Measures for Tracing Convergence of Iterative Decoding Algorithms

M. Tüchler†, S. ten Brink‡, and J. Hagenauer†

† Institute for Comm. Engineering, Munich University of Technology, Arcisstr. 21, D-80290 Munich, Germany, email: {micha,hag}@lnt.ei.tum.de

‡ Bell Laboratories, Lucent Technologies, 101 Crawfords Corner Road, Holmdel, NJ 07733, USA, email: stenbrink@bell-labs.com

Abstract— **We study the convergence behavior of turbo decoding, turbo equalization, and turbo bit-interleaved coded modulation in a unified framework, which is to regard all three principles as instances of iterative decoding of two serially concatenated codes. There is a collection of measures in the recent literature, which trace the convergence of iterative decoding algorithms based on a single parameter. This parameter is assumed to completely describe the behavior of the soft-in soft-out decoders being part of the iterative algorithm. The measures observe different parameters and were originally applied to different types of decoders. In this paper, we show how six of those measures are related to each other and we compare their convergence prediction capability for the decoding principles mentioned above. We observed that two measures predict the convergence very well for all regarded decoding principles and others suffer from systematic prediction errors independent of the decoding principle.**

I. Introduction

Today, there exist error correction codes (ECCs) for data transmission over standard channels, e.g., the additive white Gaussian noise (AWGN) channel, yielding a decoding performance close to theoretical limits. Among those is the family of the concatenated codes, whose potential was soon discovered [1], but efficient decoding was considered prohibitively complex. Berrou et al. showed in 1993 that a concatenated code can be decoded almost optimally with low computational burden using iterative decoding [2]. This finding spawned a huge amount of research on these "turbo codes" (as a reference, see e.g. [3]). The decoding principle was applied later to other data transmission systems, which can be regarded as a concatenation of two or more "encoders" processing the data to be transmitted, e.g., coded data transmission over an inter-symbol interference (ISI) channel [4], [5], [6], bit–interleaved coded modulation (BiCM) [7], trellis coded modulation (TCM) [8], [9], or coded code-division multiple-access (CDMA) [10], [11]. The receiver of one of the first three systems is said to perform turbo equalization, turbo TCM, or turbo BiCM, respectively.

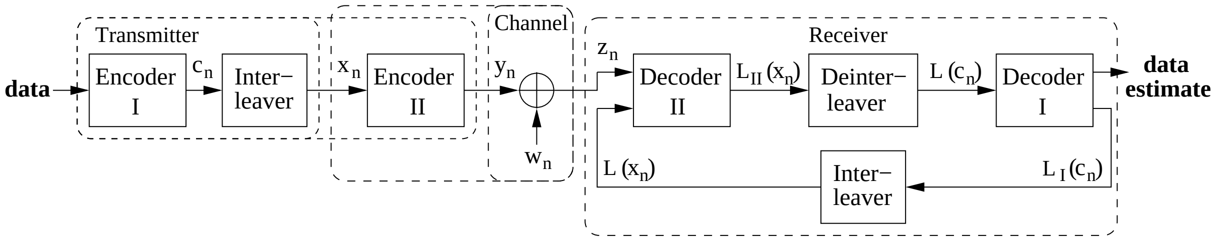
A large body of research has been undertaken to provide tools for choosing design parameters for turbo codes and suitable codes have been found, e.g., in [12], [13], [14]. A long time open problem was to understand the convergence behavior of the iterative decoding algorithm. A major question was to explain the regions in signal-to-noise ratio (SNR), where performance improvement over the iterations occurs quickly or not at all. These regions are separated by a transition called "waterfall". In [15], [16], the turbo decoder was modeled as a high-dimensional nonlinear dynamic system. By characterizing the fixed points of the system, the mentioned SNR regions could be explained and determined. Another approach was to investigate the probability density functions (PDFs) of the communicated information between the decoders. Based on the evolution of these PDFs, thresholds on the convergence of the iterative decoding algorithm for low density parity check (LDPC) codes have been obtained [17]. A simplified approach assumes that the considered PDFs are from a single parameter Gaussian family, an idea, which has been applied to study quite successfully the decoding performance of LDPC codes [18], turbo codes [19], [14], [20], turbo BiCM [7], and turbo equalization [6]. A similar analysis tool based on the observation of a single parameter was used in [11], [21].

It turns out that these analyses on decoding convergence are based on the same assumptions but they extract different parameters of the considered PDFs. This leads to different results and to different ranges of the charts depicting the parameter evolution. A drawback of all existing solutions is that the transmitted data is required to derive the parameter evolution charts. Thus, the analysis must be offline. However, it can be useful to obtain these charts online in a receiver, e.g., to select a suitable equalization algorithm in turbo equalization depending on the quality of the feedback information as demonstrated in [6]. In this paper, we compare algorithms predicting the decoding convergence of iterative algorithms in an identical parameter range, which simplifies the comparison. We introduce an approach, which does not require the transmitted data.

The paper is organized as follows. In Section II, we define three data transmission systems used for the comparison. In Section III, we introduce six measures for predicting the decoding convergence, whose convergence prediction capability is investigated in Section IV. From the results obtained there we draw some conclusions in Section V.

II. A generalized serially concatenated system

Figure 1 depicts three serially concatenated systems consisting of two component codes separated by an interleaver, which all transmit data over an AWGN channel. The concatenation of two encoders for a convolutional ECC, one



configuration	en-/decoder I (outer code)	en-/decoder II (inner code)
turbo code	ECC en-/decoder	ECC en-/decoder
turbo equalization	ECC en-/decoder	ISI channel/detector
turbo BiCM	ECC en-/decoder	mapper/demapper

Fig. 1. A generalized serially concatenated system.

encoder for a convolutional ECC and an ISI channel, or one encoder for a convolutional ECC and a signal mapper is considered and the receiver performs iterative decoding, iterative equalization and decoding, or iterative demapping and decoding, respectively. We refer to these systems as turbo code system, turbo equalization system, and turbo BiCM system.

