On some Properties of the Mutual Information between Extrinsics with Application to Iterative Decoding

Florence Alberge

On some Properties of the Mutual Information between Extrinsics with Application to Iterative Decoding

Florence Alberge
LSS (UMR8506), Univ. Paris-Sud, CNRS, CentraleSupelec,
3 rue Joliot-Curie, 91192 Gif-sur-Yvette cedex, France
Tel: +33 1 69851757; fax: +33 1 69851765 ; e-mail:alberge@lss.supelec.fr

***Abstract*—Iterative decoding is an efficient error-correction tool based on the exchange of extrinsic probabilities between the constituent decoders. In this paper, the properties of the mutual information between the extrinsic LLR at the output of two constituent decoders are analyzed with application to turbo and LDPC codes. This is a bridge between information-theoretic analysis and practical implementations. It is proved here that the mutual information between extrinsics is a lower bound of the mutual information between each extrinsic and the transmitted message. In addition, an efficient online evaluation is provided in the paper with accuracy validated through numerical experiments. As an application, the mutual information between extrinsics is used for designing efficient stopping criterion and error detection rules at the decoder side. Two online methods for the estimation of optimal scaling factor to be applied to the extrinsic LLR are also derived. In contrast with most references, an analytical expression is obtained that does not require estimation of the actual transmitted bits. All results in the paper are derived for Gaussian distributed LLR with independent mean and variance.**

***Index Terms*—Mutual Information, Iterative Decoding, Stopping Rule, EXIT Charts.**

I. Introduction

Extrinsic information transfer (EXIT) charts were first introduced in [1] for the analysis of iterative decoding. The principle is quite general and has been applied to turbo-codes [2], LDPC codes [3] or repeat-accumulate codes [4] and extended to non-binary iterative decoding in [5]. In EXIT analysis, the mutual information $I(L, X)$ between the extrinsic log-likelihood ratios (LLR) L and the binary message X is computed for performance evaluation. This is an offline tool which allows to compute a performance rating on the reliability of a solution and is usually intended to design capacity-approaching codes before implementing them. In an actual transmission, the receiver decodes the transmitted message via an iterative turbo decoding process based on extrinsic propagation [6]. In this iterative receiver, the extrinsic information is updated within each individual decoding block and passed to the other decoder over many iterations. This is a pragmatic approach since iterative decoding was not originally introduced as the solution to an optimization problem rendering its analysis difficult. As a consequence, the stopping criterion of the iterative process can not easily be derived from the evaluation of an objective function which would converge to some limit value. In addition, the mutual information $I(L, X)$ which is the performance indicator in the Exit Charts is not easily computable at the receiver. Thus, different classes of approximations and rules have been developed for error detection and stopping criteria. One important class of stopping criteria is concerned with the detection of a stationary point. This class encompasses cross entropy [7] stopping rule, sign change rule [8], [9] or hard-decision-aided [10] rule as special cases. These rules have difficulties stopping the decoding if the iterative process enters an oscillatory behavior [11] and do not have any error detection capability. Another class is based on a performance evaluation like in [12] where the stopping rule is based on the mean of the absolute values of the log-likelihood ratios at the output of the component decoders or on the instantaneous values as in [13]. It is well-known that the reliability of a solution is connected with the magnitude of the LLR however the choice of the threshold is not trivial.

Independently, several publications such as [14], [15], [16] considered the issue of optimal LLR scaling. Indeed, it was frequently observed in turbo-codes and LDPC codes that the imperfections in the receiver can substantially lower the performance of the decoder. Linear LLR correction can potentially compensate the degradation. The LLR-distribution is analyzed for both LDPC and turbo-decoder in [17], [18]. It is pointed out that the LLR are well approximated by Gaussian distributions but it is inappropriate to use the mean of the density only to model the iterative decoding process. Using this result, the LLR are modelled in this paper using Gaussian distribution including a scaling factor which is equivalent to considering independent mean and variance.

This paper shows that the two issues (stopping rule and optimal scaling) can be addressed jointly by introducing a new performance indicator. This indicator is the mutual information between extrinsics $I(L_y, L_z)$ where L_y and L_z denote the extrinsic LLRs at the output of two individual component decoders. The contributions of the paper are listed below. The mutual information between extrinsics is proved to be a lower bound of the mutual information between the extrinsics LLR and the message X to be retrieved. As a consequence, $I(L_y, L_z)$ is an indicator for a successful/unsuccessful decoding and a good metric for stopping criterion. The mutual information between extrinsics can be measured offline via an histogram method as in Exit Charts. A single curve Exit

Chart [19] is used here since this configuration is suitable for tracking the evolution of a given system as a whole and enables an easy evaluation of $I(L_y, L_z)$. In parallel, an efficient online computation of $I(L_y, L_z)$ is derived. Comparison with the results of the Exit Charts shows a good accuracy of the online estimation. In addition to the previous results, two methods are proposed for the online estimation of the optimal scaling factor. These methods allow closed-form expression and do not require an estimation of the actual message bits as in [16], [20]. It should be mentioned however that the latter references do not require Gaussian assumption and are more general than the methods proposed here. It will be shown, through simulations that, in the context of Gaussian LLR, the methods proposed here outperform [16] in particular in early iterations and at low SNR. The connection between the optimal scaling factor and the mutual information between extrinsics is also emphasized. It is well understood that the optimal scaling factor depends on the SNR, the iteration number and the code structure. The evolution of the mutual information between extrinsics shares the same characteristics. We will see in this paper that the optimal scaling factor can be viewed as a function of a single parameter: the mutual information $I(L_y, L_z)$ between extrinsics. The main contribution of the paper is thus to provide a theoretical analysis and a practical method for the full exploitation of the information contained in the extrinsics for a proper design of the receiver enabling self-diagnosis (successful decoding, early stop) and well-informed choices (scaling factor). The proposed methodology is quite general and rely partially on the assumption that the LLR are Gaussian distributed and that the symmetry property holds.

II. System model and notation

In this paper, random variables are denoted with upper-case letters and their corresponding realizations with lower-case letters. Sequences of random variables or vectors are indicated by boldface letters. Fig. 1 depicts the decoding model used in this paper and previously in [21]. The a priori channel models the a priori information used at the constituent decoders and the communication channel models the transmission medium between the transmitter and the receiver. Both serial and parallel turbo coding schemes can be seen as specific instantiations of this generic scheme. For the iterative decoding of an outer decoder in a serial concatenation, the switch in Fig. 1 is in position 1 and the communication channel is inactive. When the switch is in position 2, the system can be used for modeling the iterative decoding of a constituent decoder in a parallel concatenation setting. This system can also be used for modeling the input/output relationship at the Variable Node or at the Check Node decoder in LDPC. In both systems, the iterative decoding aims at finding the binary sequence $\mathbf{X} = (X_1, X_2, ..., X_K)$ with length K. A constituent decoder takes as input an a priori information $\mathbf{L_y}$ (resp. $\mathbf{L_z}$) which denotes log-likelihood ratios (LLR) of an extrinsic probability. The length of $\mathbf{L_y}$ (resp. $\mathbf{L_z}$) is K and $L_{y,k}$ is the k^{th} element of $\mathbf{L_y}$. At iteration i, the first decoder receives the prior information $\ell_y^{(i)}$ and outputs the regenerated prior information $\ell_z^{(i)}$ which comes as an input to the second

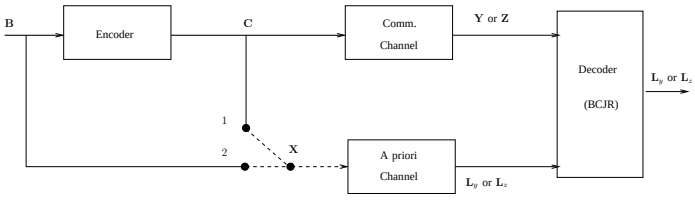

Figure 1. General Encoding/Decoding model

decoder and gives $\ell_y^{(i+1)}$ as an output. We then focus on the statistical properties of the extrinsic LLR. In other words, $L_{y,k}$ (or $L_{z,k}$) is considered as an outcome of random variable L_y (or L_z). In the following, index k will be omitted (except when needed for clarification) and we will use simply X, L_y, L_z or even L when an equation or a property hold for both L_y and L_z. Unless stated otherwise, the next properties hold [22]: **Symmetry**: $p_L(\ell|X) = p_L(-\ell| - X)$, **Generalized Consistency**: $\frac{p_L(\ell|X=1)}{p_L(-\ell|X=1)} = e^{\alpha\ell}$ with $\alpha \in \mathbb{R}^+$, **Range**: $\ell \in]-\infty; +\infty[$. In the seminal works on EXIT Charts and density evolution [2], [19], [23], the assumptions considered are symmetry and consistency which corresponds to $\alpha = 1$ in the Generalized Consistency property above. This comes from the assumption that L is a noisy version of $\mathbf{X}$ with expression:

$$L = \frac{\sigma^2}{2}X + \sigma n \tag{1}$$

with $n \sim \mathcal{N}(0, 1)$. In that case, $\frac{p_L(\ell|X=1)}{p_L(-\ell|X=1)} = e^{\ell}$. A stochastic analysis of iterative decoding is available in [17] where it is shown that the input-output signals in a turbo-decoder, when expressed using LLR, are indeed well approximated by a Gaussian distribution but with independent mean and variance. These results also hold for LDPC decoders [18]. In this paper, we will consider that

$$L = \alpha\frac{\sigma^2}{2}X + \sigma n \tag{2}$$

with $n \sim \mathcal{N}(0, 1)$. In that case, $\frac{p_L(\ell|X=1)}{p_L(-\ell|X=1)} = e^{\alpha\ell}$ and the Generalized consistency condition holds. This model is in accordance with the true LLR distribution that can be observed in iterative decoding of turbo or LDPC codes.

