On Achieving an Asymptotically Error-Free Fixed-Point of Iterative Decoding for Perfect *A Priori* Information

Jörg Kliewer, *Senior Member, IEEE*, and Daniel J. Costello, Jr., *Life Fellow, IEEE*

***Abstract*—In this paper we provide necessary and sufficient conditions for constituent codes in (multiple) concatenated and graph-based coding schemes to achieve an asymptotically error-free iterative decoding fixed-point if the maximum possible *a priori* information is available. At least one constituent code in an iterative decoding scheme must satisfy these conditions in order to ensure an asymptotically vanishing bit error probability at the convergence point of the decoder. Our results are proved for arbitrary binary-input symmetric memoryless channels (BISMCs) and thus can be universally applied to many transmission scenarios. Specifically, using a factor graph framework, it is shown that non-inner codes in a serial concatenation or check nodes in generalized LDPC codes achieve perfect extrinsic information if and only if the minimum Hamming distance between codewords is two or greater. For inner codes in a serial concatenation, constituent codes in a parallel concatenation, or variable nodes in doubly-generalized LDPC codes the corresponding encoder condition for acquiring perfect extrinsic information is an infinite codeword weight for a weight-one input sequence. For this case we provide a general proof which holds for all linear encoders and BISMCs. We also show that these results can improve the performance of concatenated coding schemes.**

***Index Terms*—Extrinsic information transfer functions, iterative decoding, factor graphs, code concatenation.**

I. INTRODUCTION

METHODS of characterizing the transfer of mutual information through a soft-input soft-output (SISO) decoder or, more generally, through factor nodes in a factor graph representation, such as extrinsic information transfer (EXIT) charts [2], [3], information processing charts (IPCs) [4], density evolution [5], [6], and generalized EXIT (GEXIT) charts [7] have emerged as useful tools for predicting the convergence properties of iterative decoding schemes for low density parity check (LDPC) codes or concatenated codes in the asymptotic block length regime. In EXIT or GEXIT charts, each constituent decoder or specific type of node in a factor graph is associated with such a transfer function that relates the mutual information of the *a priori* input to the extrinsic mutual information at the decoder output. By combining the mutual information from all the transfer functions, the transfer of average mutual information between the individual decoders or factor nodes in a factor graph can be evaluated and the decoding convergence behavior can be assessed.

One important goal in designing a concatenated coding scheme is to arrive at a virtually error-free reconstruction at the convergence point of the iterative decoder. In fact, achieving perfect extrinsic information, i.e., the maximum mutual information between *a priori* bits and extrinsic soft-information, at the output of at least one constituent decoder represents a necessary and sufficient condition for obtaining an asymptotically vanishing bit error probability at the convergence point. This was already observed in [2] for the special case of a binary input AWGN channel. Conversely, this means that asymptotically error-free reconstruction cannot be achieved if perfect extrinsic information cannot be obtained for at least one constituent decoder at the point of convergence.

In this paper we address the required conditions for the constituent codes of (multiple) concatenated and graph-based coding schemes to achieve perfect extrinsic information if perfect *a priori* information is available. This is equivalent to an analysis of the fixed-point of iterative decoding for perfect *a priori* information. Our considerations are carried out for arbitrary binary-input symmetric memoryless channels (BISMCs) and thus can be universally applied to many transmission scenarios. A BISMC randomly maps the input symbols $X_i \in \{0,1\}$ into output symbols Y_i taken from a Q-ary set $\mathbb{Y} = \{0, 1, \ldots, Q-1\}$ according to the transition probabilities $P(Y = y|X = x)$. A channel is said to be symmetric if it can be decomposed into strongly symmetric subchannels [5].

First, we extend an earlier result for non-inner codes in serial concatenation, namely that a minimum Hamming distance of two or greater between codewords is necessary and sufficient in order to achieve perfect extrinsic information [8], to non-inner factor nodes in factor graphs, i.e., to factor nodes without access to channel information. This result holds for arbitrary linear and non-linear codes and *a priori* channels and is, for example, useful for improving the ability of joint source-channel decoding schemes to exploit residual source redundancy [9]–[11] or outer entropy encoding [12], [13], which often exhibit a minimum distance of only one between two source codewords. It also applies to check nodes in generalized LDPC codes[1], where generic linear codes can be

Manuscript received May 24, 2012; revised November 14, 2012 and February 6, 2013. The editor coordinating the review of this paper and approving it for publication was H. Pishro-Nik.

J. Kliewer is with the Klipsch School of Electrical and Computer Engineering, New Mexico State University, Las Cruces, NM 88003, USA (e-mail: jkliewer@nmsu.edu).

D. J. Costello, Jr. is with the Department of Electrical Engineering, University of Notre Dame, 275 Fitzpatrick Hall, Notre Dame, IN 46556, USA (e-mail: costello.2@nd.edu).

This work was partly supported by NSF grants CCF-0830650, CCF-0830651, CCF-0830666, and CCF-1017632. This paper was presented in part at the IEEE International Symposium on Information Theory, Seattle, WA, July 2006 [1].

Digital Object Identifier 10.1109/TCOMM.2013.042313.120359

[1]This also applies to doubly-generalized LDPC codes [14] where both check and variable nodes can be replaced by linear block codes.

used as check nodes, and thus extends the result of [14] for a binary erasure *a priori* channel to arbitrary BISMCs.

Second, we address the case of inner codes in a serial concatenation, constituent codes in a parallel concatenation, or variable nodes in doubly-generalized LDPC codes [14]. It has been shown in [15] for a serially concatenated coding scheme that recursive inner encoders are preferable since they always lead to an interleaver gain, in contrast to nonrecursive or block encoders. A similar observation has been made in [16] for the constituent encoders of a (parallel concatenated) turbo code. This can also be noticed by considering EXIT functions: If the sequence of information bits can be described as a uniformly distributed i.i.d. (independent and identically distributed) process, only recursive inner encoders in serial concatenation or recursive constituent encoders in parallel concatenation can achieve the maximum average mutual information of one bit between the information bits and the extrinsic soft-values if perfect *a priori* information is available. The same holds for inner factor nodes, i.e., for factor nodes with direct access to channel information. Note that in contrast to [15], [16], where the use of recursive encoders is motivated by an analysis of error probability for maximum likelihood (ML) decoding based on the weight distribution of concatenated codes, we show that the findings in [15], [16] also hold in the context of *iterative* decoding.

In addition, we present a formal proof for serial concatenation and arbitrary BISMCs that, for blocklengths tending to infinity, only inner encoders which generate an infinite output weight for a weight-one input sequence, such as recursive inner encoders, lead to perfect extrinsic information at the output of the corresponding SISO decoder. This fact was observed in [17] but no analysis was given. In contrast to [18], where only rate-1 convolutional codes are considered via an analysis of their state-space representation, our proof uses information theoretic considerations and universally holds for arbitrary code rates and linear encoders. This is also verified by the fact that these results can be equivalently formulated in a factor graph framework. As an example for our findings, we address iteratively-decoded bit interleaved coded modulation (BICM-ID), where, in order to eliminate the error floor associated with soft demapping, we consider adding an accumulator as a recursive precoder prior to the mapper. At the decoder, the soft demapper is then replaced by a SISO decoder working on the joint accumulator / mapper trellis.

By combining the results for both inner and non-inner codes, necessary and sufficient conditions for the existence of a fixed-point of iterative decoding for perfect *a priori* information are provided. The same conditions also hold for the existence of a fixed-point for density evolution by considering the equivalent proofs in a factor graph framework.

The paper is organized as follows. In Section II the underlying system model and notation is introduced. Section III addresses both the case of non-inner codes in a serial concatenation and non-inner factor nodes and proves, by using a factor graph framework, that a minimum distance of two or greater between codewords is necessary and sufficient for achieving perfect extrinsic information. As an example we consider iterative joint source-channel decoding of a Markov source. In Section IV the case of inner codes in a serial concatenation, constituent codes in a parallel concatenation, and inner factor nodes is addressed. Here we show that for a blocklength tending to infinity perfect extrinsic information is achieved if and only if the encoder provides an infinite output codeword weight for a weight-one information sequence. Furthermore, the finite blocklength case is addressed, where it is shown for the binary-input AWGN channel that a moderately large output weight for a weight-one input is sufficient to get close to the ideal case. Then our findings are applied to improve the performance of a BICM-ID-based transmission system. Finally, some conclusions are given in Section V.

II. System Model

In the following, random variables (r.v.'s) are denoted with upper case letters, and the corresponding realizations with lower case letters, such as A or a, respectively. Sequences are represented with upper or lower case bold letters, such as $\boldsymbol{A}$ or $\boldsymbol{a}$. Further, we use $p(\cdot)$ for probability density functions (pdfs) and $P(\cdot)$ for probabilities and probability mass functions (pmfs).

In the following we present two different constituent decoder models which will be employed in this paper, an *a priori* channel model and a model based on factor graphs and the sum-product algorithm (SPA).

A. A priori channel model of the constituent decoder

The underlying system model is depicted in Fig. 1 [3, Fig. 2]. For the iterative decoding of a non-inner code in a (multiple) serial concatenation, the connections marked with "1" are active in Fig. 1. In this case, $\boldsymbol{X}_c = \boldsymbol{X}_a$ and the communication channel is not used at all. All other scenarios for iterative decoding (inner serially concatenated code and (multiple) parallel concatenation) are modeled by Fig. 1 with the ganged switches in position "2". Note that the case where the connection marked with "1" is active also describes the situation for the check nodes of generalized LDPC codes [14], [19] (see also the description of the factor graph model in Section II-B).

