, \frac{1}{2}]$. This relation holds if the first derivative of $c(a)$ with respect to a is non-negative.

Thus, we can give the following sufficient condition for (B.1):

$$\frac{dc(a)}{da} \geq 0 \tag{B.12}$$

for $a \in [0, \frac{1}{2}]$ and $t \in [0, \frac{1}{2}]$. This derivative can be computed as

$$\begin{aligned}\frac{dc(a)}{da} &= \frac{d}{da}([1-2a] \cdot b([1-2a]t + a)) \\ &= -2 \cdot b([1-2a]t + a) + [1-2a] \cdot b'([1-2a]t + a) \cdot (1-2t) \\ &= -2 \cdot b([1-2a]t + a) \\ &\quad + (1 - 2([1-2a]t + a)) \cdot b'([1-2a]t + a).\end{aligned}$$

After substituting $s := [1-2a]t + a$, (B.12) holds if and only if

$$-2 \cdot b(s) + (1-2s) \cdot b'(s) \geq 0 \tag{B.13}$$

for $s \in [t, \frac{1}{2}]$ and $t \in [0, \frac{1}{2}]$, and thus for $s \in [0, \frac{1}{2}]$. In (B.13), we first apply (B.11) and

$$b'(t) = \frac{db(t)}{dt} = \frac{h'''(t) \cdot h'(t) - \big(h''(t)\big)^2}{\big(h'(t)\big)^2},$$

and then we apply the expressions for $h(x)$ and its derivatives, (B.3), (B.4), (B.5). This gives the following equivalent relations:

$$\begin{aligned}(1-2s) \cdot b'(s) &\geq 2 \cdot b(s) \\ (1-2s) \cdot \frac{h'''(s) \cdot h'(t) - \big(h''(s)\big)^2}{\big(h'(s)\big)^2} &\geq 2 \cdot \frac{h''(s)}{h'(s)} \\ (1-2s) \cdot \Big[h'''(s) \cdot h'(t) - \big(h''(s)\big)^2\Big] &\geq 2 \cdot h''(s) \cdot h'(s) \\ (1-2s)\Big[\frac{1-2s}{s^2(1-s)^2} \cdot \operatorname{ld}\frac{1-s}{s} - \frac{\operatorname{ld} e}{s^2(1-s)^2}\Big] &\geq 2\frac{-1}{s(1-s)} \cdot \operatorname{ld}\frac{1-s}{s} \\ (1-2s)^2 \cdot \operatorname{ld}\frac{1-s}{s} - (1-2s) \cdot \operatorname{ld} e &\geq -2s(1-s) \cdot \operatorname{ld}\frac{1-s}{s}\end{aligned}$$

$$[(1-2s)^2 + 2s(1-s)]\,\mathrm{ld}\,\frac{1-s}{s} \geq (1-2s)\cdot \mathrm{ld}\,e$$
$$\ln\frac{1-s}{s} \geq \frac{1-2s}{1-2s+2s^2}. \quad \text{(B.14)}$$

We substitute $u := (1-s)/s$. From $s \in [0, \frac{1}{2}]$, it follows that $u \in [1, \infty)$. Using $s = 1/(1+u)$ in (B.14), the left-hand side results as $\ln u$, and the right-hand side results as

$$\frac{1 - 2\frac{1}{1+u}}{1 - 2\frac{1}{1+u} + 2(\frac{1}{1+u})^2} = \frac{(1+u)^2 - 2(1+u)}{(1+u)^2 - 2(1+u) + 2} = \frac{u^2-1}{u^2+1}.$$

Thus, (B.1) holds if

$$\ln u \geq \frac{u^2-1}{u^2+1}$$

for $u \in [1, \infty)$. Since equality holds for $u = 1$, it is sufficient to show that

$$\frac{d}{du}\ln u \geq \frac{d}{du}\Big(\frac{u^2-1}{u^2+1}\Big). \quad \text{(B.15)}$$