For all systems, the (binary) data is encoded blockwise with the (binary) convolutional "outer" encoder I to N_{I} code bits $c_n \in \mathcal{B}$, $\mathcal{B}=\{0,1\}$, $n=1,2,...,N_{\mathrm{I}}$. The interleaver permutes these bits to x_n, $n=1,...,N_{\mathrm{I}}$. The deinterleaver reverses the interleaver permutation.

For the turbo code system, the x_n are encoded with the (binary) convolutional "inner" encoder II to N_{II} code symbols $y_n \in \tilde{\mathcal{B}}$, $\tilde{\mathcal{B}}=\{+1,-1\}$, $n=1,2,...,N_{\mathrm{II}}$, which are transmitted over an AWGN channel. The PDF of the noise samples w_n is given by $f_w(w)=\phi_{0,\sigma_w^2}(w)$,

$$\phi_{\mu,\sigma^2}(w) = e^{-(w-\mu)^2/(2\sigma^2)}/\sqrt{2\pi\sigma^2}, \;\; w,\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+,$$

i.e., the noise variance is σ_w^2. Received is $z_n=y_n+w_n$.

For the turbo equalization system, the x_n are mapped to the signal alphabet $\tilde{\mathcal{B}}$ and transmitted over an ISI channel with impulse response $h[n] = \sum_{k=0}^{M-1} h_k \delta[n-k]$, $h_k \in \mathbb{R}$, assumed to be known to the receiver. Received are the symbols $z_n = (\sum_{k=0}^{M-1} h_k x_{n-k}) + w_n$. For turbo equalization using higher order signal constellations see e.g. [22].

For the turbo BiCM system, the x_n are mapped to symbols $y_n \in \mathbb{C}$ from a Q-ary signal alphabet. Received is $z_n=y_n+w_n$, $w_n \in \mathbb{C}$, where the real and the imaginary part of w_n are distributed with $\phi_{0,\sigma_w^2}(w)$.

In the receiver, we assume that symbol-based maximum a-posteriori probability (MAP) decoding algorithms, e.g. the "BCJR" algorithm [23], are used. The decoder II processes the received symbols z_n and outputs the log-likelihood ratio (LLR)

$$L_{\mathrm{II}}(x_n) = \ln \frac{P(x_n=0 \mid z_1, z_2, \cdots, z_{N_{\mathrm{II}}})}{P(x_n=1 \mid z_1, z_2, \cdots, z_{N_{\mathrm{II}}})} - \ln \frac{P(x_n=0)}{P(x_n=1)},$$

which is the a-posteriori LLR minus the a-priori LLR $L(x_n) = \ln(P(x_n=0)/P(x_n=1))$. We refer to [2], [3], [24] for turbo codes, [4], [5], [6] for turbo equalization, and [7] for turbo BiCM for more information on how to obtain or approximate $L_{\mathrm{II}}(x_n)$. After deinterleaving, $L_{\mathrm{II}}(x_n)$ is considered a-priori LLR $L(c_n) = \ln(P(c_n=0)/P(c_n=1))$ for the decoder I, which outputs

$$L_{\mathrm{I}}(c_n) = \ln \frac{P(c_n=0 \mid L(c_1), \cdots, L(c_{N_{\mathrm{I}}}))}{P(c_n=1 \mid L(c_1), \cdots, L(c_{N_{\mathrm{I}}}))} - \ln \frac{P(c_n=0)}{P(c_n=1)},$$

and estimates of the transmitted data. The LLR $L_{\mathrm{I}}(c_n)$ is often called extrinsic information in the literature [3]. Applying the turbo principle, the LLRs $L_{\mathrm{I}}(c_n)$ are interleaved and regarded as a-priori LLR $L(x_n)$ for decoder I. After an initial decoding step, where $L(x_n)=0$, for all x_n, is assumed, the receiver iterates between decoder I and II until a termination criterion stops the iterative process.

III. Measures of the transfer characteristics

A. Definition

The two decoders can be modeled as devices mapping a sequence of LLRs L_{in} to a new sequence of LLRs L_{out}, where decoder I maps $L(c_n)$ to $L_{\mathrm{I}}(c_n)$ and decoder II maps $L(x_n)$ to $L_{\mathrm{II}}(x_n)$. The sequence of random variables (r.v.'s) L_{in} is assumed to be independent and identically distributed (i.i.d.) according to a single parameter PDF $f_{in}(l|X=\tilde{x})$ conditioned on the value $\tilde{x} \in \tilde{\mathcal{B}}$ of the r.v. X representing c_n (decoder I) or x_n (decoder II), respectively:

$$f_{in}(l \mid X=\tilde{x}) = \phi_{2\tilde{x}\gamma_{in}, 4\gamma_{in}}(l), \tag{1}$$

where we use the following correspondence between the alphabets $\mathcal{B}$ and $\tilde{\mathcal{B}}$:

$$\tilde{c}_n = \begin{cases} +1, & c_n=0, \\ -1, & c_n=1, \end{cases} \qquad \tilde{x}_n = \begin{cases} +1, & x_n=0, \\ -1, & x_n=1. \end{cases}$$

We denote $f_{in}(l|X=\tilde{x})$ briefly as $f_{in}(l \mid \tilde{x})$. The statistics $\mu_{in} = E(L_{in}|X=+1)$ (mean) and $\sigma_{in}^2 = Var(L_{in}|X=+1)$ (variance) of L_{in} reveal that γ_{in} is chosen to be the "signal-to-noise ratio" (SNR)

$$\mu_{in}^2/\sigma_{in}^2 = (2\gamma_{in})^2/(4\gamma_{in}) = \gamma_{in} \tag{2}$$

of the LLRs L_{in}. The PDF $f_{in}(l \mid \tilde{x})$ is motivated by the fact that the LLR

$$L(z|y) = \ln(p(z|y=+1)/p(z|y=-1)) = 2/\sigma_w^2 \cdot z \tag{3}$$

computed from the output z of an AWGN channel with noise variance σ_w^2 and input $y \in \tilde{\mathcal{B}}$ has a distribution of type (1). The SNR of $L(z|y)$ is equal to $1/\sigma_w^2$.