Assume a binary sequence $\boldsymbol{B} = [B_1, B_2, \ldots, B_\ell, \ldots, B_K]$ of length K with r.v.'s B_ℓ and corresponding realizations $b_\ell \in \{0, 1\}$ is applied to a code $\mathcal{C}$ with code rate $R_c = K/N$. Specifically, for a joint source-channel coding scenario, the sequence $\boldsymbol{B}$ may be interpreted as the binary version of a length-L index sequence $\boldsymbol{I} = [I_1, I_2, \ldots, I_k, \ldots, I_L]$ with J-bit indices $I_k = i_k \in \mathbb{I}$, $\mathbb{I} = \{0, 1, \ldots, 2^J - 1\}$ and $K = LJ$. In the simplest case such an index may be obtained by J-bit quantization of a waveform source sample, but generally it represents an arbitrary J-bit source encoder parameter.

A binary codeword $\boldsymbol{X}_c = [X_{c,1}, X_{c,2}, \ldots, X_{c,\ell}, \ldots, X_{c,N}]$ of length N is generated by the encoder and transmitted over the communication channel. At the channel output the sequence $\boldsymbol{Y}_c = [Y_{c,1}, Y_{c,2}, \ldots, Y_{c,\ell}, \ldots, Y_{c,N}]$ is observed. The *a priori* channel (or extrinsic channel) models the *a priori* information used at each constituent decoder in an iterative decoding scheme. The input sequence $\boldsymbol{X}_a = [X_{a,1}, X_{a,2}, \ldots, X_{a,\ell}, \ldots, X_{a,M}]$ has realizations $x_{a,\ell}$ from the binary alphabet $\mathbb{X}_a$, where $M = N$ if in Fig. 1 the switch is in position "1" (i.e., in the case of a non-inner decoder in serial concatenation) and $M = K$ otherwise. The *a priori* channel observation is denoted by $\boldsymbol{Y}_a = [Y_{a,1}, Y_{a,2}, \ldots, Y_{a,\ell}, \ldots, Y_{a,M}]$. The sequence $\boldsymbol{A}$ contains the

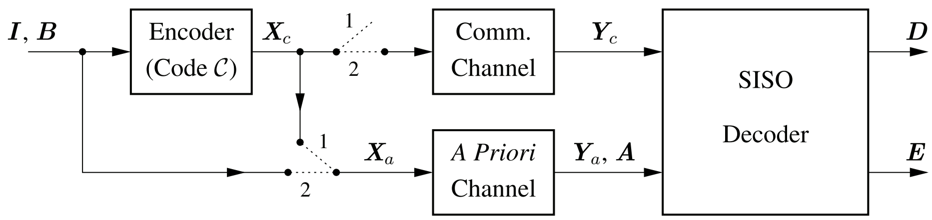

Fig. 1. Underlying *a priori* channel system model [3, Fig. 2].

corresponding *a priori* log-likelihood ratios (L-values) with elements

$$A_\ell \triangleq \ln\left(\frac{p(Y_{a,\ell}|X_{a,\ell}=0)}{p(Y_{a,\ell}|X_{a,\ell}=1)}\right), \quad \text{for} \quad \ell = 1,2,\ldots,M, \quad (1)$$

where the (*a priori*) channel probability density function (pdf) is $p(y_{a,\ell}|x_{a,\ell})$. We also define the sequences $\boldsymbol{v}_{a,\setminus\ell} \triangleq [v_{a,1},\ldots,v_{a,\ell-1},v_{a,\ell+1},\ldots,v_{a,N}]$ and $\boldsymbol{v}_{a,\ell\to q} \triangleq [v_{a,1},\ldots,v_{a,\ell-1},q,v_{a,\ell+1},\ldots,v_{a,N}]$ for any generic length N sequence $\boldsymbol{v}$.

The symbol-by-symbol SISO decoder employs both the outputs of the communication and *a priori* channels for computing *a posteriori* probabilities (APPs) $P(X_{a,\ell}=x_{a,\ell}|\boldsymbol{y}_c,\boldsymbol{y}_a)$ and extrinsic APPs $P(X_{a,\ell}=x_{a,\ell}|\boldsymbol{y}_c,\boldsymbol{y}_{a,\setminus\ell})$, $x_\ell \in \{0,1\}$, which follows the model in [3]. The sequences $\boldsymbol{D}$ and $\boldsymbol{E}$ comprise the corresponding conditional L-values and each element is defined as r.v.s

$$\begin{aligned} D_\ell &\triangleq \ln\left(\frac{P(X_{a,\ell}=0|\,\boldsymbol{Y}_c,\,\boldsymbol{Y}_a)}{P(X_{a,\ell}=1|\,\boldsymbol{Y}_c,\,\boldsymbol{Y}_a)}\right) \quad \text{and} \\ E_\ell &\triangleq \ln\left(\frac{P(X_{a,\ell}=0|\,\boldsymbol{Y}_c,\,\boldsymbol{Y}_{a,\setminus\ell})}{P(X_{a,\ell}=1|\,\boldsymbol{Y}_c,\,\boldsymbol{Y}_{a,\setminus\ell})}\right), \end{aligned} \quad (2)$$

resp., for $\ell = 1,2,\ldots,N$, where $\boldsymbol{Y}_c$ and $\boldsymbol{Y}_{a,\setminus\ell}$ denote the random vectors corresponding to the realizations $\boldsymbol{y}_c$ and $\boldsymbol{y}_{a,\setminus\ell}$.

Denoting the mutual information between two r.v.'s X and Y as $I(X;Y)$, we define the quantities

$$I_A \triangleq \frac{1}{N}\sum_{\ell=1}^{N} I(X_{a,\ell};A_\ell), \quad 0 \le I_A \le 1, \quad (3)$$

$$I_E \triangleq \frac{1}{N}\sum_{\ell=1}^{N} I(X_{a,\ell};E_\ell), \quad 0 \le I_E \le 1, \quad (4)$$

where I_A is defined as the average *a priori* information and I_E refers to the average extrinsic information, respectively. These quantities are related to the average sequence or word-wise mutual information per channel input bit by the expressions

$$\frac{1}{N}\,I(\boldsymbol{X}_a;\boldsymbol{A}) \le I_A, \quad \frac{1}{N}\,I(\boldsymbol{X}_a;\boldsymbol{E}) \le I_E. \quad (5)$$

By defining a continuous mapping between *a priori* and extrinsic information, we obtain the EXIT function (transfer characteristic) T for the SISO decoder in the model of Fig. 1 as $I_E = T(I_A, I_{\text{ch}})$, where $I_{\text{ch}} \triangleq I(\boldsymbol{X}_c;\,\boldsymbol{Y}_c)$.

B. Factor graph model

Optimal decoding for a channel code or redundant source code includes computing the marginals of a global function. The idea of factor graphs [20]–[22] is to alleviate the burden of this computation by breaking the global function up into smaller local functions [23]. Thus we obtain bipartite graphs which consist of factor nodes representing the computation of the local functions [24]. A well-known example is the factor graph representation of LDPC codes in conjunction with the SPA [5], [19], [21]. Here the factor nodes associated with the individual parity check equations described by the rows of the parity check matrix $\mathbf{H}$ are called check nodes, and the factor nodes associated with the "distribution" of the received channel information are called variable nodes, respectively. Factor graph representations are also known, e.g., for convolutional and turbo codes [20], for Markov sources [20], [25], and for non-linear codes such as variable-length codes [25], to which the SPA can be applied as well.

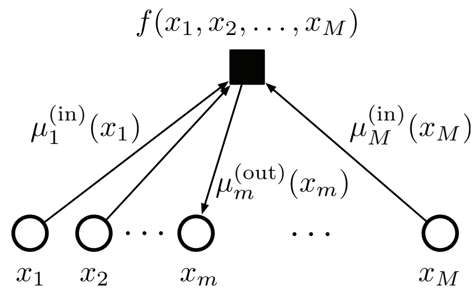

Fig. 2. Underlying factor graph model for a fixed choice of m.

In the following, we will employ a simple update rule for the SPA on a factor graph. Consider a factor node of degree M associated with a global function $f(x_1,x_2,\ldots,x_M)$, where x_1, x_2, ..., x_M are the M independent variables associated with this function. Note that $f(x_1,x_2,\ldots,x_M)$ can be factored into a product of local functions [24]. The SPA update rule states that the outgoing message along the edge directed towards x_m, $m=1,\ldots,M$, is given as

$$\mu_m^{(\text{out})}(x_m) = \gamma_m \sum_{x_1,\ldots,x_{m-1},x_{m+1},\ldots,x_M} f(x_1,x_2,\ldots,x_M) \prod_{\ell\neq m} \mu_\ell^{(\text{in})}(x_\ell), \quad (6)$$

where $\gamma_m > 0$ is a normalization factor. Herein, the messages $\mu_\ell^{(\text{in})}(\cdot)$ and $\mu_m^{(\text{out})}(\cdot)$ are defined as $\mu_\ell^{(\text{in})}(x_\ell) \triangleq p(y_{a,\ell}|x_{a,\ell})$ and $\mu_m^{(\text{out})}(x_m) \triangleq P(x_{a,m}|\boldsymbol{y}_{a,\setminus m},\boldsymbol{y}_c)$, and with this definition decoding on the factor graph corresponds to decoding on the SISO decoder in Fig. 1. The update rule in (6) is illustrated in Fig. 2.

III. Non-Inner Codes in Serial Concatenation

A. Condition for perfect extrinsic information

As a starting point, consider the average extrinsic information I_E at the output of the SISO decoder in Fig. 1 given by

$$I_E = \frac{1}{N}\sum_{\ell=1}^{N} I(X_{a,\ell}; E_\ell) = I_{E,\max} - \frac{1}{N}\sum_{\ell=1}^{N} H(X_{a,\ell}|E_\ell) \in [0,1], \quad (7)$$

where $I_{E,\max} \triangleq \frac{1}{N}\sum_{\ell=1}^{N} H(X_{a,\ell})$. It can be seen from (7) that in order to maximize I_E we require $H(X_{a,\ell}|E_\ell) = 0$ for all ℓ.