The left-hand side results as $1/u$, and the right-hand side results as

$$\frac{d}{du}\Big(\frac{u^2-1}{u^2+1}\Big) = \frac{2u(u^2+1) - (u^2-1)2u}{(u^2+1)^2} = \frac{4u}{(u^2+1)^2}.$$

Therefore (B.15) can equivalently be written as

$$\frac{1}{u} \geq \frac{4u}{(u^2+1)^2}$$
$$\Leftrightarrow\ (u^2+1)^2 \geq 4u^2$$
$$\Leftrightarrow\ (u^2-1)^2 \geq 0.$$

To sum up, a sufficient condition for (B.1) is

$$(u^2-1)^2 \geq 0$$

for $u \in [1, \infty)$. Since this is the case, we have the proof. □

C

Acronyms

AWGN	additive white Gaussian noise
BEC	binary erasure channel
BEP	bit error probability
BIAWGNC	binary-input AWGN channel
BISMC	binary-input symmetric memoryless channel
BSC	binary symmetric channel
BSEC	binary symmetric error and erasure channel
EXIT	extrinsic information transfer
IPC	information processing characteristic
LDPC	low-density parity-check
MIP	mutual information profile
RA	repeat-accumulate (code)
SNR	signal-to-noise ratio

References

[1] J. Huber and S. Huettinger, "Information Processing and Combining in Channel Coding," in *Int. Symp. on Turbo Codes & Rel. Topics*, Brest, France, September 2003, pp. 95–102.

[2] 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, October 2001.

[3] S. Huettinger, "Analysis and Design of Power-Efficient Coding Schemes," Ph.D. dissertation, University Erlangen-Nürnberg, Germany, 2004.

[4] R. Gallager, "Low-Density Parity-Check Codes," *IEEE Trans. Inform. Theory*, vol. 8, no. 1, pp. 21–28, January 1962.

[5] D. J. MacKay, "Good Error-Correcting Codes Based on Very Sparse Matrices," *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 399–431, March 1999.

[6] S. ten Brink, "Convergence Behavior of Iteratively Decoded Parallel Concatenated Codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, October 2001.

[7] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of Low-Density Parity-Check Codes for Modulation and Detection," *IEEE Trans. Inform. Theory*, vol. 52, no. 14, pp. 670–678, April 2004.

[8] 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, September 2003, pp. 39–42.

[9] Ingmar Land, Simon Huettinger, Peter A. Hoeher, and Johannes Huber, "Bounds on Information Combining," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 612–619, February 2005.

[10] I. Land, S. Huettinger, P. A. Hoeher, and J. Huber, "Bounds on Mutual Information for Simple Codes Using Information Combining," *Ann. Télécommun.*, vol. 60, no. 1/2, pp. 184–214, January/February 2005.

[11] I. Land, P. A. Hoeher, and J. Huber, "Analytical Derivation of EXIT Charts for Simple Block Codes and for LDPC Codes Using Information Combining," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Vienna, Austria, September 2004.

[12] ——, "Bounds on Information Combining for Parity-Check Equations," in *Proc. Int. Zurich Seminar on Communications (IZS)*, Zurich, Switzerland, February 2004, pp. 68–71.

[13] I. Land, "Reliability Information in Channel Decoding – Practical Aspects and Information Theoretical Bounds," Ph.D. dissertation, University of Kiel, Germany, 2005. [Online]. Available: http://e-diss.uni-kiel.de/diss_1414/

[14] I. Sutskover, S. Shamai (Shitz), and J. Ziv, "Extremes of Information Combining," in *Proc. Allerton Conf. on Communications, Control, and Computing*, Monticello, Illinois, USA, October 2003.

[15] ——, "Extremes of Information Combining," *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1313–1325, April 2005.

[16] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.

[17] S. Huettinger and J. Huber, "Extrinsic and Intrinsic Information in Systematic Coding," in *Proc. Ieee Int. Symp. Inform. Theory (ISIT)*, Lausanne, Switzerland, June 2002, p. 116.