III. Mutual Information

The mutual information between extrinsics is analyzed first. The statistical model in (2) is considered and an equivalence class between LLR following model (1) and (2) is exhibited with mutual information as equivalence relation. Exit Charts are revisited with model (2).

A. Definition

The mutual information between L_y and L_z is defined as

$$I(L_y, L_z) = \frac{1}{log(2)} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p_{L_y,L_z}(\ell_y, \ell_z) \times \log\left(\frac{p_{L_y,L_z}(\ell_y,\ell_z)}{p_{L_y}(\ell_y)p_{L_z}(\ell_z)}\right) d\ell_y d\ell_z \tag{3}$$

Based on the three properties above (Symmetry, Generalized Consistency and Range), $I(L_y, L_z)$ can be written as a function of the mutual information between X and the LLR as:

$$I(L_y, L_z) = I(L_y, X) + I(L_z, X) - I(L_y + L_z, X) \quad (4)$$

where $I(L, X) = 1 - \int_{-\infty}^{+\infty} p_L(\ell|X = 1) \log_2(1 + e^{-\alpha\ell}) d\ell = 1 - \int_{-\infty}^{+\infty} p_{\alpha L}(\ell|X = 1) \log_2(1 + e^{-\ell}) d\ell$. The proof is given in section A of the appendix. Some interesting properties can be derived in the special case where $I(L, X)$ is a function of a single parameter. In that case, $I(L_y, L_z)$ is a function of two variables as

$$I(L_y, L_z) = J(a_{L_y}) + J(a_{L_z}) - J(u(a_{L_y}, a_{L_z})) \quad (5)$$

where $a_{L_y}, a_{L_y} \in \mathbb{R}^+$ and $u : \mathbb{R}^+ \times \mathbb{R}^+ \mapsto \mathbb{R}^+$. Models (1) and (2) fall under this case, this is made precise below.

Example 1 [2], [19], [23]: L is the Gaussian variable in model (1) then

$$I(L, X) = J(\sigma) \quad (6)$$

with

$$J(\sigma) = 1 - \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(\ell - \frac{\sigma^2}{2}\right)^2}{2\sigma^2}} \log_2(1 + e^{-\ell}) d\ell \quad (7)$$

and J is a monotonically increasing function of σ [24]. The mutual information $I(L_y, L_z)$ is given by (5) with $a_{L_y} = \sigma_{L_y}$, $a_{L_z} = \sigma_{L_z}$ and $u(a_{L_y}, a_{L_z}) = \sqrt{a_{L_y}^2 + a_{L_z}^2}$.

Example 2 [17]: L is the Gaussian variable in model (2) then

$$I(L, X) = J(\alpha\sigma) \quad (8)$$

where J is the function defined in (7). We can observe that even if L is described by two parameters, $I(L, X)$ is again a monotonically increasing function of a single parameter $\alpha\sigma$. The mutual information $I(L_y, L_z)$ is given by (5) with $a_{L_y} = \alpha_{L_y}\sigma_{L_y}$, $a_{L_z} = \alpha_{L_z}\sigma_{L_z}$ and $u(a_{L_y}, a_{L_z}) = \sqrt{a_{L_y}^2 + a_{L_z}^2}$. The proof is given in Appendix A. As a consequence, it is assumed here that J is a monotonically increasing function and is thus reversible and that $u(a_{L_y}, a_{L_z}) \geq \max(a_{L_y}, a_{L_z})$. These two properties hold for example 1 and 2 above. The properties of $I(L_y, L_z)$ are listed below.

- $\max(J(a_{L_y}), J(a_{L_z})) \leq J(\sqrt{a_{L_y}^2 + a_{L_z}^2}) \leq J(a_{L_y}) + J(a_{L_z})$
- $\lim_{a_{L_y} \to 0} I(L_y, L_z) = \lim_{a_{L_y} \to 0} I(L_y, X)$
- $\lim_{a_{L_y} \to \infty} I(L_y, L_z) = I(L_z, X)$ for a given a_{L_z}
- $I(L_y, L_z) \leq min(I(L_y, X), I(L_z, X))$
- If $I(L_y, L_z) = m$ then $I(L_y, X) \geq m$ and $I(L_z, X) \geq m$.

The last property proves that $I(L_y, L_z)$ is a lower bound of the mutual information between the extrinsic and the message X. Thus if, at the end of the iterative process, $I(L_y, L_z)$ is almost equal to 1 so are $I(L_y, X)$ and $I(L_z, X)$.

From this section we can conclude that, the mutual information between extrinsics $I(L_y, L_z)$ is therefore a performance indicator for the whole system and can be used as a stopping criterion or as an error-free sequence indicator at the receiver side, provided that $I(L_y, L_z)$ can efficiently be computed online. This is addressed in section IV-A.

B. On equivalent LLR classes

Let X denote a random variable with equiprobable values in $\{-1; +1\}$. Let $G(X)$ denote the set of random variables following the model in (2) with $\alpha \in \mathbb{R}^+$ and $\sigma \in \mathbb{R}^+$. Let $G_{\alpha_0}(X)$ denote a subset of $G(X)$ such that $\alpha = \alpha_0$ is a given number in $\mathbb{R}^+$. Thus, $G(X) = \bigcup_{\alpha_0 \in \mathbb{R}^+} G_{\alpha_0}(X)$. The LLR considered in (2) spans $G(X)$ whereas the LLR considered in (1) and in density evolution or in EXIT charts spans $G_1(X)$. We first characterize the correspondence between $G(X)$ and $G_1(X)$ based on the mutual information $I(L, X)$.

Result 1: Let $L \in G(X)$. Let $L_\alpha = \alpha L$. Then $I(L, X) = I(L_\alpha, X)$ and $L_\alpha \in G_1(X)$.

Proof: $L_\alpha = \alpha^2 \frac{\sigma^2}{2} X + \alpha\sigma n$ and belongs by construction to $G_1(X)$. From (8), $I(L, X) = J(\alpha\sigma)$ and from (6), $I(L_\alpha, X) = J(\alpha\sigma)$. ■

With the notations in result 1, αL is the unique element in $G_1(X)$ with mutual information with X equal to $I(L, X)$. We prove below the stability of $G_1(X)$ under addition.

Result 2: If $L_a \in G_1(X)$ and $L_b \in G_1(X)$ then $L_a + L_b \in G_1(X)$.

Proof: $L_a = \frac{\sigma_a^2}{2} X + \sigma_a n_a$ and $L_b = \frac{\sigma_b^2}{2} X + \sigma_b n_b$ where n_a and n_b are two independent Gaussian variables with mean 0 and unit variance. $L_a + L_b = \frac{\sigma_a^2 + \sigma_b^2}{2} X + \sqrt{\sigma_a^2 + \sigma_b^2} n$ where n is a Gaussian variable with mean 0 and unit variance. ■

As a consequence, if the extrinsic LLRs of the constituent components are in $G_1(X)$, the a posteriori log-ratio is also in $G_1(X)$. From result 1, it is clear that the scaling has no impact on the mutual information of the actual LLR vector with the message. However it was observed in [14], [15], [16] and in many other references that, if a proper scaling is not used, sub-optimal LLRs may be obtained in future iterations and propagated resulting in worse performance. This point is illustrated below. The system under consideration is a serially-concatenated turbo-code (SCTC) with a $(5, 7)_8$ outer code. The inner code is a convolutional code with generator $\frac{1}{1+D}$. This SCTC will be used through the paper as an illustrative example. The two decoders are considered separately. Let L_{prior} denote the LLR at the input of a decoder with $L_{prior} \in G_\alpha(X)$. The mutual information $I(L_{prior}, X)$ is kept fixed while α is increased from 0.4 to 2.5. The mutual information is measured at the output of each decoder and plotted as a function of α in Fig. 2. We observe for the inner code that, independently of the fixed value of $I(L_{prior}, X)$, the mutual information at the output reaches a maximum when $L_{prior} \in G_1(X)$ ($\alpha = 1$). For the outer code, a maximum is also observed for low value of $I(L_{prior}, X)$ at $\alpha \approx 1$. We can conclude that EXIT charts give a prediction of the highest mutual information that can be obtained at the output of a decoder for a given $I(L_{prior}, X)$ and that this value can be reached provided that the LLR are properly scaled in order to be in $G_1(X)$. Note that the inner decoder receives as an input the LLR computed from the received data which are in $G_1(X)$ whereas L_{prior} is the sole input of the outer decoder which explains the difference observed in Fig. 2 between the inner and outer decoder.