In [3], [26] it was shown that for a SISO decoder emitting APPs the mutual information terms $I(X_{a,\ell}; E_\ell)$ can also be written as

$$I(X_{a,\ell}; E_\ell) = I(X_{a,\ell}; \boldsymbol{Y}_{a,\setminus\ell}, \boldsymbol{Y}_c) = H(X_{a,\ell}) - H(X_{a,\ell}|\, \boldsymbol{Y}_{a,\setminus\ell}, \boldsymbol{Y}_c), \quad (8)$$

where $\boldsymbol{Y}_{a,\setminus\ell} = [Y_{a,1}, \ldots, Y_{a,\ell-1}, Y_{a,\ell+1}, \ldots, Y_{a,N}]$ describes the received sequence at the output of the *a priori* channel. Since for non-inner decoders in a serial concatenation the communication channel is not present in the model of Fig. 1, the conditional entropy $H(X_{a,\ell}|\, \boldsymbol{Y}_{a,\setminus\ell}, \boldsymbol{Y}_c)$ in (8) does not depend on $\mathbf{Y}_c$. Without loss of generality, we consider discrete output channels in the following[2]. Then, the conditional entropy $H(X_{a,\ell}|\, \boldsymbol{Y}_{a,\setminus\ell})$ may be further expanded as

$$H(X_{a,\ell}|\, \boldsymbol{Y}_{a,\setminus\ell}) = -\sum_{\boldsymbol{y}_{a,\setminus\ell}} \sum_{q=0}^{1} \sum_{\boldsymbol{x}_a \in \mathbb{X}_a^M : x_{a,\ell}=q} \left[P(\boldsymbol{y}_{a,\setminus\ell}|\boldsymbol{x}_a)\, P(\boldsymbol{x}_a)\right] \cdot \log_2\left(\frac{1}{P(\boldsymbol{y}_{a,\setminus\ell})} \sum_{\boldsymbol{x}_a \in \mathbb{X}_a^M : x_{a,\ell}=q} P(\boldsymbol{y}_{a,\setminus\ell}|\boldsymbol{x}_a)\, P(\boldsymbol{x}_a)\right), \quad (9)$$

where

$$P(\boldsymbol{y}_{a,\setminus\ell}) = \sum_{\boldsymbol{x}'_a \in \mathbb{X}_a^M : x'_{a,\ell}=0} P(\boldsymbol{y}_{a,\setminus\ell}|\boldsymbol{x}'_a)\, P(\boldsymbol{x}'_a) + \sum_{\boldsymbol{x}'_a \in \mathbb{X}_a^M : x'_{a,\ell}=1} P(\boldsymbol{y}_{a,\setminus\ell}|\boldsymbol{x}'_a)\, P(\boldsymbol{x}'_a). \quad (10)$$

From (8) we can see that, for blocklengths tending to infinity, $X_{a,\ell}$ can be estimated from $\boldsymbol{Y}_{a,\setminus\ell}$ (and $\boldsymbol{Y}_c$ for the case of inner decoders), with vanishing (bit) error probability P_e when $I_A = I_{A,\max}$, if and only if $H(X_{a,\ell}|\, \boldsymbol{Y}_{a,\setminus\ell}) = 0$. By applying Fano's inequality and considering that $x_{a,\ell} \in \{0,1\}$, P_e can be lower bounded as $P_e \geq h^{-1}\left(H(X_{a,\ell}|\, \boldsymbol{Y}_{a,\setminus\ell})\right)$, where $h(p) = -p\log(p) - (1-p)\log(1-p)$ is the binary entropy function for $p \in [0, 0.5]$. This shows that if $H(X_{a,\ell}|\, \boldsymbol{Y}_{a,\setminus\ell}) > 0$, asymptotically a non-zero error probability remains even when $I_A = I_{A,\max}$.

For (multiple) serial concatenation and a (not necessarily symmetric) binary-input memoryless *a priori* channel model (see Fig. 1), we restate the following theorem from [8], which provides a relationship between the minimum Hamming distance of the codewords $\boldsymbol{X}_a$ and the achievability of perfect extrinsic information at the output of a non-inner SISO decoder. In contrast to [8], where a proof based on the *a priori* channel model is given, we provide a new simpler proof which generalizes the validity of Theorem 1 to any non-inner factor node in a factor graph model employing the SPA, for example to the check nodes of generalized LDPC codes.

Theorem 1 ([8]): For any binary-input memoryless *a priori* channel, let $I_E = \frac{1}{N}\sum_{\ell=1}^{N} I(X_{a,\ell}; E_\ell)$ denote the average extrinsic information at the output of a non-inner SISO decoder in a (multiple) serial concatenation. Furthermore, let $I_{A,\max} = I_{E,\max} = \frac{1}{N}\sum_{\ell=1}^{N} H(X_{a,\ell})$ and let $d_{\min}$ denote the minimum Hamming distance between the 2^K N-bit codewords $\boldsymbol{X}_a = \boldsymbol{x}_a \in \mathbb{X}_a^M$ of the code $\mathcal{C}$. Then, $d_{\min} \geq 2$ is a necessary and sufficient condition such that $I_E(I_A = I_{A,\max}) = I_{E,\max}$ at the decoder output.

Proof: Consider a non-inner factor node of degree $M = N$ with associated variables $x_{a,1}, x_{a,2}, \ldots, x_{a,M}$, a corresponding function defined as

$$f(x_{a,1}, x_{a,2}, \ldots, x_{a,M}) = \begin{cases} 1 & \text{if } [x_{a,1}, x_{a,2}, \ldots, x_{a,M}] \in \mathcal{C}, \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

and the SPA update equation in (6), where $\mathcal{C}$ is a (not necessarily linear) binary code with block length M. Theorem 1 states that if

$$\mu_\ell^{(\text{in})}(x_{a,\ell}) = 1 \quad \text{and} \quad \mu_\ell^{(\text{in})}(x_{a,\ell\to(1-x_{a,\ell})}) = 0 \quad (12)$$

for all $\ell = 1, \ldots, M$, which is equivalent to $I_A = I_{A,\max}$, then $\mu_m^{(\text{out})}(q) = 1 - q$ for $q \in \{0,1\}$ and all $m = 1, \ldots, M$, which is equivalent to $H(X_{a,m}|\, \boldsymbol{Y}_{a,\setminus m}) = 0$, if and only if $d_{\min} \geq 2$.

By using the incoming messages in (12), the SPA update rule in (6) simplifies significantly, and we obtain the following outgoing messages:

$$\mu_m^{(\text{out})}(0) = f(x_{a,1}, \ldots, x_{a,m-1}, 0, x_{a,m+1}, \ldots, x_{a,M}),$$
$$\mu_m^{(\text{out})}(1) = f(x_{a,1}, \ldots, x_{a,m-1}, 1, x_{a,m+1}, \ldots, x_{a,M}).$$

If $d_{\min} \geq 2$ we have $\mu_m^{(\text{out})}(q) = 1 - q$ for $q \in \{0,1\}$ and all $m = 1, \ldots, M$ by considering the function definition in (11). However, if $d_{\min} = 1$ there exists at least one m for some codeword $[x_{a,1}, \ldots, x_{a,M}] \in \mathcal{C}$ such that $\mu_m^{(\text{out})}(0) = \mu_m^{(\text{out})}(1) = 0.5$ after normalization. ■

Remark: The theorem can also be shown to hold for non-binary channel input alphabets in a straightforward way by modifying the incoming and outgoing messages in the SPA appropriately.

B. Implications for joint source-channel coding

In general, $d_{\min} \geq 2$ holds for non-inner encoders in serial concatenation. However, Theorem 1 has important consequences for joint source-channel coding schemes with an outer source code and an inner channel encoder, as discussed in the following two cases.

The first case considers a variable-length code (VLC) as a (nonlinear) mapping $\mathcal{C}$ in Fig. 1. Here, the free distance of a VLC is defined as the minimum Hamming distance $d_{\min}$ between all possible codeword combinations of any length [27]. A free distance of $d_{\text{free}} = 1$ occurs, for example, if $\mathcal{C}$

[2] For continuous-output channels the probability mass functions (pmfs) need to be replaced by probability density functions (pdfs) and the sums over the channel outputs by integrals, respectively.

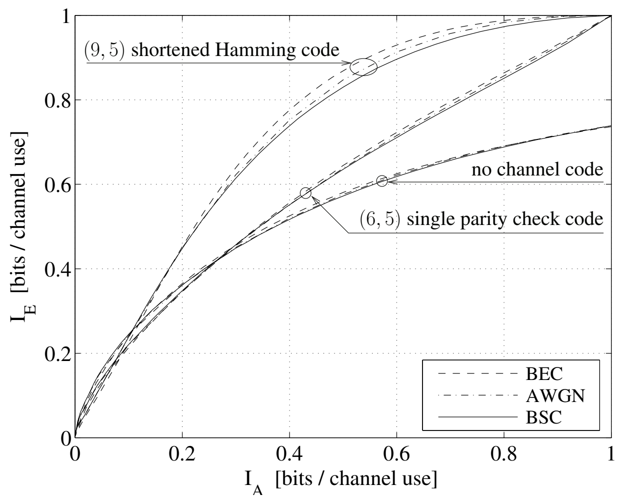

Fig. 3. EXIT functions for uniform scalar quantization with $J = 5$ bits of a first-order autoregressive source with correlation coefficient $a = 0.95$ and different (n, k) block codes $\mathcal{C}$.

is chosen as a Huffman VLC, and it follows that $I_E(I_A = 1\,\text{bit}) = 1\,\text{bit}$ cannot be achieved by a VLC SISO decoder [13]. A solution in this case is to insert explicit redundancy into the VLC codewords, as in the case of reversible VLCs [28], leading to $d_{\text{free}} = 2$ and to $I_E(I_A = 1\,\text{bit}) = 1\,\text{bit}$.