The crucial observation is that the PDF $f_{out}(l|X=\tilde{x})$, denoted as $f_{out}(l|\tilde{x})$, of the output LLRs L_{out} is reasonably well approximated by a PDF of type (1) for a second parameter γ_{out}. Since an output LLR is input LLR to the following decoder, the PDF of the LLRs communicated between the decoders retains to be of type (1) and, thus, the iterative process is completely described by the evolution of the parameter γ_{out}. Another assumption is that the input LLRs $L(x_n)$ and $L(c_n)$ are i.i.d. samples of the r.v. L_{in}, which is plausible in the receiver for large interleaver block lengths, at least for several iterations. With these findings it is possible to describe both decoders with transfer functions mapping a single real-valued input to a single real-valued output parameter. These functions are obtained by generating a-priori LLRs $L(x_n)$ or $L(c_n)$, respectively, according to the PDF $f_{in}(l|\tilde{x})$ for some γ_{in} and presenting them to each decoder separately. After decoding, the PDF $f_{out}(l|\tilde{x})$ is estimated using a sufficient number of output LLRs $L_{\mathrm{I}}(c_n)$ or $L_{\mathrm{II}}(x_n)$ conditioned on c_n or x_n, respectively. We do not attempt an analysis of $f_{out}(l|\tilde{x})$ or even γ_{out}, which is rather challenging and furthermore different for each system introduced here due to the different system components involved.

There are several measures to extracting and displaying information from $f_{out}(l|\tilde{x})$ in the literature [20], [19], [14], [25], [11], which are differently related to γ_{out}. Six measures are presented in the following.

M1: In [20], the average mean

$$\mu_{out} = \sum_{\tilde{x}\in\tilde{\mathcal{B}}} P(X=\tilde{x}) \cdot \int_{-\infty}^{\infty} \tilde{x} \cdot l \cdot f_{out}(l|\tilde{x})\, \mathrm{d}\, l$$

from the two output PDFs $f_{out}(l|\tilde{x})$ is computed, which is the average of the mean of the output LLRs conditioned on c_n or x_n, respectively. Assuming that the c_n or x_n are equally likely 0 or 1 yields that $P(X=+1) = P(X=-1) = 1/2$. Estimating $f_{out}(l|\tilde{x})$ using N_{I} LLRs $L_{\mathrm{I}}(c_n)$ from decoder I yields that

$$\mu_{out} \approx \frac{1}{N_{\mathrm{I}}} \sum_{n=1}^{N_{\mathrm{I}}} \tilde{c}_n \cdot L_{\mathrm{I}}(c_n). \tag{4}$$

Similarly, using N_{II} LLRs $L_{\mathrm{II}}(x_n)$ from decoder II, μ_{out} is approximated by $\frac{1}{N_{\mathrm{II}}}\sum_{n=1}^{N_{\mathrm{II}}} \tilde{x}_n L_{\mathrm{II}}(x_n)$. Since $f_{out}(l|\tilde{x})$ is assumed to be of type (1), the variance at the output is a function of μ_{out} and need not be computed. Thus, the output SNR γ_{out} is given by

$$\gamma_{out} = \mu_{out}^2/(2\mu_{out}) = \mu_{out}/2,$$

which follows from (2).

M2: In [14], the average variance

$$\sigma_{out}^2 = \frac{1}{2} \sum_{\tilde{x}\in\tilde{\mathcal{B}}} \int_{-\infty}^{\infty} \tilde{x} \cdot (l^2 - \mu_{out}^2) \cdot f_{out}(l|\tilde{x})\, \mathrm{d}\, l,$$

i.e., the variance of the output LLRs, is computed as a derivation of the measure M1. Estimating $f_{out}(l|x)$ using, e.g., N_{I} LLRs $L_{\mathrm{I}}(c_n)$ from decoder I yields that

$$\sigma_{out}^2 \approx \frac{1}{N_{\mathrm{I}}} \sum_{n=1}^{N_{\mathrm{I}}} L_{\mathrm{I}}(c_n)^2 - \mu_{out}^2,$$

where μ_{out} is computed as in (4). Since $f_{out}(l|\tilde{x})$ is assumed to be of type (1), the output SNR γ_{out} is given by

$$\gamma_{out} = (\sigma_{out}^2/2)^2/\sigma_{out}^2 = \sigma_{out}^2/4,$$

i.e., the mean which follows from (2) is used instead of μ_{out} to compute γ_{out}.

M3: In [19], the error probability

$$P_b = \frac{1}{2} \sum_{\tilde{x}\in\tilde{\mathcal{B}}} \int_{-\infty}^{0} \tilde{x} \cdot f_{out}(l|\tilde{x})\, \mathrm{d}\, l$$

of a wrong decision $\mathrm{sign}(L_{\mathrm{I}}(c_n)) \neq \tilde{c}_n$ or $\mathrm{sign}(L_{\mathrm{I}}(x_n)) \neq \tilde{x}_n$, respectively, is computed. Estimating $f_{out}(l|\tilde{x})$ using, e.g., N_{I} LLRs $L_{\mathrm{I}}(c_n)$ yields that

$$P_b \approx \frac{1}{N_{\mathrm{I}}} \sum_{n=1}^{N_{\mathrm{I}}} 1/2 \cdot (1 - \tilde{c}_n \cdot \mathrm{sign}(L_{\mathrm{I}}(c_n))).$$