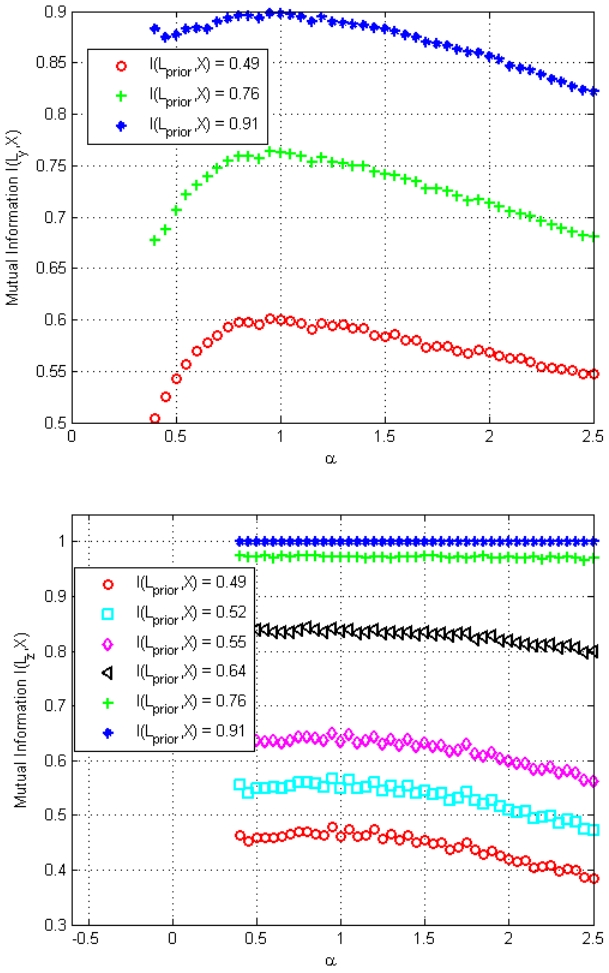

Figure 2. SCTC - Inner decoder (up) - Outer decoder (down) $EbN0 = 2dB$

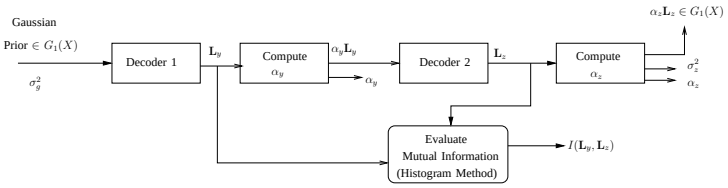

Figure 3. SC-EXIT Chart scheme for turbo-codes

C. Offline estimation : single curve exit charts

This section is devoted to the offline analysis of an iterative decoder. The performance of the association of two component decoders is usually evaluated by tracking the mutual information $I(L, X)$ at the output of each decoder considered separately and for different values of the mutual information at the input. For tracking the evolution of $I(L_y, L_z)$, both components decoders must be considered jointly. Following [19], a single curve EXIT chart (SC-EXIT) is considered here. The principle is described below for turbo and LDPC codes.

1) SC-EXIT for turbo-codes: The principle of SC-EXIT is depicted in Fig. 3. The input (a priori) is modeled with a Gaussian random variable with known mutual information $I(L, X) = J(\sigma)$ with X and is chosen in $G_1(X)$. The process is repeated for different a priori corresponding to different values of σ in the range of interest. The a priori enters decoder 1. The output L_y of decoder 1 is scaled with α_y such that $I(\alpha_y L_y, X) = I(L_y, X)$ and $\alpha_y L_y \in G_1(X)$. Then $\alpha_y L_y$ is given as an input to decoder 2. The output of decoder 2 is L_z, the corresponding scaled LLR is $\alpha_z L_z \in G_1(X)$ with parameter σ_z (σ in (1)). The mutual information $I(L_y, L_z)$ between the extrinsics is evaluated using (4) through an histogram method. The result is denoted $I_M^{(hist)}$ in the following. The computation of α is trivial since X is known. The SC-EXIT gives the evolution of $I(L_y, L_z)$ as a function of σ_g^2 and also the evolution of σ_z^2 as a function of σ_g^2. The system is able to recover the message X if $I(L_y, L_z)$ is an increasing function and if the maximum value is reached or equivalently if $\sigma_z^2 > \sigma_g^2$ [19]. If this condition is not met an oscillatory behaviour may be observed or convergence towards a local minimum may occur. In standard EXIT Charts, the curves of the individual decoders intersect. The SC-EXIT proposed here is very similar to the one in [19] except that $I(L_y, L_z)$ is not considered in [19] (tracking of the variance only) and that in [19] the non-consistency of the LLR is observed without further analysis. The translation of the SC-EXIT into an efficient stopping criterion is straightforward. This will be made explicit in the examples of section V.

2) SC-EXIT for LDPC codes: An LDPC code with m parity-check equations and length-K codewords is considered here. A soft decision algorithm such as sum-product [25] or min-sum [26], [27] is assumed at the receiver. Both algorithms are based on the exchange of extrinsics across the iterations. Tracking the evolution of $I(L_y, L_z)$ is thus also meaningful here for performance evaluation. The decoding rule is briefly recalled and connection with notations L_y and L_z is also given. Denote by $R = (R_1, R_2, ..., R_K)$ the LLR of the received signal. Denote by E_{ji} the extrinsic LLR from check node j to bit node i and M_{ji} the extrinsic LLR from bit node i to check node j. M_{ji} and E_{ji} are not defined if parity-check node j and variable i are not connected in the Tanner graph, otherwise M_{ji} and E_{ji} are updated as:

Initialization

$$M_{j,i} = R_i \tag{9}$$

Check messages

$$E_{ji} = 2tanh^{-1}\left(\prod_{i' \neq i} tanh\left(\frac{M_{ji'}}{2}\right)\right) \quad \text{Sum-Product} \tag{10}$$

$$E_{ji} = \prod_{i' \neq i} \text{sign}\left(M_{ji'}\right) \min_{i' \neq i} |M_{ji'}| \quad \text{Min-Sum} \tag{11}$$

Variable messages

$$M_{ji} = \sum_{j' \neq j} E_{j',i} + R_i \tag{12}$$

E_{ji} can be interpreted as the check node j's opinion on the probability that $X_i = 1$. We will consider also here $E_i = \frac{1}{m_i} \sum_{j \in \mathcal{A}_i} E_{ji}$ where $\mathcal{A}_i$ is the set of check nodes connected with bit i and m_i is the cardinality of $\mathcal{A}_i$. Then, E_i can be interpreted as the opinion of the check nodes on the probability

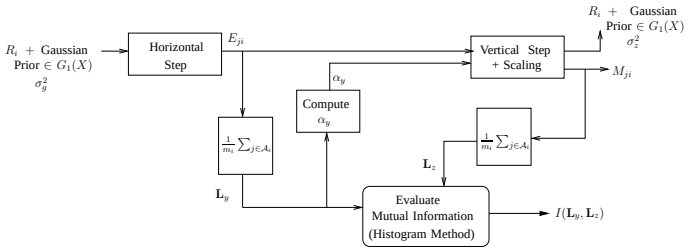

Figure 4. SC-EXIT Chart scheme for LDPC

that $X_i = 1$. In the same way, M_i is defined as $M_i = \frac{1}{m_i}\sum_{j\in\mathcal{A}_i} M_{ji}$. The SC-EXIT chart is given in Fig. 4. The input/output of the SC-EXIT is M_{ji}. To guarantee that $M_{j,i}$ remains in $G_1(X)$ it is sufficient to guarantee that $\sum_{j'\neq j} E_{j'i} \in G_1(X)$ since $R_i \in G_1(X)$ (perfect knowledge of channel characteristics or mismatch correction) and $G_1(X)$ is stable under addition. This is the role of α_y. Let L_y (resp L_z) denote the random variable associated to E_i (resp. to M_i) then L_y and L_z are noisy variables on the message X. As in the turbo-code case, the SC-EXIT Chart in Fig. 4 gives the evolution of $I(L_y, L_z)$ as a function of σ_g^2 and also the evolution of σ_z^2 as a function of σ_g^2.

IV. Efficient computation for finite-length sequences

A. Online computation of the mutual information

Obviously, the convergence behavior of the algorithm involves global quantities (on the whole sequence). Therefore, we are interested in evaluating the mutual information averaged over the whole sequence. As a first step, we consider Q frames each of length B such that $K = BQ$. The average mutual information between two extrinsics is defined as

$$I_M := \frac{1}{Q}\sum_{k=1}^{Q} I(L_{Y_k}, L_{Z_k}) \tag{13}$$

The average extrinsic information is defined as

$$I_E := \frac{1}{Q}\sum_{k=1}^{Q} I(L_k, X_k) \tag{14}$$

We demonstrate below that the average mutual information can be computed based on quantities available at the receiver, namely the extrinsics and the a posteriori probabilities. First recall two key results from [21], [28]. These results (Lemma 1 and Theorem 1) are given in [28] for binary sequences and extended to non-binary iterative decoding in [21]. They are recalled below for the binary case which corresponds to our setting.

Lemma 1: Let $\ell_{y,k}$ be the log ratio of an extrinsic probability at the output of a constituent decoder, let $\ell_{z,k}$ denote the log ratio of an a priori probability at the input of the same decoder and let $\mathbf{y}$ denote the observed sequence. Let $\ell_{z,[k]}$ denote the sequence ℓ_z from which index k is excluded. Then for all X_k in $\{-1;1\}$, $p(X_k = x_k|L_{Y,k} = \ell_{y,k}) = p(X_k = x_k|\mathbf{y}, \ell_{z,[k]})$. Lemma 1 proves that the extrinsics at the output of a constituent decoder contain the same amount of information on X_k than the observation and the a priori. The extrinsic LLR are computed as $\log\left(\frac{p(X_k=1|\mathbf{y},\ell_{z,[k]})}{p(X_k=0|\mathbf{y},\ell_{z,[k]})}\right)$. Based on lemma 1, we have $\log\left(\frac{p(X_k=1|\mathbf{y},\ell_{z,[k]})}{p(X_k=0|\mathbf{y},\ell_{z,[k]})}\right) = \log\left(\frac{p(X_k=1|L_{Y,k}=\ell_{y,k})}{p(X_k=0|L_{Y,k}=\ell_{y,k})}\right)$ which is equal to $\alpha L_{y,k}$ from the generalized consistency and symmetry assumptions. Hence, Lemma 1 proves that $\alpha\mathbf{L}_y$ (resp $\alpha\mathbf{L}_z$) should be propagated through the iterations instead of $\mathbf{L}_y$ or $\mathbf{L}_z$.