The second case involves the use of the implicit source redundancy inherent in the sequence $\boldsymbol{I}$ for error correction, e.g., when $\boldsymbol{I}$ is emitted from a Markov source. Transmitting uncoded source indices over the *a priori* channel in the absence of a code $\mathcal{C}$ leads to $d_{\text{min}} = 1$. Thus, according to Theorem 1, the maximum possible average extrinsic information cannot be obtained at the decoder output by solely exploiting residual source redundancy, even if there is a strong correlation between adjacent source indices I_k and I_{k-1}. However, a code $\mathcal{C}$ can be additionally used to increase the distance between codewords, as can be seen from the following example.

Example 1: Fig. 3 shows EXIT functions for a first-order Markov process $\boldsymbol{I}$, which is obtained by uniform scalar quantization with $J = 5$ bits of a first-order autoregressive process with correlation coefficient $a = 0.95$. The code $\mathcal{C}$ is either absent or chosen as an (n, k) block code, where n and k denote the length of a single code block and information block, respectively. A natural mapping from quantizer reconstruction levels to J-bit source indices with $J = k$ is used in all experiments, leading to $I_{E,\text{max}} = 1\,\text{bit}$. The resulting EXIT functions are shown in Fig. 3, where the block code and the Markov source with known index transition probabilities $P(I_k = i_k | I_{k-1} = i_{k-1})$ are jointly decoded using an index-based BCJR algorithm [29]. The overall rate R is defined as $R = R_s \cdot R_c$, with $R_s = H(I_k|I_{k-1})/J$ denoting the rate contribution from the residual source redundancy. We can observe that, despite the strong source correlation resulting in a source coding rate of $R_s = 0.47$, $d_{\text{min}} = 1$ implies that perfect average extrinsic information cannot be achieved even if perfect *a priori* information is applied at the decoder input. If d_{min} is increased to two, e.g., by adding a single parity check to the source indices, we can observe from Fig. 3 that $I_E(I_A = 1\,\text{bit}) = 1\,\text{bit}$, which is consistent with Theorem 1. A further reduction of the code rate by using a shortened $(9, 5)$ Hamming code leads to a larger area $\mathscr{A}$ under the EXIT function compared to the $(6, 5)$ single parity check code.

Note that for a memoryless source, a BEC *a priori* channel[3], and $H(X_{a,\ell}) = 1\,\text{bit}$, it is shown in [3, Theorem 1] that the area $\mathscr{A}$ is connected to the total code rate R by the expression $\mathscr{A} = 1 - R$. In general, for an L-th order Markov source $\boldsymbol{I}$, an (n, k) block code, and a BEC *a priori* channel, the area A under the EXIT function can be easily obtained from [3, Theorem 1] as

$$\mathscr{A} = I_{A,\max}^2 \left(1 - \frac{H(\mathcal{I})}{\sum_{\ell=1}^{n} H(X_{a,\ell})}\right), \tag{13}$$

where $H(\mathcal{I})$ denotes the entropy rate of $\boldsymbol{I}$. Here, the total code rate R includes both the rate contribution of the block code and the residual redundancy of the Markov source (see, e.g., [30]–[32]). ■

IV. Parallel Concatenation and Inner Codes in Serial Concatenation

A. Condition for perfect extrinsic information

For the constituent encoders in parallel concatenation and for the inner encoder in serial concatenation, the output $\boldsymbol{Y}_c$ of the communication channel must be considered. In the model of Fig. 1, the switch is in position "2" since the constituent decoder has *a priori* information on the information bits. In the following we assume linear encoders $\mathcal{C}$.

By considering the bit-wise extrinsic information in (8) for a continuous-output communication channel and assuming perfect *a priori* information $I_A = I_{A,\text{max}}$ for a given distribution of the source bits, the conditional entropy can be written as

$$H(X_{a,\ell}|\,\boldsymbol{Y}_{a,\setminus\ell},\,\boldsymbol{Y}_c)\Big|_{I_A=I_{A,\max}} = H(X_{a,\ell}|\boldsymbol{X}_{a,\setminus\ell},\,\boldsymbol{Y}_c) \tag{14}$$

$$= -\sum_{\boldsymbol{y}_c}\sum_{\boldsymbol{x}_{a,\setminus\ell}} p(\boldsymbol{y}_c,\boldsymbol{x}_{a,\setminus\ell}) \sum_{q=0}^{1} P(X_{a,\ell}=q|\boldsymbol{x}_{a,\setminus\ell},\boldsymbol{y}_c)\cdot \log_2(P(X_{a,\ell}=q|\boldsymbol{x}_{a,\setminus\ell},\boldsymbol{y}_c)). \tag{15}$$

We now show that $H(X_{a,\ell}|\,\boldsymbol{Y}_{a,\setminus\ell},\,\boldsymbol{Y}_c)\Big|_{I_A=I_{A,\max}} = 0$ can be obtained for inner recursive convolutional codes in serial concatenation and for recursive constituent convolutional encoders in parallel concatenation. The following lemma is an extension of the result in [1] for the binary-input AWGN (BIAWGN) channel to arbitrary BISMCs.

Lemma 1: Let the all-zero information word $\boldsymbol{x}_a^{(0)}$ of length K be applied to a linear encoder $\mathcal{C}$ of rate $R_c = K/N$. Let the resulting BPSK-modulated codeword $\boldsymbol{c}_0^{(0)} = 1 - 2\,\mathscr{C}(\boldsymbol{x}_a^{(0)})$ (an all '+1' sequence), where $\mathscr{C}(\boldsymbol{x}_a)$ denotes the mapping induced by code $\mathcal{C}$, be transmitted over a BISMC with capacity C, leading to the channel observation $\boldsymbol{y}_c$. All information bits except the one at bit position ℓ', $\ell' = 1, 2, \ldots, K$, are assumed to be perfectly known at the decoder. Then the fact that $\mathcal{C}$ generates an infinite output weight for a weight-one input sequence $\boldsymbol{x}_{a,\ell\to 1}^{(0)}$, for $K \to \infty$, $N \to \infty$, and fixed R_c, represents a necessary and sufficient condition for the

[3]An extension of the area theorem to all BISMCs is given in [7].

following APP evaluations for $N \to \infty$:

$$P(X_{a,\ell}^{(0)} = 0|\boldsymbol{x}_{a,\backslash\ell}^{(0)}, \boldsymbol{y}_c) = 1, \quad P(X_{a,\ell}^{(0)} = 1|\boldsymbol{x}_{a,\backslash\ell}^{(0)}, \boldsymbol{y}_c) = 0,$$
for $0 < C \le 1$,
$$P(X_{a,\ell}^{(0)} = 0|\boldsymbol{x}_{a,\backslash\ell}^{(0)}, \boldsymbol{y}_c) = \frac{1}{2}, \quad P(X_{a,\ell}^{(0)} = 1|\boldsymbol{x}_{a,\backslash\ell}^{(0)}, \boldsymbol{y}_c) = \frac{1}{2},$$
for $C = 0$.

Proof: In the following we state the proof based on the *a priori* channel model in Fig. 1. We need a result from [33], [34] stating that every BISMC with input $x \in \{0, 1\}$ and a discrete or continuous output y can be decomposed into binary symmetric subchannels (BSCs).

This can be obtained by decomposing the output alphabet $\mathbb{Y}$ of a BISMC into disjoint subsets $\mathbb{Y}(z)$ with one or two elements. The subsets $\mathbb{Y}(z)$ represent the output alphabets of the individual BSCs associated with a realization $Z = z \in \mathbb{Z}$ of a so called *subchannel indicator* Z independent of X, and $\mathbb{Z}$ denotes the alphabet of z. For channels with discrete outputs, Z is a discrete r.v. and the channel pmf can be decomposed as

$$P(y|x) = P(y', z|x) = P(z)\, P(y'|x, z), \quad z \in \mathbb{Z},\ y \in \mathbb{Y},\ y' \in \mathbb{Y}(z),\ x \in \{0, 1\}, \quad (16)$$

where $P(y'|x, z) = P(y'|x)$ is the channel pmf of the BSC associated with $Z = z$. We assume non-zero selection probabilities $P(z)$, which means that $\mathbb{Z}$ contains only subchannels with non-zero capacity. For BISMCs with continuous outputs (such as the BIAWGN channel), Z is a continuous r.v. and the pmfs in (16) are replaced by pdfs[4]. Eq. (16) holds as well if x, y, and z are replaced by the sequences $\boldsymbol{x} = [x_1, x_2, \ldots, x_N]$, $\boldsymbol{y} = [y_1, y_2, \ldots, y_N]$, and $\boldsymbol{z} = [z_1, z_2, \ldots, z_N]$.

For the sake of simplicity, the following proof addresses BISMCs with discrete outputs (BISDMCs). By employing Bayes' theorem and assuming equiprobable information sequences we can write

$$P(X_{a,\ell}^{(0)} = q|\boldsymbol{x}_{a,\backslash\ell}^{(0)}, \boldsymbol{y}_c) = \frac{P(\boldsymbol{y}_c|\boldsymbol{x}_{a,\ell\to q}^{(0)})}{P(\boldsymbol{y}_c|\boldsymbol{x}_a^{(0)}) + P(\boldsymbol{y}_c|\boldsymbol{x}_{a,\ell\to 1}^{(0)})}, \quad (17)$$

$q \in \{0, 1\}$. By inserting the sequence-based version of (16) in (17) we obtain, for the all-zero input hypothesis,

$$P(X_{a,\ell}^{(0)} = 0|\boldsymbol{x}_{a,\backslash\ell}^{(0)}, \boldsymbol{y}_c) = \frac{1}{1 + \dfrac{P(\boldsymbol{y}_c|\boldsymbol{x}_{a,\ell\to 1}^{(0)}, \boldsymbol{z}_c)}{P(\boldsymbol{y}_c|\boldsymbol{x}_a^{(0)}, \boldsymbol{z}_c)}}, \quad (18)$$

where $\boldsymbol{z}_c$ represents the subchannel indicator sequence for the communication channel in Fig. 1.