From (1) and (2) follows that $P_b = Q(\mu_{in}/\sigma_{in}) = Q(\sqrt{\gamma_{in}})$, where $Q(x) = \int_x^{\infty} \phi_{0,1}(l)\, \mathrm{d}\, l$, and that the output SNR γ_{out} is given by

$$\gamma_{out} = Q^{-1}(P_b)^2.$$

M4: In [14], the mutual information

$$I(X; L_{out}) = \frac{1}{2} \sum_{\tilde{x}\in\tilde{\mathcal{B}}} \int_{-\infty}^{\infty} f_{out}(l|\tilde{x}) \cdot \log_2 \frac{2\, f_{out}(l|\tilde{x})}{f_{out}(l|+1) + f_{out}(l|-1)}\, \mathrm{d}\, l, \tag{5}$$

between the r.v. X and L_{out} is computed without imposing assumption (1) on $f_{out}(l|\tilde{x})$. The integral above is evaluated by numerical integration using a histogram of, e.g., N_{I} LLRs $L_{\mathrm{I}}(c_n)$, to estimate $f_{out}(l|\tilde{x})$.

M5: In [25], the fidelity

$$\phi_{out} = \frac{1}{2} \sum_{\tilde{x}\in\tilde{\mathcal{B}}} \int_{-\infty}^{\infty} \tilde{x} \cdot \tanh(l/2) \cdot f_{out}(l|\tilde{x})\, \mathrm{d}\, l,$$

is computed. Estimating $f_{out}(l|\tilde{x})$ using, e.g., N_{I} LLRs $L_{\mathrm{I}}(c_n)$ from decoder I yields that

$$\phi_{out} \approx \frac{1}{N_{\mathrm{I}}} \sum_{n=1}^{N_{\mathrm{I}}} \tilde{c}_n \cdot \tanh(L_{\mathrm{I}}(c_n)/2)). \tag{6}$$

This measure is computed from the correlation between a symbol $\tilde{c}_n$ or $\tilde{x}_n$ and its soft estimate $E_L(\tilde{c}_n)$ or $E_L(\tilde{x}_n)$, respectively, given the output LLR $L_{\mathrm{I}}(c_n)$ or $L_{\mathrm{II}}(x_n)$. For example, the soft estimate $E_L(\tilde{c}_n)$ of $\tilde{c}_n$ is given by

$$\begin{aligned} E_L(\tilde{c}_n) &= \sum_{\tilde{c}\in\tilde{\mathcal{B}}} \tilde{c} \cdot P(\tilde{c}_n = \tilde{c} | L_{\mathrm{I}}(c_n)) \\ &= \frac{e^{L_{\mathrm{I}}(c_n)}}{1+e^{L_{\mathrm{I}}(c_n)}} - \frac{1}{1+e^{L_{\mathrm{I}}(c_n)}} = \tanh(L_{\mathrm{I}}(c_n)/2) \end{aligned}$$

and the measure ϕ_{out} is the expectation $E(\tilde{c}_n \cdot E_L(\tilde{c}_n))$ with respect to LLRs $L_{\mathrm{I}}(c_n)$ distributed with $f_{out}(l|\tilde{x})$. The fidelity measure is similar to the measure used in [11], [21], where the variance $Var(\tilde{c}_n - E_L(\tilde{c}_n))$ is observed. A negative ϕ_{out} results if the number of wrong decisions, e.g., $\mathrm{sign}(L_{\mathrm{I}}(c_n)) \neq \tilde{c}_n$, outweighs the number of correct decisions. This corresponds to a error probability P_b larger than $1/2$. The fidelity ϕ_{out} is thus restricted to the range $[0,1]$, where $\phi_{out}=0$ corresponds to a maximally unreliable estimate ($L_{\mathrm{I}}(c_n)=0$) and $\phi_{out}=1$ to a maximally reliable estimate ($|L_{\mathrm{I}}(c_n)| \to \infty$). We note that negative ϕ_{out} might occur if it is approximated using (6), which is due to numerical inaccuracies or an insufficient amount of data.

M6: The measures M1-M5 are based on the conditional PDFs $f_{out}(l|\tilde{x})$, which requires the knowledge of c_n or x_n, respectively. This disqualifies their application in a receiver. However, using convergence prediction could be advantageous, e.g., to select suitable decoding algorithms depending on the current state of the iterative process [6]. Assuming that the c_n and x_n are equally likely 0 or 1, an approach using the PDF $f_{out}(l) = (f_{out}(l|{+1}) + f_{out}(l|{-1}))/2$ would not require these symbols. We introduce a measure M6 related to M2, which computes the second moment

$$\eta_{out} = \int_{-\infty}^{\infty} l^2 \cdot f_{out}(l)\,\mathrm{d}\,l$$

of $f_{out}(l)$ to obtain the parameter γ_{out}. Estimating $f_{out}(l)$ using, e.g., N_{I} LLRs $L_{\mathrm{I}}(c_n)$ yields that

$$\eta_{out} \approx \frac{1}{N_{\mathrm{I}}} \sum_{n=1}^{N_{\mathrm{I}}} L_{\mathrm{I}}(c_n)^2.$$

Since $f_{out}(l|\tilde{x})$ is assumed to be of type (1), the output LLRs can be thought of being generated from the output of an AWGN channel with noise variance $1/\gamma_{out}$ using the equivalence (3). An approach to estimate this variance from a sequence of LLRs without requiring the transmitted (binary) data is available in [26], where the estimate was used in a stopping criterion for the iterative process. The estimation formula $(2+2\sqrt{1+\eta_{out}})/\eta_{out}$ is used here to estimate the inverse of this variance, the output SNR γ_{out}:

$$\gamma_{out} = \eta_{out}/(2 + 2\sqrt{1+\eta_{out}}) = (\sqrt{1+\eta_{out}} - 1)/2.$$