Theorem 1: Let $p_z(x_k) = p(X_k = x_k|\mathbf{y}, \ell_{z,[k]})$ represent an ergodic random variable for all X_k in $\{-1;1\}$, then the average mutual information between X and L_z is given by

$$I_E = 1 + \lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\sum_{x_k} p_z(x_k)\log_2(p_z(x_k)) \tag{15}$$

The average extrinsic information can be obtained through time-averaging over a sequence of $K = BQ$ binary symbols. Theorem 1 and (4) lead to a practical evaluation of I_M:

$$\begin{aligned} I_M &= 1 + \lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\sum_{x_k}\Bigg(p_z(x_k)\log_2(p_z(x_k)) \\ &+ p_y(x_k)\log_2(p_y(x_k)) \\ &- \frac{p_y(x_k)p_z(x_k)}{s_k}\log_2\left(\frac{p_y(x_k)p_z(x_k)}{s_k}\right)\Bigg) \end{aligned} \tag{16}$$

where $\sum_{x_k} p_y(x_k) = \sum_{x_k} p_z(x_k) = 1$ and $s_k = \sum_{x_k} p_y(x_k)p_z(x_k)$. The average mutual information can be approximated by time-averaging. The definition of the mutual information leads to another expression for an efficient computation of I_M.

$$\begin{aligned} I_M = \tfrac{1}{Q}\textstyle\sum_{k=1}^{Q}\int_{\ell_{y,k}}\int_{\ell_{z,k}} p(\ell_{y,k},\ell_{z,k})\times \\ \log_2\left(\frac{p(\ell_{y,k},\ell_{z,k})}{p(\ell_{y,k})p(\ell_{z,k})}\right)d\ell_{y,k}d\ell_{z,k} \end{aligned} \tag{17}$$

$$I_M = \tfrac{1}{Q}\textstyle\sum_{k=1}^{Q} E\left[\log_2\left(\frac{p(\ell_{y,k},\ell_{z,k})}{p(\ell_{y,k})p(\ell_{z,k})}\right)\right] \tag{18}$$

The variable X_k can be introduced in the equation above by using the fact that

$$p(\ell_{y,k},\ell_{z,k}) = \sum_{x_k}\frac{p(x_k|\ell_{y,k})p(x_k|\ell_{z,k})}{p(x_k)}p(\ell_{y,k})p(\ell_{z,k})$$

Then the average mutual information reads

$$I_M = \frac{1}{Q}\sum_{k=1}^{Q} E\left[\sum_{x_k}\log_2\left(2\sum_{x_k} p(x_k|\ell_{y,k})p(x_k|\ell_{z,k})\right)\right] \tag{19}$$

By assuming that the random variables involved in I_M are ergodic, we have

$$\begin{aligned} I_M &= \lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\log_2\left(2\sum_{x_k} p(x_k|\ell_{y,k})p(x_k|\ell_{z,k})\right) \\ &= \lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\log_2(2s_k) \end{aligned} \tag{20}$$

By using (16) and (20), we obtain

$$\lim_{K\to\infty} \frac{1}{K}\sum_{k=1}^{K}\sum_{x_k=0}^{1}\Bigg(p_z(x_k)\log_2(p_z(x_k)) + p_y(x_k)\log_2(p_y(x_k)) - \frac{p_y(x_k)p_z(x_k)}{s_k}\log_2\Big(p_y(x_k)p_z(x_k)\Big)\Bigg) = 0 \qquad (21)$$

At this step, two different methods can be used for an online evaluation of I_M. The first method is based on equation (16) evaluated over a sequence of finite length K. This method, called $I_M^{(a)}$, requires the knowledge of the extrinsic probabilities. The computational complexity in terms of number of operations grows linearly with K. The second method is based on equation (20) evaluated over a sequence of finite length K. This method, called $I_M^{(b)}$, has the same characteristics than $I_M^{(a)}$ with the additional advantage of a lower complexity.

The accuracy of these estimators is addressed here through an example. The system under consideration is the serially-concatenated turbo-code (SCTC) of section III-B. The block-length K is a parameter which is increased from 500 to 8000. In this example, $EbN0 = 1dB$ or $EbN0 = 2dB$ (bounds of the waterfall region). We use as a reference the exact expression of I_M in (4) evaluated through a standard histogram method as in section III-C1 and denoted $I_M^{(hist)}$. The comparison involves $I_M^{(a)}$ and $I_M^{(b)}$. The Mean Square Error (MSE) is plotted in Fig. 5. It can be observed that (20) is a better approximation of I_M than (16) which is impaired by the slow convergence of (21). If a precision of 10^{-2} is required on I_M, (20) is an accurate approximation provided that $K \geq 1500$. This precision can not be reached with (16) as long as $K \leq 10000$. For all these reasons, I_M will be approximated at the receiver side by $I_M^{(b)}$ with

$$I_M \approx I_M^{(b)} = 1 + \frac{1}{K}\sum_{k=1}^{K}\log_2(s_k) \qquad (22)$$

The standard average mutual information I_E could also be computed online through approximation (15) giving $I_E^{(a)}$. This approximation is compared with the value given by a histogram method ($I_E^{(hist)}$). The MSE is also plotted in Fig. 5. We can again observe the slower convergence of (15) with the block-length compared to (20).

We proved in this section that the mutual information $I(L_y, L_z)$ can be efficiently estimated online (at the receiver) through (22) without requiring estimation of X_i (hard decisions). The computational complexity of this approximation is linear with K. The mutual information $I(L, X)$ can also be estimated online in a similar manner but with lower accuracy. It is clear that the stopping rule and error detection device should be based on $I(L_y, L_z)$ rather than on $I(L, X)$. Note that the results in this section hold for both Gaussian and non-Gaussian distributed LLRs.

B. Online estimation of α and σ

The scaling factor and its estimation was considered in [14], [15], [16] and in many other references. In most of them, the optimal scaling factor is computed offline and for a given SNR value. These methods are not applicable if the channel conditions are not known with sufficient accuracy at the receiver. Recently, an online estimation method was proposed in [16] for BICM requiring an online numerical search. This method is general and does not require the LLR to be Gaussian distributed or symmetric. In [20], online LLR scaling is also derived for SNR mismatch compensation. The effectiveness of these two methods depend on the reliability of the decoder decisions which may be inaccurate during the first iteration. In the following result, an online estimation method for α and σ is provided. In contrast with most references, closed-form expressions are derived that do not require accurate decoder decisions. Our method is designed to be efficient in the whole range of SNR and at any stage of the iterative process.

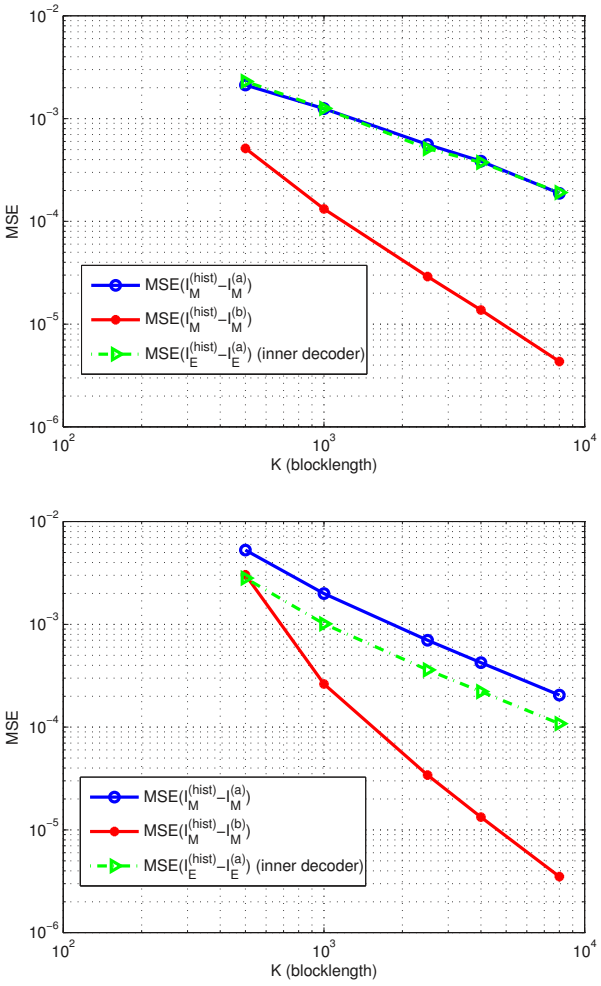

Figure 5. MSE of the difference between $I_M^{(hist)}$ and $I_M^{(a)}$ or $I_M^{(b)}$. MSE of the difference between $I_E^{(hist)}$ and $I_E^{(a)}$ at the output of the inner decoder with $EbN0 = 1dB$ (up) or $EbN0 = 2dB$ (down).

Result 3: Let $L \in G(X)$ with parameters α and σ. Let $I = I(L, X)$ then

$$\sigma^2 = \frac{V(L)}{1+0.25\Big(J^{-1}(I)\Big)^2} \qquad (23)$$

$$\alpha = \frac{J^{-1}(I)}{\sigma} \qquad (24)$$

where J^{-1} is the inverse function of J in (7) and $V(L)$ stands for the variance of L.

Proof: From (2), and considering that X and n are independent random variables, $V(L) = \alpha^2 \frac{\sigma^4}{4} + \sigma^2$. Since $I = J(\alpha\sigma)$ and J is reversible, $\alpha\sigma = J^{-1}(I)$ leading to $V(L) = \left(\frac{(J^{-1}(I))^2}{4} + 1\right)\sigma^2$ which completes the proof. ■

The method in result 3 can be extended to non-Gaussian situation provided that $I(L, X)$ is a non-decreasing function of a single variable and provided that L is given by (2) in which n is not necessarily Gaussian. In the Gaussian case, $J^{-1}(I)$ can be approximated as in [29].

V. Numerical experiments

A. Example 1 : Turbo-Code

As an example, the SCTC described in section III-B is again considered here. It is explained how the mutual information can help for choosing efficiently a stopping criterion. Then, the online estimation of the scaling factor is applied to the SCTC with comparison to alternative methods from the literature. Finally, both scaling factor estimation and stopping rule are included at the decoder, the performance of the decoder is evaluated in terms of BER and number of iterations.