[18] C. Berrou and A. Glavieux, "Near Optimum Error Correcting Coding and Decoding: Turbo-Codes," *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1261–1271, October 1996.

[19] Johannes Huber, "Grundlagen der Wahrscheinlichkeitsrechnung für iterative Decodierverfahren," *e&i: Elektrotechnik und Informationstechnik*, vol. 119, no. 11, pp. 386–394, November 2002.

[20] 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, January 2002, pp. 81–90.

[21] S. Huettinger and J. Huber, "Performance Estimation for Concatenated Coding Schemes," in *Proc. IEEE Inform. Theory Workshop*, Paris, France, March/April 2003, pp. 123–126.

[22] ——, "Design of Multiple–Turbo–Codes with Transfer Characteristics of Component Codes," in *Proc. Conf. Inform. Sciences and Systems (CISS)*, Priniceton University, Princeton, NJ, USA, March 2002.

[23] ——, "Analysis and Design of Power Efficient Coding Schemes with Parallel Concatenated Convolutional Codes," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1251–1258, July 2006.

[24] N. Chayat and S. Shamai (Shitz), "Expansion of an Entropy Property for Binary Input Memoryless Symmetric Channels," *IEEE Trans. Inform. Theory*, vol. 35, no. 5, pp. 1077–1079, September 1989.

[25] 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, November 1973.

[26] J. Huber, I. Land, and P. A. Hoeher, "Information Combining: Models, Bounds and Applications — A Tutorial (Invited Talk)," in *Proc. IEEE Int. Symp. Inform. Theory and Its Applications (ISITA)*, Parma, Italy, October 2004.

[27] D. Burshtein and G. Miller, "Bounds on the Performance of Belief Propagation Decoding," *IEEE Trans. Inform. Theory*, vol. 48, no. 1, pp. 112–122, January 2002.

[28] C. Wang, S. Kulkarni, and H. Poor, "On Finite-Dimensional Bounds for LDPC-like Codes with Iterative Decoding," in *Proc. IEEE Int. Symp. Inform. Theory and its Applications (ISITA)*, Parma, Italy, October 2004.

[29] H. Jin, A. Khandekar, and R. J. McEliece, "Irregular Repeat-Accumulate Codes," in *Proc. Int. Symp. on Turbo Codes & Rel. Topics*, Brest, France, September 2000, pp. 1–8.

[30] K. Bhattad and K. Narayanan, "An MSE Based Transfer Chart to Analyze Iterative Decoding Schemes," in *Proc. Allerton Conf. on Communications, Control, and Computing*, Allerton House, Monticello, IL, USA, 2004.

[31] M. Tuechler, S. ten Brink, and J. Hagenauer, "Measures for Tracing Convergence of Iterative Decoding Algorithms," in *Proc. Int. ITG Conf. on Source and Channel Coding*, Berlin, Germany, January 2002, pp. 53–60.

[32] I. Sutskover, S. Shamai (Shitz), and J. Ziv, "Constrained Information Combining: Theory and Applications for LDPC Coded Systems," submitted to IEEE Trans. Inform. Theory.

[33] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* John Wiley & Sons, 1991.

[34] M. Hellman and J. Raviv, "Probability of Error, Equivocation, and the Chernoff Bound," *IEEE Trans. Inform. Theory*, vol. 16, no. 4, pp. 368–372, July 1970.

[35] 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, March 1996.

[36] J. Huber, personal communication, July 2004.

[37] 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, March 1974.

[38] P. Robertson, E. Villebrun, and P. Hoeher, "A Comparison of Optimal and Sub-Optimal MAP Decoding Algorithms Operating in the Log-Domain," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Seattle, USA, 1995, pp. 1009–1013.

[39] 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, March/April 1997.

[40] F. Jelinek, "A Fast Sequential Decoding Algorithm Using a Stack," *IBM J. Res. and Develop.*, vol. 13, pp. 675–685, February 1969.

[41] J. B. Anderson and S. Mohan, "Sequential Coding Algorithms: A Survey and Cost Analysis," *IEEE Trans. Inform. Theory*, vol. 32, pp. 169–176, February 1984.