First consider an arbitrary channel observation $\boldsymbol{y}_c$ of length N at the output of the BISDMC. Each element $y_{c,\ell}$, $\ell = 1, \ldots, N$, is associated with a specific BSC subchannel with subchannel indicator $z_{c,\lambda}$ and transition probability $\rho(z_{c,\lambda})$, $\lambda = 1, \ldots, |\mathbb{Z}|$, where $|\mathbb{Z}|$ denotes the cardinality of the set $\mathbb{Z}$. Thus it suffices to show that $\lim_{N\to\infty} P(X_{a,\ell}^{(0)} = 0|\boldsymbol{x}_{a,\backslash\ell}^{(0)}, \boldsymbol{y}_c) = 1$ holds asymptotically for all BSCs with arbitrary transition probabilities $\rho(z_{c,\lambda}) \in (0, 1/2)$.

[4]Some examples are given, e.g., in [35, Chapter 2].

Now, for each BSC subchannel $z_{c,\lambda}$ of the BISDMC with transition probability $\rho(z_{c,\lambda})$, consider the following error sequences:

$$\boldsymbol{e}_{0,\lambda} = \boldsymbol{y}_{c,\lambda} \oplus \boldsymbol{x}_a^{(0)},\ \ \boldsymbol{e}_{1,\lambda} = \boldsymbol{y}_{c,\lambda} \oplus \boldsymbol{x}_{a,\ell\to 1}^{(0)} = \boldsymbol{x}_a^{(0)} \oplus \boldsymbol{x}_{a,\ell\to 1}^{(0)} \oplus \boldsymbol{e}_{0,\lambda}, \quad (19)$$

where '$\oplus$' indicates binary addition and where $\boldsymbol{y}_{c,\lambda}$ contains the part of the observed sequence $\boldsymbol{y}_c$ at the output of the BISDMC which corresponds to the subchannel $\lambda \in 1, 2, \ldots, |\mathbb{Z}|$. Let us also define N_λ as the number of symbols in $\boldsymbol{y}_{c,\lambda}$, where the block length $N = \sum_{\lambda=1}^{|\mathbb{Z}|} N_\lambda$. By using $\boldsymbol{d} \triangleq \boldsymbol{x}_a^{(0)} \oplus \boldsymbol{x}_{a,\ell\to 1}^{(0)}$ the two channel pmfs for the *same* observation $\boldsymbol{y}_c$ can be written as

$$P(\boldsymbol{y}_c|\boldsymbol{x}_a^{(0)}, \boldsymbol{z}_c) = \prod_{\lambda=1}^{|\mathbb{Z}|} (1 - \rho(z_{c,\lambda}))^{N_\lambda - w(\boldsymbol{e}_{0,\lambda})}\ \rho(z_{c,\lambda})^{w(\boldsymbol{e}_{0,\lambda})}, \quad (20)$$

$$P(\boldsymbol{y}_c|\boldsymbol{x}_{a,\ell\to 1}^{(0)}, \boldsymbol{z}_c) = \prod_{\lambda=1}^{|\mathbb{Z}|} (1 - \rho(z_{c,\lambda}))^{N_\lambda - w(\boldsymbol{d}\oplus\boldsymbol{e}_{0,\lambda})} \cdot \rho(z_{c,\lambda})^{w(\boldsymbol{d}\oplus\boldsymbol{e}_{0,\lambda})}, \quad (21)$$

where $w(\boldsymbol{x})$ denotes the Hamming weight of sequence $\boldsymbol{x}$. By inserting (20) and (21) into (18) we obtain

$$P(X_{a,\ell}^{(0)} = 0|\boldsymbol{x}_{a,\backslash\ell}^{(0)}, \boldsymbol{y}_c) = \frac{1}{1 + \prod_{\lambda=1}^{|\mathbb{Z}|} \left(\dfrac{\rho(z_{c,\lambda})}{1 - \rho(z_{c,\lambda})}\right)^{w(\boldsymbol{d}\oplus\boldsymbol{e}_{0,\lambda}) - w(\boldsymbol{e}_{0,\lambda})}}. \quad (22)$$

We now show that $\lim_{N_\lambda\to\infty}(w(\boldsymbol{d} \oplus \boldsymbol{e}_{0,\lambda}) - w(\boldsymbol{e}_{0,\lambda}))$ is unbounded for infinite output weight $w(\boldsymbol{d})$. The Strong Law of Large Numbers implies that $P(\lim_{N_\lambda\to\infty} w(\boldsymbol{e}_{0,\lambda})/N_\lambda = \rho(z_{c,\lambda})) = 1$ for any $\lambda \in \{1, 2, \ldots |\mathbb{Z}|\}$. Thus, the term $w(\boldsymbol{d} \oplus \boldsymbol{e}_{0,\lambda})$ can asymptotically be written as

$$\lim_{N_\lambda\to\infty} w(\boldsymbol{d} \oplus \boldsymbol{e}_{0,\lambda}) = w(\boldsymbol{d}) - w(\boldsymbol{d})\rho(z_{c,\lambda}) + (N_\lambda - w(\boldsymbol{d}))\rho(z_{c,\lambda}), \quad (23)$$
$$= (1 - 2\rho(z_{c,\lambda}))w(\boldsymbol{d}) + N_\lambda \rho(z_{c,\lambda}), \quad (24)$$

where in (23) the second term on the right hand side indicates the number of bits that are flipped from 1 to 0, and the third term indicates the number of bits which are flipped from 0 to 1, respectively. From (24) we can now see that

$$\lim_{N_\lambda\to\infty} (w(\boldsymbol{d} \oplus \boldsymbol{e}_{0,\lambda}) - w(\boldsymbol{e}_{0,\lambda})) = (1 - 2\rho(z_{c,\lambda}))w(\boldsymbol{d}), \quad (25)$$

which, for $\rho(z_{c,\lambda}) \in (0, 1/2)$, tends to infinity since our assumption $\lim_{N\to\infty} w(\boldsymbol{d}) \to \infty$ implies that $\lim_{N_\lambda\to\infty} w(\boldsymbol{d}) \to \infty$. By combining (22) and (25) and assuming that $0 < C_{\text{BISDMC}} \le 1$, which implies $\rho(z_{c,\lambda})/(1 - \rho(z_{c,\lambda})) < 1$ for all λ, we finally obtain

$$\lim_{N\to\infty} P(X_{a,\ell}^{(0)} = 0|\boldsymbol{x}_{a,\backslash\ell}^{(0)}, \boldsymbol{y}_c) = \begin{cases} 1 & \text{for} \quad 0 < C_{\text{BISDMC}} \le 1, \\ \frac{1}{2} & \text{for} \quad C_{\text{BISDMC}} = 0. \end{cases} \quad (26)$$

This proves the sufficient part of Lemma 1. The necessary part can easily be shown by forcing the quotient in the denominator of (18) to tend to zero for $K \to \infty$, $N \to \infty$, and fixed R_c.

For BISMCs with continuous outputs the proof can be carried out in a similar way by considering the limiting case $|\mathbb{Z}| \to \infty$. ■

Remarks:

- The above proof can be extended to inner factor nodes in factor graphs. The proof is related to the above proof for the *a priori* channel model, and a sketch for BISDMCs is presented in the Appendix.
- Lemma 1 says that for any $C > 0$ the missing bit $x_{a,\ell}^{(0)}$ can be perfectly estimated asymptotically by an ML decoder if all other bits in the word are known at the decoder. Intuitively, this works since the effective code rate R_{eff} for the single missing bit x_ℓ is $1/N$, and therefore there exists an N' such that $R_{\text{eff}} < C$ can always be guaranteed for any blocklength $N > N'$.

We can now state the main result of this section in the following theorem for the case where $I_{A,\max} = I_{E,\max} = 1\,\text{bit}$.

Theorem 2: For an inner encoder in a serial concatenation or a constituent encoder in a parallel concatenation of two or more codes, a recursive convolutional encoder $\mathcal{C}$ is required to achieve $I_E(I_A = 1\,\text{bit}) = 1\,\text{bit}$ for an information blocklength tending to infinity and transmission over a BPSK-modulated BISMC with non-zero capacity.

Proof: First note that for recursive convolutional encoders the generator matrix $\mathbf{G}(D)$ in the transform domain has entries which are rational functions in D. Thus, since for weight-one input sequences the rational polynomial $P(D) = 1/(\sum_{k=1}^{m} a_k D^k)$, $a_k \in \{0, 1\}$, with finite degree m in the denominator polynomial has an unbounded degree, the corresponding (time-domain) output sequence $\boldsymbol{p}$ has infinite weight.