1) SC-EXIT chart and early stopping criterion: The SC-EXIT, for this particular system, are given in Fig. 6 when $EbN0 = 1dB$ or $EbN0 = 2dB$. We observe the following. When $\sigma_g^2 > 5$, the convergence condition ($\sigma_z^2 > \sigma_g^2$) towards the correct decision is fulfilled even when $EbN0 = 1dB$ whereas an oscillatory behavior may be observed when $\sigma_g^2 \leq 5$ preventing the convergence of the iterative process. But at the opposite, if this threshold ($\sigma_z^2 > 5$) is reached the iterative process is likely to converge in few iterations. We take advantage of this fact and propose an accurate stopping criterion as follows. The iterative process for this SCTC should be stopped when either (SC1) or (SC2) below are met:

$$(SC1) \quad I(L_y, L_z) > 1 - \epsilon \quad \text{or} \quad i > i_{max} \tag{25}$$

$$(SC2) \quad i = i_{early} \quad \text{and} \quad \sigma_z < \sqrt{5} \tag{26}$$

where ϵ is a threshold to be fixed, i_{max} is the maximum number of iterations and i_{early} is the number of iterations for deciding of an early stop. The first condition (SC1) is designed to detect error free sequences whereas (SC2) is intended to stop the process at an early stage when the algorithm is unlikely to converge to the global solution. The equivalence between different forms of the stopping rule is given below.

Result 4: Let $L \in G_1(X)$ with parameter σ, the following inequalities are equivalent :

$$\sigma > \sigma_0 \tag{27}$$

$$I(L, X) > J(\sigma_0) \tag{28}$$

$$E(|L|) > M(\sigma_0) \text{ with } M(\sigma_0) = \sigma_0\sqrt{2\pi}e^{\frac{-\sigma_0^2}{8}} + \frac{\sigma_0^2}{2} erf\left(\sqrt{\frac{\sigma_0^2}{8}}\right) \tag{29}$$

$$E(L^2) > \frac{\sigma_0^4}{4} + \sigma_0^2 \tag{30}$$

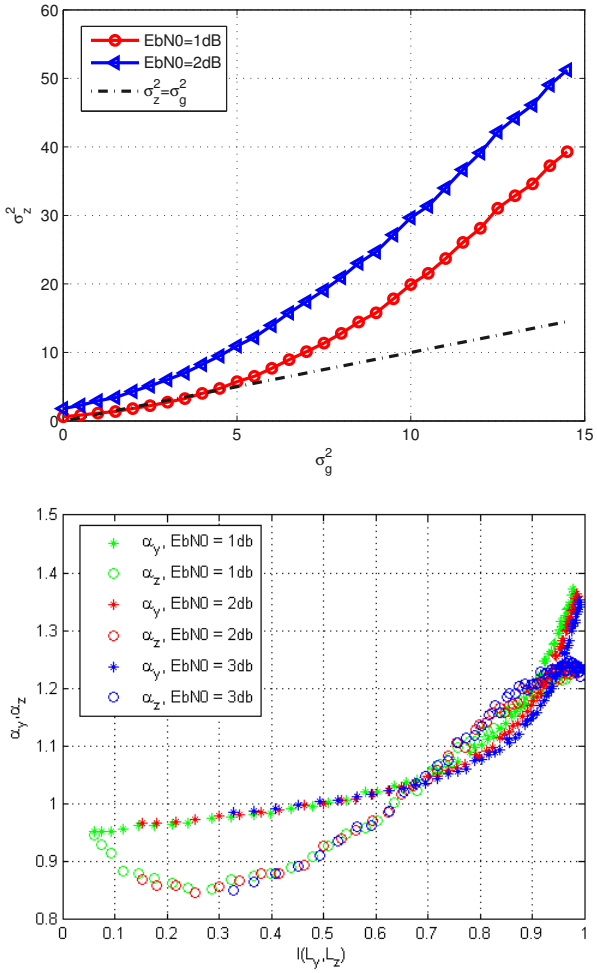

Figure 6. SC-EXIT Chart for the SCTC, up : σ_z^2 vs σ_g^2, down: α_y and α_z vs $I(L_y, L_z)$.

Let $L_y, L_z \in G_1(X)$. If $I(L_y, L_z) > 1 - \epsilon$ then $I(L_y, X) > 1 - \epsilon$ and $I(L_z, X) > 1 - \epsilon$.

Proof: Eq. (28) is equivalent to (27) because J is a strictly increasing function. When $L \in G_1(X)$, simple computation leads to $E(|L|) = M(\sigma)$ where M is a strictly increasing function which proves the equivalence between (29) and (27). The equivalence between (30) and (27) is obvious. The latter statement comes from the properties of $I(L_y, L_z)$. ■

Result 4 proves the equivalence between various stopping criterion commonly used in practice. Stopping rules (27-30) are based on the reliability of the LLR at the output of a given decoder. Alternative methods such as Cross-entropy [7] measure the similarity of the distributions at the output of the two decoders. Once the system has converged to the correct solution, the magnitude of the LLRs has reached a high level (reliability) and the two decoders agree on the decisions (similarity). It is proved below that I_M the estimator of $I(L_y, L_z)$ reaches its maximum if and only if the two conditions are met.

Result 5: $I(L_y, L_z) \approx I_M = 1$ if and only if $\text{sign}(L_{y,i}) = \text{sign}(L_{z,i})$ and $|L_{y,i}| = |L_{z,i}| = +\infty \ \forall i \in \{1, 2..., K\}$.

Proof: $I_M = 1 + \frac{1}{K}\sum_{i=1}^{K} \log_2\left(\frac{1+e^{L_{y,i}+L_{z,i}}}{(1+e^{L_{y,i}})(1+e^{L_{z,i}})}\right)$ and $I_M = 1$ is equivalent to $\frac{1+e^{L_{y,i}+L_{z,i}}}{(1+e^{L_{y,i}})(1+e^{L_{z,i}})} = 1 \; \forall i \in \{1, 2..., K\}$. This condition is met if and only if $\text{sign}(L_y, i) = \text{sign}(L_z, i)$ and $|L_y| = |L_z| = +\infty \; \forall i \in \{1, 2..., K\}$. ■

As an illustration, the shape of I_M and of cross-entropy (CE) are compared in Fig. 7 for a single bit with probability measures p_y anf p_z. We can seen that $1-CE$ reaches its maximum when $p_y(x_k) = p_z(x_k)$ (similarity) whereas I_M reaches its maximum when $p_y(x_k) = p_z(x_k) = 0$ or $p_y(x_k) = p_z(x_k) = 1$ (reliability and similarity).

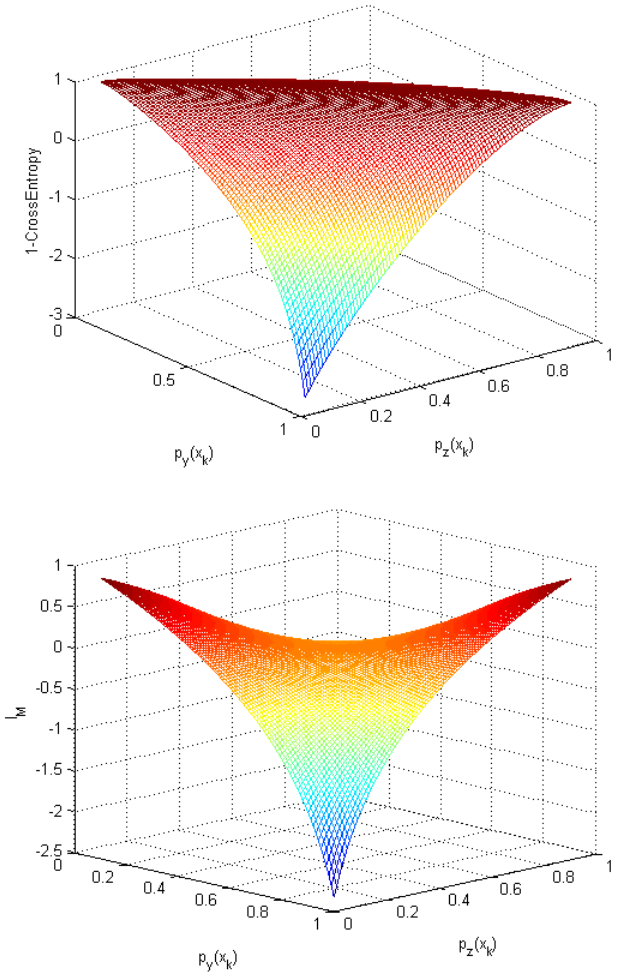

Figure 7. Comparison between $1 - CE$ (up) and I_M (down).

From the SC-EXIT, the typical values of α_y and α_z can also be obtained. They are plotted in Fig. 6 as a function of $I(L_y, L_z)$ and for different values of the SNR. We observe that the value of α varies with the iteration number and with the SNR but this information is captured in $I(L_y, L_z)$. As a consequence, the scaling factor α is a function of a single parameter $I(L_y, L_z)$. We conclude that an alternative method for estimating the value of α consists in evaluating online the mutual information $I(L_y, L_z)$ with (22) and using Fig. 6 to obtain the corresponding α_y and α_z. The coefficients of the polynomials that fits each of the two curves $\alpha_y = f(I(L_y, L_z))$ and $\alpha_z = g(I(L_y, L_z))$ in the least square sense are given in Appendix B.