B. Implementation

Repeating the measurements explained above for several input SNRs $\gamma_{in} \in \mathbb{R}^+$ yields a transfer function $\gamma_{in} \to \gamma_{out}$, $\gamma_{out} \in \mathbb{R}^+$, (M1-M3, M6), $\gamma_{in} \to I(X; L_{out})$, $I(X; L_{out}) \in [0,1]$, (M4) or $\gamma_{in} \to \phi_{out}$, $\phi_{out} \in [0,1]$, (M5) for each decoder. To study the convergence behavior based on the trajectory of the iterative algorithm, i.e., the sequence of the considered parameter observed after each decoding task (two per iteration), transfer functions with identical domain and range are required, since the output parameter of one decoder becomes the input parameter for the following decoder. This is solved for measure M4 by defining the mutual information $I(X; L_{in}) = G(\gamma_{in})$ between the r.v. X and L_{in} distributed with $f_{in}(l|\tilde{x})$, which is a function of γ_{in} only:

$$G_1(\gamma_{in}) = \frac{1}{2} \sum_{\tilde{x} \in \tilde{\mathcal{B}}} \int_{-\infty}^{\infty} \phi_{2\tilde{x}\gamma_{in},4\gamma_{in}}(l) \cdot \log_2 \frac{2\,\phi_{2\tilde{x}\gamma_{in},4\gamma_{in}}(l)}{\phi_{2\gamma_{in},4\gamma_{in}}(l) + \phi_{-2\gamma_{in},4\gamma_{in}}(l)}\,\mathrm{d}\,l,$$

Using $G_1(\gamma_{in})$, a transfer function from $I(X; L_{in}) \in [0,1]$ to $I(X; L_{out}) \in [0,1]$ is defined, where $\gamma_{in} = G_1^{-1}(I(X; L_{in}))$. For measure M5, an input fidelity $\phi_{in} = G_2(\gamma_{in})$, the correlation of estimates obtained from LLRs distributed with $f_{in}(l|\tilde{x})$ with its associated symbol, is defined:

$$G_2(\gamma_{in}) = \frac{1}{2} \sum_{\tilde{x} \in \tilde{\mathcal{B}}} \int_{-\infty}^{\infty} \tilde{x} \cdot \tanh(l/2) \cdot \phi_{2\tilde{x}\gamma_{in},4\gamma_{in}}(l)\,\mathrm{d}\,l.$$

Using $G_2(\gamma_{in})$, a transfer function from $\phi_{in} \in [0,1]$ to $\phi_{out} \in [0,1]$ is defined, where $\gamma_{in} = G_2^{-1}(\phi_{in})$.

In order to compare the six measures, an identical domain and range is desired. Rather than $\mathbb{R}^+$ for the measures M1-M3 and M6, we propose to use the interval $[0,1]$, since here we are able to observe the state of convergence where the decoders decode error-free corresponding to $\gamma_{out} \to \infty$ (M1-M3, M6) and $\phi_{out} = I(X; L_{out}) = 1$ (M4, M5). For the measures M1-M3 and M6 we thus define a transfer function from $G_1(\gamma_{in}) \in [0,1]$ to $G_1(\gamma_{out}) \in [0,1]$ using the mapping $G_1(\cdot)$.

In the sequel, we will specify two transfer functions $\theta_{out} = T_k(\theta_{in})$, $k \in \{\mathrm{I}, \mathrm{II}\}$, with input $\theta_{in} \in [0,1]$ and output $\theta_{out} \in [0,1]$ for decoder I ($k=\mathrm{I}$) and II ($k=\mathrm{II}$) to predict the convergence of iterative algorithms. The transfer function $T_k(\theta_{in})$ is obtained by $\gamma_{in} = G^{-1}(\theta_{in})$, $\gamma_{in} \to \gamma_{out}$, and $\theta_{out} = G_1^{-1}(\gamma_{out})$ using the measures M1-M3 and M6, by $\gamma_{in} = G_1^{-1}(\theta_{in})$, $\gamma_{in} \to I(X; L_{out})$, and $\theta_{out} = I(X; L_{out})$ using measure M4, and by $\gamma_{in} = G_2^{-1}(\theta_{in})$, $\gamma_{in} \to \phi_{out}$, and $\theta_{out} = \phi_{out}$ using measure M5.

IV. Comparison of convergence predictions

We computed $T_{\mathrm{I}}(\theta_{in})$ and $T_{\mathrm{II}}(\theta_{in})$ for the three concatenated systems introduced in Section II. All systems use a rate $R=1/2$, memory 4 outer convolutional code (encoder I) with the generator $G(D) = [1+D+D^4 \;\; 1+D^2+D^3+D^4]$. The interleaver permutation was obtained randomly.

The turbo code system uses a rate 1, memory 1, recursive inner convolutional code (encoder II) with the generator $1/(1+D)$. The interleaver size is $N_{\mathrm{I}} = N_{\mathrm{II}} = 200000$. The $y_n \in \tilde{\mathcal{B}}$ are transmitted over an AWGN channel at 1.4 dB E_b/N_0 given by $E_s/(2R\sigma_w^2) = 1/\sigma_w^2$.

For the turbo equalization system, the x_n mapped to $\tilde{\mathcal{B}}$ are transmitted over a length 5 ISI channel with impulse response $h[n] = 0.227\,\delta[n] + 0.46\,\delta[n-1] + 0.688\,\delta[n-2] + 0.46\,\delta[n-3] + 0.227\,\delta[n-4]$ at 5 dB E_b/N_0 defined by $1/\sigma_w^2$. The interleaver size is $N_{\mathrm{I}} = N_{\mathrm{II}} = 65536$.

For the turbo BiCM system, the x_n are mapped to symbols y_n from the 8-ASK alphabet $\{-7, -5, -3, ..., +5, +7\}$ using an anti-Gray mapping, i.e., the 8 amplitude levels correspond to $\{000, 111, 001, 110, 010, 101, 011, 100\}$ The interleaver size is $N_{\rm I} = 3N_{\rm II} = 200004$. The y_n are transmitted over an AWGN channel at 8.5 dB E_b/N_0 defined by $2/8 \cdot (1^2+3^2+5^2+7^2)/(2R\sigma_w^2) = 21/\sigma_w^2$.