[42] C. Kuhn and J. Hagenauer, "Iterative List-Sequential (LISS) Detector for Fading Multiple-Access Channels," in *Proc. IEEE Globecom Conf.*, Dalles, Texas, USA, November/December 2004.

[43] S. ten Brink, "Design of Serially Concatenated Codes Based on Iterative Decoding Convergence," in *Proc. Int. Symp. on Turbo Codes & Rel. Topics*, Brest, France, September 2000, pp. 319–322.

[44] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic Information Transfer Functions: Model and Erasure Channel Properties," *IEEE Trans. Inform. Theory*, 2004.

[45] Ingmar Land and Johannes Huber, "Information Processing in Ideal Coding Schemes with Code-Symbol Decoding," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Adelaide, Australia, September 2005.

[46] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon Limit Error-Correcting Coding and Decoding: Turbo Codes," *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 1064–1070, May 1993.

[47] S. ten Brink, "Code Characteristic Matching for Iterative Decoding of Serially Concatenated Codes," *Ann. Télécommun.*, vol. 56, no. 7–8, pp. 394–408, 2001.

[48] F. Brännström, "Convergence Analysis and Design of Multiple Concatenated Codes," Ph.D. dissertation, Chalmers University of Technology, Göteborg, Sweden, 2004.

[49] H.-A. Loeliger, "An Introduction to Factor Graphs," *IEEE Signal Processing Mag.*, vol. 21, no. 1, pp. 28–41, January 2004.

[50] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor Graphs and the Sum-Product Algorithm," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 498–519, February 2001.

[51] T. J. Richardson and R. L. Urbanke, "The Capacity of Low-Density Parity-Check Codes under Message-Passing Decoding," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 599–618, Feb 2001.

[52] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of Capacity-Approaching Irregular Low-Density Parity-Check Codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 619–637, February 2001.

[53] R. G. Gallager, "Low Density Parity Check Codes," Ph.D. dissertation, Cambridge, MA, USA, 1963.

[54] D. J. MacKay and R. Neal, "Near Shannon Limit Performance of Low Density Parity Check Codes," *Electron. Lett.*, vol. 33, no. 6, pp. 457–458, March 1997.

[55] S.-Y. Chung, T. J. Richardson, and R. L. Urbanke, "Analysis of Sum-Product Decoding of Low-Density Parity-Check Codes Using a Gaussian Approximation," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 657–670, February 2001.

[56] E. Sharon, A. Ashikhmin, and S. Litsyn, "EXIT Functions for Binary Input Memoryless Symmetric Channels," *IEEE Trans. Inform. Theory*, vol. 54, no. 7, pp. 1207–1214, July 2006.

[57] D. Divsalar, H. Jin, and R. J. McEliece, "Coding Theorems for Turbo-Like Codes," in *Proc. Allerton Conf. on Communications, Control, and Computing*, Allerton House, Monticello, IL, USA, September 1998, pp. 201–210.

[58] D. Divsalar, S. Dolinar, and F. Pollara, "Low Complexity Turbo-like Codes," in *Proc. Int. Symp. on Turbo Codes & Rel. Topics*, Brest, France, September 2000, pp. 73–80.

[59] S. ten Brink and G. Kramer, "Design of Repeat-Accumulate Codes for Iterative Detection and Decoding," *IEEE Trans. Signal Processing*, vol. 51, no. 11, pp. 2764–2772, November 2003.

[60] I. Land, J. Sayir, and P. A. Hoeher, "Bounds on Information Combining for the Accumulator of Repeat-Accumulate Codes without Gaussian Assumption," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Chicago, USA, June/July 2004, p. 443.

[61] Y. Jiang, A. Ashikhmin, R. Koetter, and A. C. Singer, "Extremal Problems of Information Combining," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Adelaide, Australia, September 2005, pp. 1236–1240.

[62] 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, March 1996.

[63] S. Shamai (Shitz), personal communication, September 2003.