2) Scaling factor estimation: Two methods for an online estimation of the scaling factor α are proposed in this paper. The efficiency of these methods is measured here for the SCTC. Comparison is also provided with two other methods from the literature:

- Generalized Mutual Information (GMI) [16]. The optimal scaling factor $\hat{s}$ reads:

$$\hat{\alpha} = \arg\min_{\alpha}\{1 - \frac{1}{K}\sum_{i=1} \log_2(1 + \exp(-\text{sign}(\hat{X}_i L_i \alpha)))\}$$

  where $\hat{X}_i$ is the hard decision of the decoder that provides an estimate for X_i. This online method maximizes the generalized mutual information and is proved to be equivalent to offline explicit method for deriving the scaling factor. The method does not require the Gaussian assumption to hold. The scaling factor is obtained from online numerical search (iterative method). The accuracy depends on decoder decision and may be inaccurate at low SNR and at the beginning of the iterative process.
- LLR absolute value [30]. It is assumed here that L is Gaussian distributed as in (2). The scaling factor is obtained by measuring the mean $E(|L|)$ and variance $V(|L|)$ of the absolute value of the LLR and by solving the system of equations below :

$$V(|L|) = \frac{\alpha^2\sigma^4}{4} + \sigma^2 \quad (31)$$

$$\frac{V(|L|)}{E^2[|L|]} = \frac{1+\frac{\alpha^2\sigma^2}{4}}{\left(\sqrt{\frac{2}{\pi}}e^{-\frac{\alpha^2\sigma^2}{8}} + \frac{\alpha\sigma}{2}erf(\frac{\alpha\sigma}{2\sqrt{2}})\right)^2} = h(\alpha\sigma) \quad (32)$$

  A second order polynomial is found in [30] to approximate $h^{-1}\left(\frac{V(|L|)}{E^2[|L|]}\right)$ in the range $0 - 6dB$ avoiding numerical integration.
- Mutual information (MI). This is the method in result 3 where I reads (Eq. (15)):

$$I = I(L, X) \approx 1 + \frac{1}{\log(2)K}\sum_{i=1}^{K}\left(\frac{L_i e^{L_i}}{1+e^{L_i}} - \log(1+e^{L_i})\right)$$

  This method holds as long as J is a strictly increasing function of the variable $\alpha\sigma$. In this paper, J is obtained by assuming Gaussian distributed LLR and $J^{-1}(I)$ is computed with the approximation in the appendix of reference [29].
- SC-EXIT. The scaling factor is estimated with (39) for the inner code and (40) for the outer code. From Eq. (22), the mutual information $I(L_y, L_z)$ reads:

$$I(L_y, L_z) \approx 1 + \frac{1}{K}\sum_{i=1}^{K} \log_2\left(\frac{1+e^{L_{y,i}+L_{z,i}}}{(1+e^{L_{y,i}})(1+e^{L_{z,i}})}\right)$$

The accuracy of the proposed methods is quantified with the normalized mean error as $E(\frac{|\hat{\alpha}-\alpha|}{\alpha})$ where $\hat{\alpha}$ is the estimate provided by one of the methods under consideration and α is the desired scaling factor. The latter is obtained assuming true transmitted bits are available. The MSE is plotted in Fig. 8 as a function of the mutual information in order to

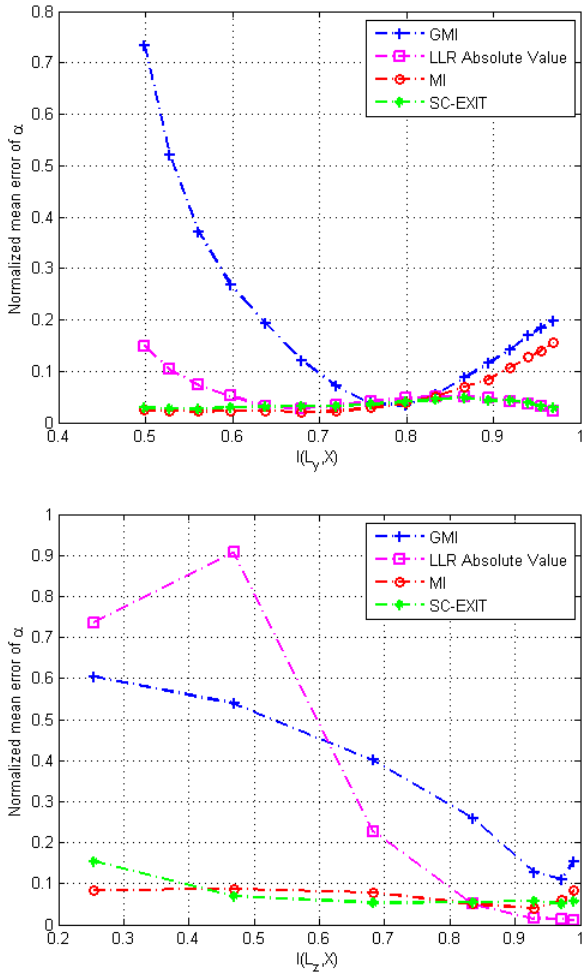

Figure 8. Scaling factor estimation - Output of inner code (up) - Output of outer code (down).

evaluate the efficiency of the method for various SNR and at different stages of the iterative process. As expected, GMI is not sufficiently accurate for low values of the mutual information and is dismissed here since accurate estimation of α is mandatory in early iterations (see Fig. 2). The method in [30] is very sensitive to violation of the Gaussian assumption. This may occur in early iterations especially at the outer code that does not admit the channel observations as an input. At the opposite, we observe that the two methods proposed in this paper are valid over the whole range of SNR and are robust, to a certain extent, to non-Gaussian data. It can be noticed that the Gaussian assumption is not required in the derivation of (15) and (22).

3) BER curves: The optimal parameters deduced from the SC-EXIT are now included at the decoder side. Precisely, we now consider as stopping criterion (SC1) in (25) combined with the early stopping criterion (SC2) in (26) with $i_{early} = 15$. We proved in the previous section that stopping rules can take several equivalent expressions. Two configurations are considered here: *(a)* (SC1) and (SC2) involve $I(L_y, L_z)$ as stopping indicator and α_y and α_z are estimated with the method called SC-EXIT in the previous section, *(b)* (SC1) and (SC2) involve σ as stopping indicator and α_y and α_z are estimated with the method in (24). The thresholds for (SC2) are based on the EXIT curve ($\sigma \leq \sqrt{5}$, $I(L_y, L_z) \leq 0.4$) whereas the thresholds in (SC1) are chosen in order to obtain the same performance in terms of BER for the two methods ($\epsilon = 10^{-2}$ and $\sqrt{\sigma_y^2 + \sigma_z^2} > 20$). The performance is evaluated in terms of BER and WER and is plotted in Fig. 9 with label $SCTC_{optim}$. Comparison is given with the standard decoder ($SCTC_{standard}$) with unscaled LLR and a fixed number of iterations ($n = 15$). Two different codeword lengths are considered: $K = 1000$ and $K = 8000$. In both cases, we observe a slight improvement in the performance with $SCTC_{optim}$. This is due to the propagation of scaled LLR and also to a better repartition of the iterations thanks to the stopping criterion. The average number of iterations is reported in Fig. 9 confirming the validity of (SC1) as an accurate stopping rule based on a performance criterion (error free). At low $EbN0$, stopping rule (SC2) maintains the average number of iterations at a reasonable level even if the maximum number of iterations allowed is large ($i_{max} = 50$ here) with a better efficiency when (SC2) is based on $I(L_y, L_z)$.

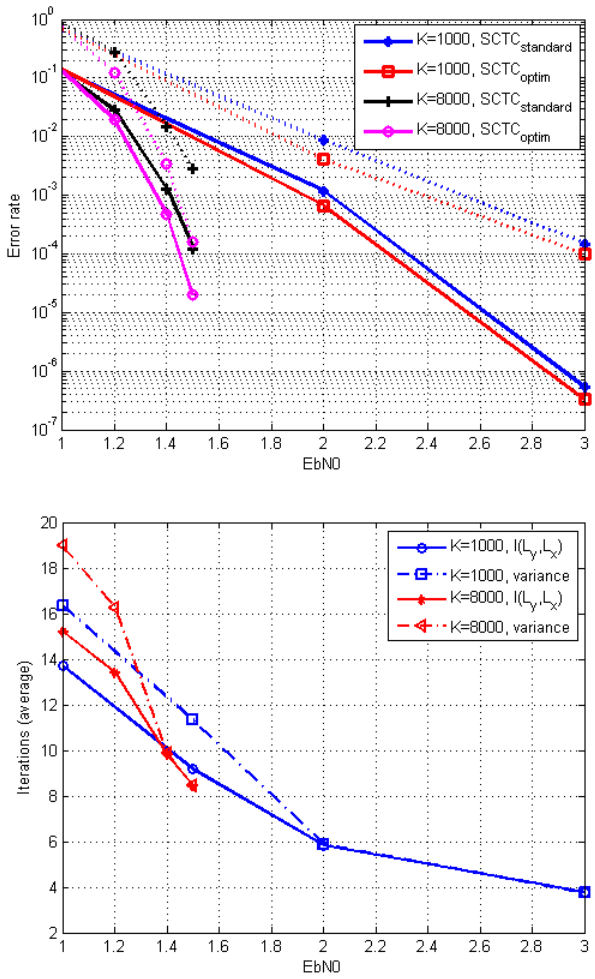

Figure 9. SCTC with $K = 1000$ or $K = 8000$ - Up: BER (solid line), WER (dotted line)- Down: Average number of iterations.

B. Example 2 : LDPC codes

In this section, a (3-6) regular LDPC code and a (4,8) regular LDPC are considered. The SC-Exit charts are given in the section below.