The transfer function $T_{\rm I}(\theta_{in})$ of the outer code was obtained by decoding 10^7 a-priori LLRs $L(c_n)$ per γ_{in} and computing θ_{out} from the output LLRs $L_{\rm I}(c_n)$. The transfer function $T_{\rm II}(\theta_{in})$ of the inner code was obtained by encoding and transmitting 10^7 symbols x_n per γ_{in}. The received z_n and 10^7 a-priori LLRs $L(x_n)$ were decoded to specify θ_{out} using the output LLRs $L_{\rm II}(x_n)$. The c_n and x_n were equally likely 0 and 1.

To obtain system trajectories of the real systems, $\theta_{out}(j,k)$ was computed from the output LLRs $L_{\rm I}(c_n)$ ($k{=}$I) or $L_{\rm II}(x_n)$ ($k{=}$II), respectively, available during the iterative process after the jth, $j=0,1,...,$ iteration. For all measures except M6 the transmitted symbols are required to compute $\theta_{out}(j,k)$ for the system trajectory, which is the sequence $\{\theta_{out}(0,{\rm II}), \theta_{out}(0,{\rm I}), \theta_{out}(1,{\rm II}), \theta_{out}(1,{\rm I}), ...\}$. Since $L_{\rm II}(x_n)$ is input to decoder I and $L_{\rm I}(c_n)$ is input to decoder II (next iteration), we have $\theta_{in}(j,{\rm I}) = \theta_{out}(j,{\rm II})$ and $\theta_{in}(j{+}1,{\rm II}) = \theta_{out}(j,{\rm I})$. Initially, no information about the transmitted symbols is available, i.e., $\theta_{in}(0,{\rm II}) = 0$ for all measures. A measure is accurate when the real transfer characteristic $\theta_{in}(j,k) \rightarrow \theta_{out}(j,k)$, $k \in \{{\rm I},{\rm II}\}$, is well predicted by the transfer function $T_k(\theta_{in}(j,k))$, i.e. $\theta_{out}(j,k) - T_k(\theta_{in}(j,k))$ is small for all j.

Figures 2, 3, 4, show the transfer functions $T_k(\theta_{in})$, $k \in \{{\rm I},{\rm II}\}$, and the system trajectory for the turbo code, the turbo equalization, and the turbo BiCM system. The mutual information measure M4 and the fidelity measure M5 are most accurate in all systems followed by the error rate measure M3. The latter tends to be too pessimistic, since the system trajectory is outside the predicted range. Convergence to low BERs ($\theta_{in}(j,k)$ approaching 1) might still be possible even though the transfer functions predict that no such trajectory exists. The measures M1, M2, and M6 are least accurate and tend to be too optimistic, since the system trajectory is inside the predicted range. We think that this inaccuracy originates from the Gaussian assumption (1), which is applied at the input *and* the output of a SISO decoder. Indeed, the convergence prediction is by far least accurate for the MAP equalizer in Figure 3 and the demapper in Figure 4, which both violate (1). Still, these measures provide useful results for turbo codes and for turbo equalization using linear equalizers [6], where (1) approximately holds.

We note that all measures apply (1) at the input of the SISO decoder but only the measures M1-M3 and M6 at the output. Given the accurate results for the measures M4 and M5, it seems that a SISO decoder, once excited with LLRs corresponding to some $I(X;L_{in})$ or ϕ_{in}, outputs the same $I(X;L_{out})$ or ϕ_{out}, respectively, regardless (to some extent) of the actual distribution of L_{in}. This was observed for measure M4 in [14]. This view also explains the less accurate prediction performance of the measures M1, M2, and M6, where we first force the distribution of the output LLRs to be of type (1) and map the parameter γ_{out} of this distribution to the mutual information $G_1^{-1}(\gamma_{out})$.

Another aspect for comparison is the computational burden to compute the measures. The measures M1, M2, and M6 are obtained easily but they might suffer from clipped LLRs, e.g., in fixed-point arithmetic. The measures M3, M4, and M5 are robust regarding clipping, since output LLRs with small magnitude have the strongest influence on the result of the measurement. The measure M5 shows a good trade-off between accuracy and computational burden. The measure M4 provides additional insight aside convergence prediction, e.g., about the achievable information rates using a particular system [27]. The measure M6 is the only one, which can be applied in a receiver.

V. Conclusions

The comparison shows that all introduced measures help to understand the behavior of iterative algorithms and help to select system parameters to optimize the overall performance despite the apparent assumptions. The aid of a visualized iterative process to its understanding and the still achieved accuracy led to their appearance in the literature. We showed that two measures are fairly accurate for a series of different decoding algorithms, the fidelity measure and the mutual information measure, with the latter one representing a common quantity in communication theory.

References

[1] G. Forney, *Concatenated Codes.* Cambridge, MA, U.S.A.: MIT Press, 1966.

[2] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: turbo codes," in *Proc. IEEE Intern. Conf. on Comm., Geneva*, May 1993.

[3] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Trans. on Information Theory*, pp. 429–445, March 1996.

[4] C. Douillard et al., "Iterative correction of intersymbol interference: turbo equalization," *European Trans. on Telecomm.*, vol. 6, pp. 507–511, Sep-Oct 1995.

[5] G. Bauch and V. Franz, "A comparison of soft-in/soft-out algorithms for 'turbo detection'," in *Proc. Intern. Conf. on Telecomm.*, pp. 259–263, June 1998.

[6] M. Tüchler, R. Koetter, and A. Singer, "Turbo equalization: principles and new results," *submitted to IEEE Trans. on Comm.*, August 2000.