Let $\boldsymbol{c}_0 \triangleq 1 - 2\mathscr{C}(\boldsymbol{x}_{a,\ell\to 0})$ and $\boldsymbol{c}_1 \triangleq 1 - 2\mathscr{C}(\boldsymbol{x}_{a,\ell\to 1})$, respectively. Assume that we transmit the BPSK codeword $\boldsymbol{c}_0$ and the sequence $\boldsymbol{y}_c$ is observed at the channel output. Due to the linearity of the encoder and the fact that a weight-one information word $\boldsymbol{x}_{a,\ell\to 1}^{(0)}$ asymptotically results in an infinite codeword weight at the output of a recursive convolutional encoder, we can directly apply Lemma 1. With $\boldsymbol{c}_0$ corresponding to $\boldsymbol{c}_0^{(0)}$ and $\boldsymbol{c}_1$ corresponding to $\boldsymbol{c}_1^{(0)}$, resp., we conclude that

$$\begin{aligned} \lim_{N\to\infty} P(X_{a,\ell} = 0 | \boldsymbol{x}_{a,\setminus\ell}, \boldsymbol{y}_c) &= 1, \\ \lim_{N\to\infty} P(X_{a,\ell} = 1 | \boldsymbol{x}_{a,\setminus\ell}, \boldsymbol{y}_c) &= 0 \end{aligned} \tag{27}$$

for all information word realizations $\boldsymbol{b}$ and all channel observations $\boldsymbol{y}_c$ associated with a non-zero capacity. Inserting (27) into (15) and considering that $\lim_{x\to 0} x \log x = 0$ proves the theorem. ■

It is clear from the proof of Theorem 2 that the condition $I_{E_i}(I_{A_i} = 1\,\text{bit}) = 1\,\text{bit}$ cannot be satisfied by codes for which an information word of weight one generates a finite codeword weight. This holds for all block codes and non-recursive convolutional encoders. Hence this extends the well known result that, with ML decoding, non-recursive convolutional encoders should not be used as inner encoders in a serially concatenated coding scheme [15] or as constituent encoders in parallel concatenation [16] to the case of *iterative* decoding. Further, combining the results from Theorems 1 and 2 provides necessary and sufficient conditions under which iterative decoding of concatenated linear codes has a fixed-point if the *a priori* information is perfect.

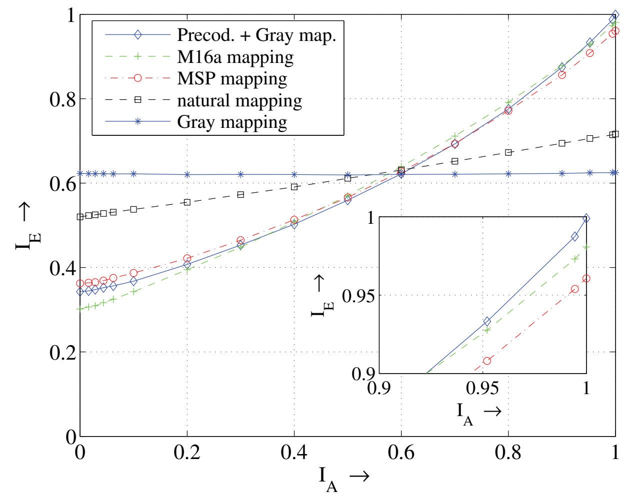

Fig. 4. Comparison of (inner) demapper EXIT functions for different 16-QAM mappings and $E_s/N_0 = 7.5\,\text{dB}$ on an AWGN channel. The M16a and MSP mappings were proposed in [41] and [36], respectively.

These results also lead to necessary and sufficient conditions for factor nodes in a factor graph framework employing the iterative SPA to achieve asymptotically perfect outgoing messages when perfect incoming messages are given. In the context of bipartite graphs for linear codes, where inner factor nodes are the variable nodes and non-inner (outer) factor nodes are the check nodes, this result suggests that, if the conditions stated in Theorem 1 and Lemma 1 are satisfied, density evolution has a fixed-point under perfect *a priori* information.

Example 2: The result from Theorem 2 is now applied to iteratively-decoded bit-interleaved coded modulation (BICM-ID) in a serial concatenation. The inner encoder simply maps a block of J bits to a complex waveform, and the corresponding constituent decoder is represented by a SISO demapper (see, e.g., [36], [37]). Since the inner encoder is non-recursive, perfect extrinsic information is never achieved for perfect *a priori* information. In order to be able to achieve an average extrinsic information of $I_E = 1\,\text{bit}$, we insert an accumulator with generator polynomial $g(D) = 1/(1 + D)$ as a recursive precoder before the mapping operation (see, e.g., [38]–[40]). SISO decoding can then be carried out on the joint trellis for the recursive precoder and the mapper. Fig. 4 shows some constituent decoder EXIT functions for selected 16-QAM mappings and the recursive precoding followed by Gray mapping for an AWGN channel with $E_s/N_0 = 7.5\,\text{dB}$. The M16a mapping was proposed in [41], where it was obtained by an optimization approach minimizing the difference $1 - I_E(I_A = 1\,\text{bit})$. In [36], the MSP mapping was suggested as a good trade-off to achieve large I_E for both zero and perfect *a priori* information. We observe from Fig. 4 that only Gray-mapped recursive precoding is able to achieve $I_E = 1\,\text{bit}$, whereas both the M16a and MSP mappings only get close to $I_E = 1\,\text{bit}$. Although for these mappings the deviation from ideal behavior is very small, this has a strong impact on performance in the error floor region. This can be seen from numerical results for a serially concatenated coding scheme where the outer encoder is a rate $R_o = 1/2$, memory-1, recursive systematic convolutional (RSC) encoder with generator matrix $\mathbf{G}(D) = [1 \;\; 1/(1+D)]$. The resulting bit error rate (BER) performance is shown in Fig. 5, where $\eta = R_o J = 2\,\text{bit/s/Hz}$ is the effective throughput and $E_b = E_s/\eta$. It is seen from Fig. 5 that, in contrast to both the MSP and M16a mappings, no error floor is observed for the Gray-mapped recursive precoding scheme, where all simulated transmissions were observed as error-free for $E_b/N_0 > 4\,\text{dB}$. ■

B. Achievable extrinsic information for finite blocklength

The proof of Theorem 2 only addresses the case of an infinite output codeword weight for an information word weight of one. In the following we address the question of how the achievable extrinsic information for perfect *a priori* information depends on a *finite* output codeword weight $w(\boldsymbol{d}_c)$, with $\boldsymbol{d}_c \triangleq \mathscr{C}(\boldsymbol{x}_a^{(0)}) - \mathscr{C}(\boldsymbol{x}_{a,\ell\to 1}^{(0)})$, where, without loss of generality, we consider the BIAWGN channel.

By considering the linearity of the code $\mathcal{C}$, (15) can be rewritten as

$$\begin{aligned} H(X_{a,\ell}|\,\boldsymbol{Y}_{a,\setminus\ell}, \boldsymbol{Y}_c)\Big|_{I_A = I_{A,\max}} &= -\int_{\boldsymbol{y}_c} p(\boldsymbol{y}_c|\boldsymbol{x}_{a,\setminus\ell}^{(0)}) \sum_{q=0}^{1} P(X_{a,\ell} = q|\boldsymbol{x}_{a,\setminus\ell}^{(0)}, \boldsymbol{y}_c)\cdot \\ &\qquad \log_2(P(X_{a,\ell} = q|\boldsymbol{x}_{a,\setminus\ell}^{(0)}, \boldsymbol{y}_c))\, d\boldsymbol{y}_c \\ &= -\,E_{p(\boldsymbol{y}_c)}\Bigg\{\sum_{q=0}^{1} P(X_{a,\ell} = q|\boldsymbol{x}_{a,\setminus\ell}^{(0)}, \boldsymbol{y}_c)\cdot \\ &\qquad \log_2(P(X_{a,\ell} = q|\boldsymbol{x}_{a,\setminus\ell}^{(0)}, \boldsymbol{y}_c))\Bigg\}, \end{aligned} \quad (28)$$

where (28) exploits the fact that the pdf $p(\boldsymbol{y}_c|\boldsymbol{x}_{a,\setminus\ell}^{(0)})$ is conditioned on the all-zero information word. For the BIAWGN channel the encoded all-zero information word $\mathscr{C}(\boldsymbol{x}_a^{(0)})$ is transmitted, where $\boldsymbol{y}_c = \boldsymbol{c}_0^{(0)} + \boldsymbol{n}$ is observed at the channel output, $\boldsymbol{n} = [n_1, \ldots, n_\ell, \ldots, n_N]$, and n_ℓ, $\ell = 1, 2, \ldots, N$, represents a zero-mean Gaussian noise sample with variance σ_n^2. Additionally, observing the same channel output sequence $\boldsymbol{y}_c$ with the BPSK-modulated codeword $\boldsymbol{c}_1^{(0)} = 1 - 2\,\mathscr{C}(\boldsymbol{x}_{a,\ell\to 1}^{(0)})$ leads to the error sequences

$$\boldsymbol{e}_0 = \boldsymbol{n}, \quad \boldsymbol{e}_1 = \boldsymbol{y}_c - \boldsymbol{c}_1^{(0)} = \boldsymbol{c}_0^{(0)} - \boldsymbol{c}_1^{(0)} + \boldsymbol{n} = \boldsymbol{d} + \boldsymbol{n}, \quad (29)$$

where $\boldsymbol{d} \triangleq \boldsymbol{c}_0^{(0)} - \boldsymbol{c}_1^{(0)} = -2\boldsymbol{d}_c$ and $w(\boldsymbol{d}) = w(\boldsymbol{d}_c)$. By employing (18) with all pmfs replaced by Gaussian pdfs we obtain

$$P(X_{a,\ell}^{(0)} = 0|\boldsymbol{x}_{a,\setminus\ell}^{(0)}, \boldsymbol{y}_c) = \frac{1}{1 + \exp\left(\dfrac{|\boldsymbol{e}_0|^2 - |\boldsymbol{e}_1|^2}{2\,\sigma_n^2}\right)}. \quad (30)$$

Inserting this into (28) and considering (7) and (8), we obtain the desired relation between $I_E(I_A = I_{A,\max})$ and the corresponding average output weight $w(\boldsymbol{d}_c)$ for a given channel noise variance σ_n^2. In Fig. 6 this relation is plotted for different $E_s/N_0 = 1/(2\,\sigma_n^2)$ and $I_{A,\max} = 1\,\text{bit}$, where the expectation in (28) is taken over 1000 channel realizations. We can see from Fig. 6 that, for example, for $E_s/N_0 \geq -1\,\text{dB}$ an average codeword weight of $w(\boldsymbol{d}_c) = 6$ suffices to achieve $I_E \approx 1\,\text{bit}$. Thus, in practice, for moderate E_s/N_0 and block or non-recursive convolutional codes with a modest value of $w(\boldsymbol{d}_c)$, the error $|I_E - 1|$ is very small, so that these codes may be used as inner codes in serial concatenation or constituent codes in parallel concatenation with only a negligible loss in performance compared to recursive convolutional codes.