1) SC-Exit chart: The SC-EXIT Chart in Fig. 4 is applied to regular LDPC codes. The evolution of $I(L_y, L_z)$ is plotted in Fig. 10 when either Sum-Product of Min-Sum is implemented. We observe that the curves obtained with Sum-Product and Min-Sum are superimposed when an optimal scaling factor is applied meaning that the two implementations yield the same performance. In several publications, Min-Sum with scaling factor is compared to Sum-Product without scaling for which a slightly worse performance is noted at low SNR. The comparison is unfair, optimal scaling with Sum-Product must correct this phenomenon. The value of the scaling factor α_y is given in Fig. 11. We can see again that α_y is a function of the single variable $I(L_y, L_z)$ since the value obtained for α_y always falls on the same curve independently of $EbN0$ and of the iteration number. From Fig. 11, we can conclude that LLR-scaling is mandatory when using Min-Sum at the decoder side. In addition, we can observe that α_y is not in the immediate vicinity of 1 even when Sum-Product is implemented at the decoder. As with Min-Sum, the LLR are overestimated in the first iterations and should be scaled to improve the performance. The performance of (3-6) and (4-8)-LDPC codes are now evaluated with different implementations at the decoder. We consider Sum-Product with either unscaled LLR (SP, $\alpha = 1$) or optimally scaled LLR (SP, α_{optim}). We also consider Min-Sum either with unscaled LLR (MS, $\alpha = 1$) or with a fixed scaling factor (SP, $\alpha = 0.8$). The value of the scale factor ($\alpha = 0.8$) is obtained from the literature [14], [15] where its optimality is assessed for regular (3-6)-LDPC codes. In this experiment, the codeword-length is $K = 1000$. The stopping criterion is $\mathbf{H}\hat{\mathbf{c}} = 0$ or $i = it_{max}$ where $\mathbf{H}$ is the parity-check matrix, $\hat{\mathbf{c}}$ is the hard-decision on the encoded bits and $it_{max} = 50$. The BER is plotted as a function of $EbN0$ in Fig. 12. We observe the following, scaling LLR always improves the BER. The gain is more important when α is far from 1 and at high $EbN0$. When properly scaled, Min-Sum and Sum-product exhibit the same performance. We can observe that $\alpha = 0.8$ is indeed optimal with (3-6)-LDPC and is close to be optimal with (4-8)-LDPC.

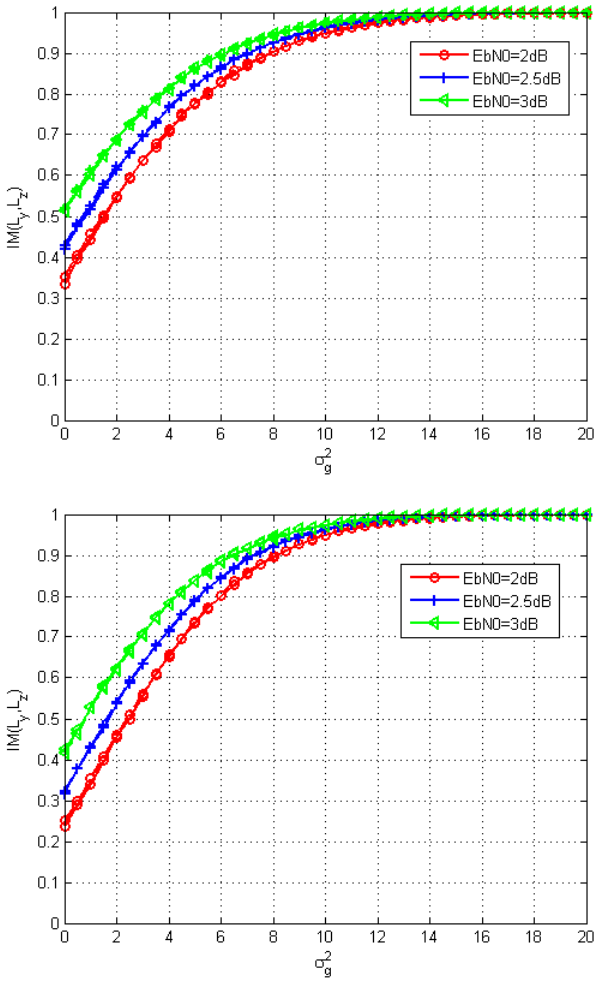

Figure 10. SC-EXIT Chart, $I(L_y, L_z)$ vs σ_g^2 for (3-6)-LDPC (up) and for (4-8)-LDPC (down)- Solid line: Sum-Product, Dotted line: Min-Sum.

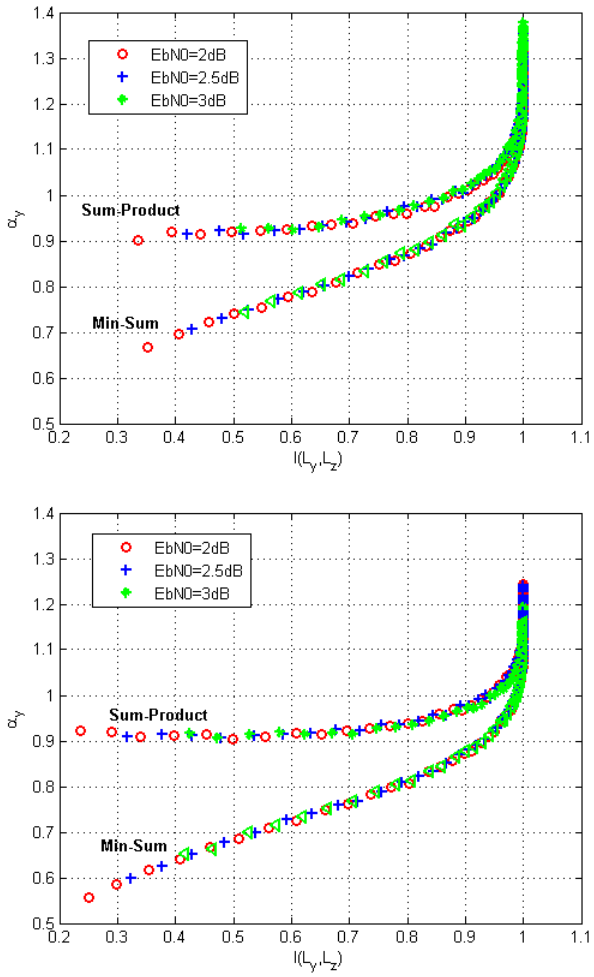

Figure 11. SC-EXIT Chart, α_y vs $I(L_y, L_z)$ for (3-6)-LDPC (up) and (4-8)-LDPC (down).

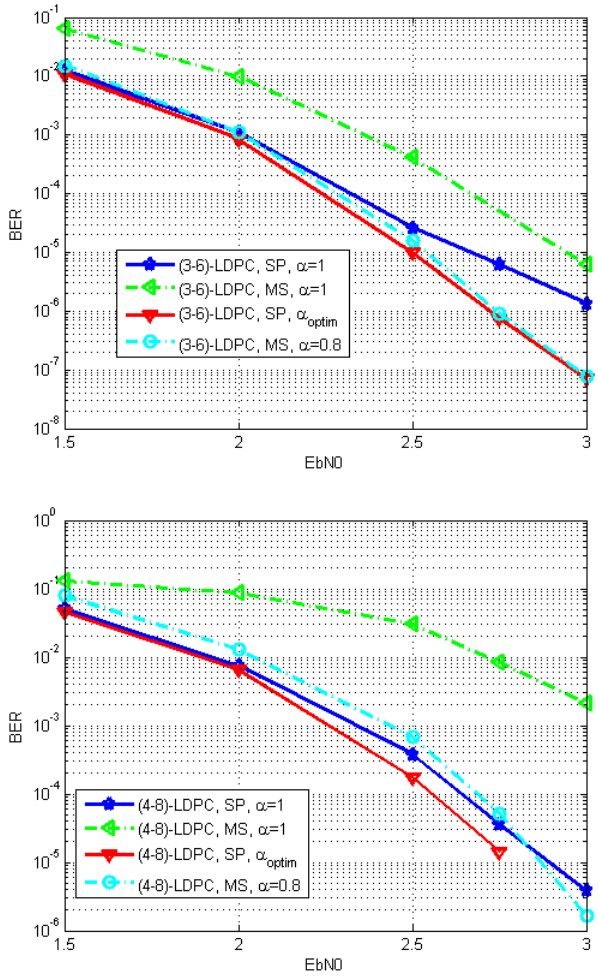

Figure 12. BER curves (3-6)-LDPC (up) and for (4-8)-LDPC (down).

VI. Conclusion

In this paper, we studied the properties of the mutual information between the extrinsics involved in an iterative decoding process. Our framework is quite general and encompasses serially or parallel concatenated turbo-codes or LDPC codes as special cases. We proved that the mutual information between extrinsics is a pertinent performance measure that can be used at the receiver side for error detection purpose. In addition, an offline evaluation, reminiscent of EXIT Charts, gives guidelines for defining efficient stopping rules. Two methods for an online efficient estimation of scaling factors are also derived. Numerical example highlighted the benefits and the generality of the proposed approach.

Acknowledgement

Thanks to Pierre Duhamel for fruitful discussions and constructive criticism.

Appendix

A. Proof of Eq. (4):

In the following, we consider that the three properties (Symmetry, Generalized Consistency and Range) hold for L_y and L_z. We will used the notation α_Y (resp. α_Z) in the Generalized Consistency property when applied to L_y (resp. L_z) and α otherwise (*ie* the result hold for both and is given for L which stands for L_y or L_z).