[7] S. ten Brink, J. Speidel, and R. Yan, "Iterative demapping and decoding for multilevel modulation," *Proc. IEEE Globecom Conf.*, pp. 579–584, Nov 1998.

[8] S. Benedetto et al., "Serial concatenated trellis coded modulation with iterative decoding: design and performance," in *Proc. IEEE Global Telecomm. Conf.*, Nov 1997.

[9] D. Divsalar, S. Dolinar, and F. Pollara, "Serial turbo trellis coded modulation with rate-1 inner code," in *Proc. Intern. Symp. on Information Theory*, p. 194, 2000.

[10] X. Wang and H. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. on Comm.*, vol. 47, no. 7, pp. 1046–1061, 1999.

[11] P. Alexander, A. Grant, and M. Reed, "Iterative detection and code-division multiple-access with error control coding," *European Trans. on Telecomm.*, vol. 9, pp. 419–425, Sep-Oct 1998.

[12] D. Divsalar and F. Pollara, "On the design of turbo codes," *TDA Progress Report*, vol. 42, pp. 99–121, Nov 1995.

[13] S. Benedetto, R. Garello, and G. Mondorsi, "A search for good convolutional codes to be used in the construction of turbo codes," *IEEE Trans. on Comm.*, vol. 46, pp. 1101–1105, Sep 1998.

[14] S. ten Brink, "Convergence behaviour of iteratively decoded parallel concatenated codes," *IEEE Trans. on Comm.*, vol. 49, Oct 2001.

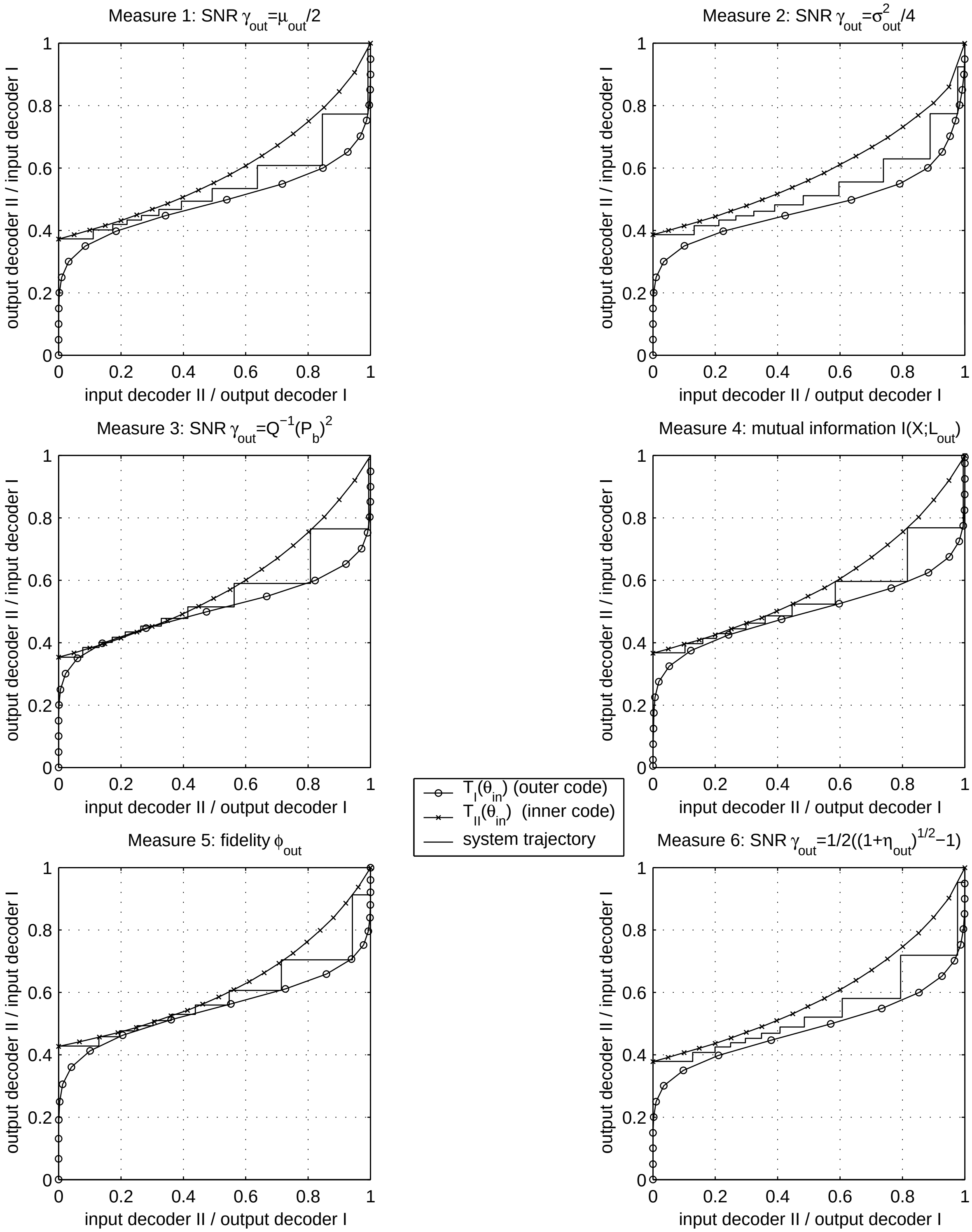


Fig. 2. Measures for tracing decoding convergence of serially concatenated codes (turbo decoding).

[15] T. Richardson, "The geometry of turbo decoding dynamics," *IEEE Trans. on Information Theory*, vol. 46, pp. 9–23, Jan 2000.

[16] D. Agrawal and A. Vardy, "The turbo decoding algorithm and its phase trajectories," *IEEE Trans. on Information Theory*, vol. 47, pp. 699–722, February 2001.