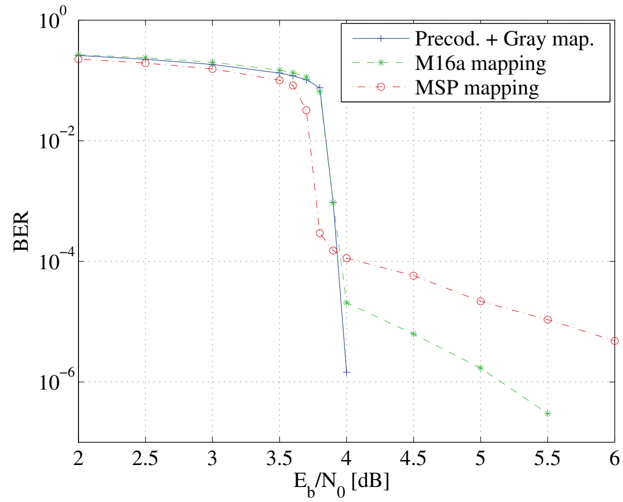

Fig. 5. BER versus E_b/N_0 for the proposed inner Gray-mapped recursive precoding scheme and the M16a [41] and MSP [36] mappings for the 16-QAM constellation and an outer memory-1 RSC encoder (1000 simulated transmissions, blocklength 20000 information bits).

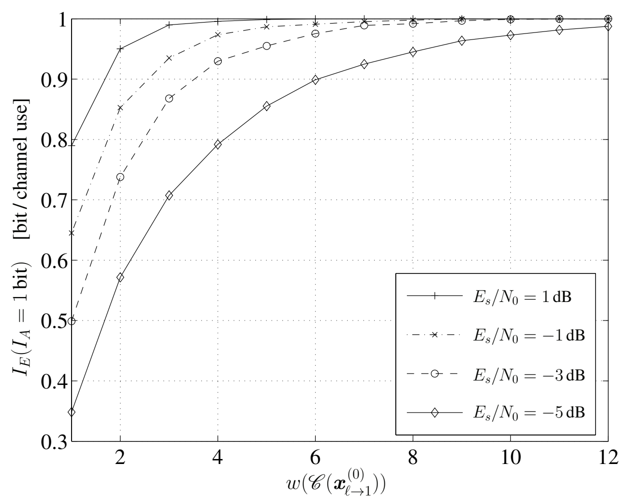

Fig. 6. Relation between $I_E(I_A = 1\,\text{bit})$ and the corresponding average codeword weight $w(\boldsymbol{d}_c)$ for a weight-one information word $\boldsymbol{x}_{a,\ell\to 1}^{(0)}$ as a function of E_s/N_0 for a BIAWGN channel.

V. Conclusion

We have proved necessary and sufficient conditions for the constituent codes of concatenated coding schemes over BISMCs that allow them to asymptotically achieve the maximum possible average extrinsic information at the output of the corresponding SISO decoder if perfect *a priori* information is available. By extending these results to a factor graph framework and the SPA the same conditions can be established for factor nodes, for example for check and variable nodes of doubly-generalized LDPC codes.

Specifically, we have addressed two cases: For non-inner codes in a (multiple) serial concatenation and non-inner factor nodes, a maximum average extrinsic information requires a minimum Hamming distance of two or greater between adjacent codewords. For linear inner codes in a (multiple) serial concatenation, linear constituent codes in a (multiple) parallel concatenation, or inner factor nodes associated with a linear code, perfect extrinsic information is obtained if and only if the encoder produces an infinite output weight for a weight-one input sequence (e.g., a recursive convolutional encoder). Combining these results provides necessary and sufficient conditions under which both iterative decoding of concatenated linear codes and density evolution on a factor graph have a fixed-point for perfect *a priori* information.

Examples for both cases have shown that the results can be used to improve overall system performance. In the non-inner code case the first example considers joint source-channel decoding schemes with an outer source mapping that produces codewords with a minimum Hamming distance of one (e.g., a Markov source). An additional high-rate block encoding after the source encoder results in $d_{\min} \geq 2$, so that the joint SISO decoder is able to achieve perfect extrinsic information. The second example for the inner code case considers BICM-ID in which the SISO demapper does not achieve perfect extrinsic information even when optimized mappings are used, leading to a noticeable error floor in the BER performance curve. By adding an accumulator as a recursive precoder combined with joint SISO mapping/decoding, perfect achievable extrinsic information is guaranteed for the inner decoder.

APPENDIX

PROOF OF LEMMA 1 FOR A FACTOR-GRAPH FRAMEWORK

For the sake of simplicity we only consider discrete-output BISMCs. The extension to arbitrary BISMCs is straightforward. Consider a factor node of degree $M = K$ and associated independent variables $x_{a,1}, \ldots, x_{a,M}$ with the function definition

$$f(x_{a,1}, x_{a,2}, \ldots, x_{a,M}) = \prod_{t=1}^{N} P(y_{c,t} | \mathscr{C}_t(\boldsymbol{x}_a)), \tag{31}$$

where $x_{c,t} = \mathscr{C}_t(\boldsymbol{x}_a)$ is the t-th component of the codeword $\boldsymbol{x}_c$ in the linear code $\mathcal{C}$. Lemma 1 states that if the all zero codeword $\boldsymbol{x}_c^{(0)}$ is transmitted, then

$$\mu_\ell^{(\text{in})}(q) = 1 - q \quad \text{for} \quad q \in \{0, 1\} \quad \text{and all} \quad \ell = 1, \ldots, M, \tag{32}$$

which is equivalent to $I_A = I_{A,\max}$. Then, with probability 1 as $N \to \infty$, the normalized outgoing messages are

$$\mu_m^{(\text{out})}(q) = 1 - q \quad \text{for} \quad q \in \{0, 1\} \quad \text{and all} \quad m = 1, \ldots, M, \tag{33}$$

which implies $H(X_{a,m} | \boldsymbol{Y}_{a,\setminus m}, \boldsymbol{Y}_c) = 0$.

By using the incoming messages in (32), the SPA update rule in (6) simplifies significantly and we obtain the following outgoing messages:

$$\mu_m^{(\text{out})}(0) = \prod_{t=1}^{N} P(y_{c,t} | \mathscr{C}_t(0, 0, \ldots, 0, \ldots, 0, 0)),$$
$$\mu_m^{(\text{out})}(1) = \prod_{t=1}^{N} P(y_{c,t} | \mathscr{C}_t(0, 0, \ldots, 1, \ldots, 0, 0)).$$

We now have

$$\begin{aligned} \ln\left(\frac{\mu_m^{(\text{out})}(0)}{\mu_m^{(\text{out})}(1)}\right) &= \sum_{t=1}^{N} \ln\left(\frac{P(y_{c,t} | \mathscr{C}_t(0, 0, \ldots, 0, \ldots, 0, 0))}{P(y_{c,t} | \mathscr{C}_t(0, 0, \ldots, 1, \ldots, 0, 0))}\right) \\ &= \sum_{t: \mathscr{C}_t(0,0,\ldots,1,\ldots,0,0)=1} \ln\left(\frac{P(y_{c,t}|0)}{P(y_{c,t}|1)}\right). \end{aligned} \tag{34}$$

Next we assume that the communication channel has a bit error probability of less than 0.5, i.e.,

$$\sum_{y_{c,t}} P(y_{c,t}|0) \ln\left(\frac{P(y_{c,t}|0)}{P(y_{c,t}|1)}\right) > 0, \quad t = 1, \ldots, N, \tag{35}$$

and that

$$\rho_m = \lim_{N \to \infty} \frac{1}{N} w(\mathscr{C}(0, 0, \ldots, 1, \ldots, 0, 0)) > 0, \qquad m = 1, \ldots, M. \tag{36}$$

This condition means that we assume the normalized output weight ρ_m of $\mathcal{C}$ grows linearly in N for an input weight of one, since $w(\mathscr{C}(0, 0, \ldots, 1, \ldots, 0, 0)) \to \infty$ for $N \to \infty$.

By applying the Strong Law of Large Numbers to the r.h.s. of (34), asymptotically we obtain with probability 1 for all $m = 1, \ldots, M$, $t = 1, \ldots, N$ that

$$\begin{aligned} \lim_{N \to \infty} \frac{1}{N} \ln\left(\frac{\mu_m^{(\text{out})}(0)}{\mu_m^{(\text{out})}(1)}\right) &= \rho_m\, E\left\{\ln\left(\frac{P(Y_{c,t}|0)}{P(Y_{c,t}|1)}\right)\right\} \\ &= \rho_m \sum_{y_{c,t}} P(y_{c,t}|0) \cdot \\ &\qquad \ln\left(\frac{P(y_{c,t}|0)}{P(y_{c,t}|1)}\right) > 0, \end{aligned} \tag{37}$$

where $Y_{c,t}$ is distributed according to $P(y_{c,t}|0)$, and the inequality on the r.h.s. of (37) follows from the assumptions in (35) and (36). From (37) finally the result in (33) follows.

ACKNOWLEDGMENT

The authors wish to thank Pascal Vontobel for contributing the idea of applying the results for the *a priori* model also to factor graphs.