The expression of $p_L(\ell)$ is given by $p_L(\ell) = \frac{1}{2}\sum_x p_L(\ell|x)$. Since L_y and L_z are independent when conditioned on X, we have

$$\log\left(\frac{p_{L_y,L_z}(\ell_y,\ell_z)}{p_{L_y}(\ell_y)p_{L_z}(\ell_z)}\right) = \log(2) + \underbrace{\log\left(\sum_x p_{L_y}(\ell_Y|x)p_{L_z}(\ell_Z|x)\right)}_{U_{L_y,L_z}} - \underbrace{\log\left(\sum_x p_{L_y}(\ell_y|x)\right)}_{U_{L_y}} - \underbrace{\log\left(\sum_x p_{L_z}(\ell_z|x)\right)}_{U_{L_z}} \quad (33)$$

U_{L_y} and U_{L_z} have similar expression $U_L = \log\left(p_L(\ell|x)\right) + \log\left(1+e^{-\alpha x\ell}\right)$ whereas $U_{L_y,L_z} = \log\left(p_{L_y}(\ell_y|x)p_{L_z}(\ell_z|x)\right) + \log\left(1+e^{-x(\alpha_y\ell_y+\alpha_z\ell_z)}\right)$. The mutual information reads

$$I(L_y,L_z) = 1 + \frac{1}{log_2(2)}\Big(\underbrace{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{L_y,L_z}(\ell_y,\ell_z)U_{L_y,L_z}dl_ydl_z}_{A_{L_y,L_z}} - \underbrace{\int_{-\infty}^{+\infty} p_{L_y}(\ell_y)U_{L_y}dl_y}_{A_{L_y}} - \underbrace{\int_{-\infty}^{+\infty} p_{L_z}(\ell_z)U_{L_z}dl_z}_{A_{L_z}}\Big) \quad (34)$$

where $A_L = \frac{1}{2}\sum_x \int_{-\infty}^{+\infty} p_L(\ell|x)\log(p_L(\ell|x))d\ell + \int_{-\infty}^{+\infty} p_L(\ell|X=1)\log\left(1+e^{-\alpha\ell}\right)d\ell$ and

$$\begin{aligned} A_{L_y,L_z} &= \tfrac{1}{2}\textstyle\sum_x \int_{-\infty}^{+\infty} p_{L_y}(\ell_y|x)\log(p_{L_y}(\ell_y|x))d\ell_y \\ &+\tfrac{1}{2}\textstyle\sum_x \int_{-\infty}^{+\infty} p_{L_z}(\ell_z|x)\log(p_{L_z}(\ell_z|x))d\ell_z \\ &+\textstyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{L_y}(\ell_y|X=1)p_{L_z}(\ell_z|X=1)\times \\ &\log(1+e^{-(\alpha_y\ell_y+\alpha_z\ell_z)})d\ell_yd\ell_z \end{aligned} \quad (35)$$

Then the mutual information reads

$$\begin{aligned} I(L_y,L_z) &= I(L_y,X) + I(L_z,X) \\ &+\textstyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{L_y}(\ell_y|X=1)p_{L_z}(\ell_z|X=1)\times \\ &\log(1+e^{-(\alpha_y\ell_y+\alpha_z\ell_z)})d\ell_yd\ell_z - 1 \end{aligned} \quad (36)$$

We prove now that the last term in (36) is the mutual information between $L_y + L_z$ and X. We can first remark that (substitution in the integral)

$$\begin{aligned} &\textstyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{L_y}(\ell_y|X=1)p_{L_z}(\ell_z|X=1)\times \\ &\log(1+e^{-(\alpha_y\ell_y+\alpha_z\ell_z)})d\ell_yd\ell_z = \\ &\textstyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{\alpha_yL_y}(\ell_y|X=1)p_{\alpha_zL_z}(\ell_z|X=1)\times \\ &\log(1+e^{-(\ell_y+\ell_z)})d\ell_yd\ell_z \end{aligned} \quad (37)$$

Let $u = \ell_y + \ell_z$ and $t = \ell_z$, we have

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{\alpha_y L_y}(\ell_y|X=1)p_{\alpha_z L_z}(\ell_z|X=1)\times \log(1+e^{-(\ell_y+\ell_z)})d\ell_y d\ell_z = \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} p_{\alpha_y L_y}(u-t|X=1)p_{\alpha_z L_z}(t|X=1)dt\right)\times \log(1+e^{-u})du$$

The convolution leads to

$$\begin{aligned} I(L_y,L_z) &= I(L_y,X)+I(L_z,X)+ \\ &\int_{-\infty}^{+\infty} p_{\alpha_y L_y+\alpha_z L_z}(\ell_y+\ell_z|1)\log_2(1+e^{-\ell_y-\ell_z})d\ell_y d\ell_z - 1 \\ &= I(L_y,X)+I(L_z,X)-I(L_y+L_z,X) \end{aligned} \quad (38)$$

which concludes the proof.

B. Approximation of $\alpha = f(I(L_y, L_z)$ for the SCTC

The two curves in Fig. 6 can be approximated with the polynomials below:

$$\begin{aligned} \alpha_y = {} & 5.6574x^6 - 12.3492x^5 + 10.3204x^4 - 3.7379x^3 \\ & +0.5703x^2 + 0.0671x + 0.9461 \end{aligned} \quad (39)$$

$$\begin{aligned} \alpha_z = {} & -9.1055x^5 + 24.933x^4 - 26.7916x^3 \\ & +14.8465x^2 - 3.8655x + 1.2148 \end{aligned} \quad (40)$$

where $x = I(L_y, L_z)$.

REFERENCES

[1] S. ten Brink, "Convergence of iterative decoding," *Electronics Letters*, vol. 35, no. 13, pp. 1117–1119, 1999.

[2] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. on Commun.*, vol. 49, pp. 1727–1737, Oct 2001.

[3] E. Sharon, A. Ashikhmin, and S. Litsyn, "Analysis of low-density parity-check codes based on EXIT functions," *IEEE Transactions on Communications*, vol. 54, no. 8, pp. 1407–1414, 2006.

[4] S. ten Brink and G. Kramer, "Design of repeat-accumulate codes for iterative detection and decoding," *IEEE Transactions on Signal Processing*, vol. 51, no. 11, pp. 2764–2772, 2003.

[5] A. Grant, "Convergence of non-binary iterative decoding," in *GLOBECOM*, Nov. 2001, vol. 2, pp. 1058–1062.

[6] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo codes," in *Proc. IEEE Int. Conf. Commun*, 1993, pp. 1064–1070.

[7] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 429–445, 1996.

[8] R. Shao, S. Lin, and M. Fossorier, "Two simple stopping criteria for turbo decoding," *IEEE Trans. Commun.*, vol. 47, pp. 1117–1120, 1999.

[9] Y Wu, B. Woerner, and J. Ebel, "A simple stopping criterion for iterative decoding," *IEEE Communications Letters*, vol. 4, pp. 258–260, 2000.

[10] T. Ngatched and R. Takawira, "Simple stopping criterion for iterative decoding," *Electronics Letters*, vol. 37, pp. 1350–1351, 2001.

[11] L. Kocarev, F. Lehmann, G.M. Maggio, B. Scanavino, Z. Tasev, and A. Vardy, "Nonlinear dynamics of iterative decoding systems: analysis and applications," *IEEE Trans. on Infor. Theory*, vol. 52, no. 4, pp. 1366–1384, 2006.

[12] F. Zhai and I.J. Fair, "New error detection techniques and stopping criteria for turbo decoding," in *Canadian Conference on Electrical and Computer Engineering*, 2000, vol. 1, pp. 58–62 vol.1.

[13] Fan-Min Li and An-Yeu Wu, "On the new stopping criteria of iterative turbo decoding by using decoding threshold," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5506–5516, 2007.

[14] Jinghu Chen, A. Dholakia, E. Eleftheriou, M.P.C. Fossorier, and Xiao-Yu Hu, "Reduced-Complexity Decoding of LDPC Codes," *IEEE Transactions on Communications*, vol. 53, no. 8, pp. 1288–1299, Aug 2005.

[15] A. Alvarado, V. Nunez, L. Szczecinski, and E. Agrell, "Correcting suboptimal metrics in iterative decoders," in *Communications, 2009. ICC '09. IEEE International Conference on*, June 2009, pp. 1–6.

[16] Jinhong Wu, M. El-Khamy, Jungwon Lee, and Inyup Kang, "BICM performance improvement via online LLR optimization," in *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*, April 2013, pp. 3850–3855.

[17] M. Fu, "Stochastic analysis of turbo decoding," *IEEE trans on Inform. Theory*, vol. 51, no. 1, pp. 81–100, 2005.

[18] M. Fu, "On Gaussian approximation for density evolution of low-density parity-check codes," in *IEEE International Conference on Communications, 2006. ICC '06.*, June 2006, vol. 3, pp. 1107–1112.

[19] D. Divsalar, S. Dolinar, and F. Pollara, "Iterative turbo decoder analysis based on density evolution," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 5, pp. 891–907, 2001.

[20] M. El-Khamy, J. Wu, J. Lee, and I. Kang, "Online log-likelihood ratio scaling for robust turbo decoding," *Communications, IET*, vol. 8, no. 2, pp. 217–226, January 2014.

[21] J. Kliewer, S. Xin Ng, and L. Hanzo, "Efficient computation of EXIT functions for nonbinary iterative decoding," *IEEE Trans. on Commun.*, vol. 54, no. 12, pp. 2133–2136, December 2006.

[22] J. Hagenauer, "The EXIT chart: Introduction to extrinsic information transfer in iterative processing," in *Proceedings of Eur. Signal Process. Conf.*, Vienna, Austria, Sept 2004, pp. 1541 –1548.

[23] H. El-Gamal and Jr. Hammons, A.R., "Analyzing the turbo decoder using the gaussian approximation," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 671–686, 2001.

[24] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, New York, Wiley, 1991.

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

[26] M.P.C. Fossorier, M. Mihaljevic, and H. Imai, "Reduced complexity iterative decoding of low-density parity check codes based on belief propagation," *IEEE Transactions on Communications*, vol. 47, no. 5, pp. 673–680, May 1999.

[27] N. Wiberg, *Codes and decoding on general graphs*, Ph.D. thesis, Link oping Univ., Link oping, Sweden, 1996.

[28] I. Land, P. Hoeher, and S. Gligorević, "Computation of symbol-wise mutual information in transmission system with log APP decoders and application to EXIT charts," in *in Proc. Int. ITG Conf. Source Channel Coding*, Erlangen, Germany, Jan. 2004, pp. 195–202.

[29] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Transactions on Communications*, vol. 52, no. 4, pp. 670–678, April 2004.

[30] T.A. Summers and S.G. Wilson, "SNR mismatch and online estimation in turbo decoding," *IEEE Transactions on Communications*, vol. 46, no. 4, pp. 421–423, Apr 1998.