REFERENCES

[1] J. Kliewer, A. Huebner, and D. J. Costello, Jr., “On the achievable extrinsic information of inner decoders in serial concatenation,” in *Proc. 2006 IEEE Int. Sympos. on Information Theory*, pp. 2680–2684.

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

[3] A. Ashikhmin, G. Kramer, and S. ten Brink, “Extrinsic information transfer functions: model and erasure channel properties,” *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2657–2673, Nov. 2004.

[4] S. Huettinger and J. Huber, “Information processing and combining in channel coding,” in *Proc. 2003 Int. Symp. on Turbo Codes & Rel. Topics*, pp. 95–102.

[5] R. G. Gallager, *Low-Density Parity-Check Codes*. MIT Press, 1963.

[6] T. J. Richardson and R. L. Urbanke, “The capacity of low-density parity-check codes under message-passing decoding,” *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.

[7] C. Méasson, A. Montanari, T. Richardson, and R. Urbanke, “Life above threshold: from list decoding to area theorem and MSE,” in *Proc. 2004 IEEE Information Theory Workshop*.

[8] J. Kliewer, N. Goertz, and A. Mertins, “Iterative source-channel decoding with Markov random field source models,” *IEEE Trans. Signal Process.*, vol. 34, no. 10, pp. 3688–3701, Oct. 2006.

[9] K. Sayood and J. C. Borkenhagen, “Use of residual redundancy in the design of joint source/channel coders,” *IEEE Trans. Commun.*, vol. 39, no. 6, pp. 838–846, June 1991.

[10] H. H. Otu and K. Sayood, “A joint source/channel coder with block constraints,” *IEEE Trans. Commun.*, vol. 47, no. 22, pp. 1615–1618, Nov. 1999.

[11] N. Görtz, “On the iterative approximation of optimal joint source-channel decoding,” *IEEE J. Sel. Areas Commun.*, vol. 14, no. 9, pp. 1662–1670, Sept. 2001.

[12] J. Kliewer and R. Thobaben, “Iterative joint source-channel decoding of variable-length codes using residual source redundancy,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 919–929, May 2005.

[13] R. Thobaben and J. Kliewer, “Low complexity iterative joint source-channel decoding for variable-length encoded Markov sources,” *IEEE Trans. Commun.*, vol. 53, no. 12, pp. 2054–2064, Dec. 2005.

[14] E. Paolini, M. P. C. Fossorier, and M. Chiani, “Generalized and doubly generalized LDPC codes with random component codes for the binary erasure channel,” *IEEE Trans. Inf. Theory*, vol. 54, no. 4, pp. 1651–1672, Apr. 2010.

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

[16] S. Benedetto and G. Montorsi, “Unveiling turbo codes: some results on parallel concatenated coding schemes,” *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 409–428, Mar. 1996.

[17] D. Divsalar, S. Dolinar, and F. Pollara, “Iterative turbo-decoder analysis based on density evolution,” *IEEE J. Sel. Areas Commun.*, vol. 19, no. 5, pp. 891–907, May 2001.

[18] M. Tuechler, “Design of serially concatenated codes depending on the block length,” *IEEE Trans. Commun.*, vol. 52, no. 2, pp. 209–218, Feb. 2004.

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

[20] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, “Factor graphs and the sum-product algorithm,” *IEEE Trans. Inf. Theory*, vol. 2, no. 47, pp. 498–519, Feb. 2001.

[21] N. Wiberg, “Codes and decoding on general graphs,” Ph.D. thesis, Linköping University, Sweden, 1996.

[22] G. D. Forney, Jr., “Codes on graphs: normal realizations,” *IEEE Trans. Inf. Theory*, vol. 2, no. 47, pp. 520–548, Feb. 2001.

[23] S. M. Aji and R. J. McEliece, “The generalized distributive law,” *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 325–343, Feb. 2000.

[24] H.-A. Loeliger, “An introduction to factor graphs,” *IEEE Signal Process. Mag.*, pp. 28–41, Jan. 2004.

[25] A. Guyader, E. Fabre, C. Guillemot, and M. Robert, “Joint source-channel turbo decoding of entropy-coded sources,” *IEEE J. Sel. Areas Commun.*, vol. 19, no. 9, pp. 1680–1696, Sept. 2001.

[26] I. Land, P. Hoeher, and S. Gligorević, “Computation of symbol-wise mutual information in transmission systems with logAPP decoders and application to EXIT charts,” in *Proc. 2004 International ITG Conference on Source and Channel Coding*, pp. 195–202.

[27] R. Bauer and J. Hagenauer, “On variable length codes for iterative source/channel decoding,” in *Proc. 2001 IEEE Data Compression Conference*, pp. 273–282.

[28] Y. Takishima, M. Wada, and H. Murakami, “Reversible variable length codes,” *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 158–162, Feb./Mar./Apr. 1995.

[29] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, “Optimal decoding of linear codes for minimizing symbol error rate,” *IEEE Trans. Inf. Theory*, pp. 294–287, Mar. 1974.

[30] F. I. Alajaji, N. C. Phamdo, and T. E. Fuja, “Channel codes that exploit the residual redundancy in CELP-encoded speech,” *IEEE Trans. Speech and Audio Process.*, vol. 4, no. 5, pp. 325–336, Sept. 1996.

[31] T. Fazel and T. Fuja, “Robust transmission of MELP-compressed speech: an illustrative example of joint source-channel decoding,” *IEEE Trans. Commun.*, vol. 51, no. 6, pp. 973–982, June 2003.

[32] T. Fingscheidt and P. Vary, “Softbit speech decoding: a new approach to error concealment,” *IEEE Trans. Speech and Audio Process.*, vol. 9, no. 3, pp. 240–251, Mar. 2001.

[33] N. Chayat and S. Shamai, “Extension of an entropy property for binary input memoryless symmetric channels,” *IEEE Trans. Inf. Theory*, vol. 35, no. 5, pp. 1077–1079, Sept. 1989.

[34] I. Land, S. Huettinger, P. A. Hoeher, and J. B. Huber, “Bounds on information combining,” *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 612–619, Feb. 2005.

[35] I. Land and J. Huber, *Information Combining*, Foundations and Trends in Communications and Information Theory. NOW Publishers, 2006.

[36] A. Chindapol and J. A. Ritchey, “Design, analysis, and performance evaluation for BICM-ID with square QAM constellations in Rayleigh fading channels,” *IEEE J. Sel. Areas Commun.*, vol. 19, no. 5, pp. 944–957, May 2001.

[37] X. Li, A. Chindapol, and J. A. Ritcey, “Bit-interleaved coded modulation with iterative decoding and 8PSK signaling,” *IEEE Trans. Commun.*, vol. 50, no. 8, pp. 1250–1257, Aug. 2002.

[38] N. Gresset, J. J. Boutros, and L. Brunel, “Optimal linear precoding for BICM over MIMO channels,” in *Proc. 2004 IEEE Int. Sympos. on Information Theory*, p. 66.

[39] F. Simoens, H. Wymeersch, and M. Moeneclaey, “Linear precoders for bit-interleaved coded modulation on AWGN channels: analysis and design criteria,” *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 87–99, Jan. 2008.

[40] D. Duyck, J. J. Boutros, and M. Moeneclaey, “Precoding for word error rate minimization of LDPC coded modulation on block fading channels,” *IEEE Trans. Wireless Commun.*, vol. 11, no. 7, pp. 2457–2467, July 2012.

[41] F. Schreckenbach, N. Goertz, J. Hagenauer, and G. Bauch, “Optimization of symbol mappings for bit-interleaved coded modulation with iterative decoding,” *IEEE Commun. Lett.*, vol. 7, no. 12, pp. 593–595, Dec. 2003.

Jörg Kliewer (S’97-M’99-SM’04) received the Dipl.-Ing. (M.Sc.) degree in electrical engineering from Hamburg University of Technology, Hamburg, Germany, in 1993 and the Dr.-Ing. degree (Ph.D.) in electrical engineering from the University of Kiel, Germany, in 1999, respectively. From 1993 to 1998, he was a Research Assistant at the University of Kiel, and from 1999 to 2004, he was a Senior Researcher and Lecturer with the same institution. In 2004, he visited the University of Southampton, U.K., for one year, and from 2005 until 2007, he was with the University of Notre Dame, IN, as a Visiting Assistant Professor. In August 2007, he joined New Mexico State University, Las Cruces, NM, as an Assistant Professor. His research interests include error-correcting codes, network coding, wireless communications, and communication networks. Dr. Kliewer was the recipient of a Leverhulme Trust Award and a German Research Foundation Fellowship Award in 2003 and 2004, respectively. He was a Member of the Editorial Board of the *EURASIP Journal on Advances in Signal Processing* from 2005-2009 and is Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS since 2008.

Daniel J. Costello, Jr. (S’62-M’69-SM’78-F’86-LF’08) received his Ph.D. in Electrical Engineering from the University of Notre Dame in 1969. Since 1985, he has been a Professor of Electrical Engineering at Notre Dame and from 1989 to 1998 served as Chair of the Department. In 2000, he was named the Leonard Bettex Professor of Electrical Engineering. Dr. Costello has been a member of IEEE since 1969 and was elected Fellow in 1985. In 2000, the IEEE Information Theory Society selected him as a recipient of a Third-Millennium Medal, and he was a co-recipient of the 2009 IEEE Donald G. Fink Prize Paper Award and the 2012 ComSoc & Information Theory Society Joint Paper Award. Dr. Costello’s research interests are in the area of digital communications, with emphasis on error control coding and coded modulation. He has numerous technical publications in his field, and in 1983 he co-authored a textbook entitled *Error Control Coding: Fundamentals and Applications*, the 2nd edition of which was published in 2004.