[17] T. Richardson and R. Urbanke, "The capacity of low density parity-check codes under message passing decoding," *IEEE Trans. on Information Theory*, vol. 47, pp. 599–618, Feb 2001.

[18] S. Chung, T. Richardson, and R. Urbanke, "Analysis of sum-product decoding of low-density-parity-check codes using a Gaussian approximation," *IEEE Trans. on Information Theory*, vol. 47, pp. 657–670, Feb 2001.

[19] H. El Gamal and A. Hammons, Jr., "Analyzing the turbo decoder using the Gaussian approximation," *IEEE Trans. on Information Theory*, vol. 47, pp. 671–686, Feb 2001.

[20] D. Divsalar, S. Dolinar, and F. Pollara, "Low complexity turbo-like codes," in *Proc. 2nd Intern. Symp. on Turbo codes, Brest, France*, pp. 73–80, Sep 2000.

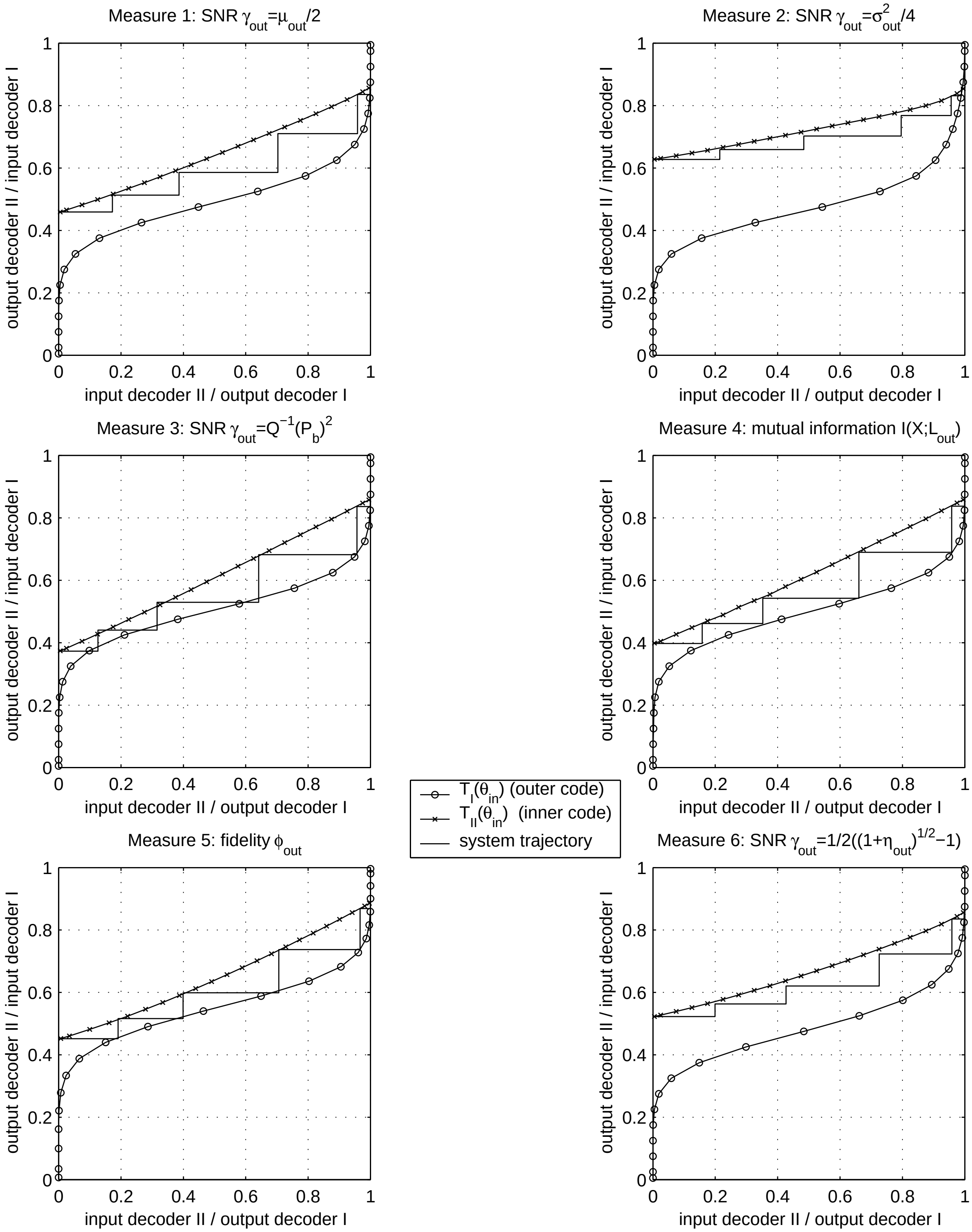


Fig. 3. Measures for tracing decoding convergence of iterative equalization and decoding (turbo equalization).

[21] M. Reed and C. Schlegel, "An iterative receiver for the partial response channel," in *Proc. Intern. Symp. on Information Theory*, p. 63, Aug 1998.

[22] M. Tüchler and J. Hagenauer, "Linear time and frequency domain turbo equalization," in *Proc. 53rd IEEE Vehicular Technology Conf., Rhodes, Greece*, pp. 1449–1453, May 2001.

[23] L.R. Bahl et al., "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. on Information Theory*, vol. 20, pp. 284–287, March 1974.

[24] S. ten Brink, "Code characteristic matching for iterative decoding of serially concatenated codes," *Annals of Telecommunications*, vol. 56, pp. 394–408, April 2001.

[25] K. Narayanan, "Effect of precoding on the convergence of turbo equalization for partial response channels," *IEEE Journal on Sel. Areas in Comm.*, vol. 19, pp. 686–698, April 2001.

[26] P. Robertson, "Illuminating the structure of code and decoder of parallel concatenated resursive systematic (turbo) codes," in *Proc. IEEE Globecom Conf.*, pp. 1298–1303, Dec 1994.

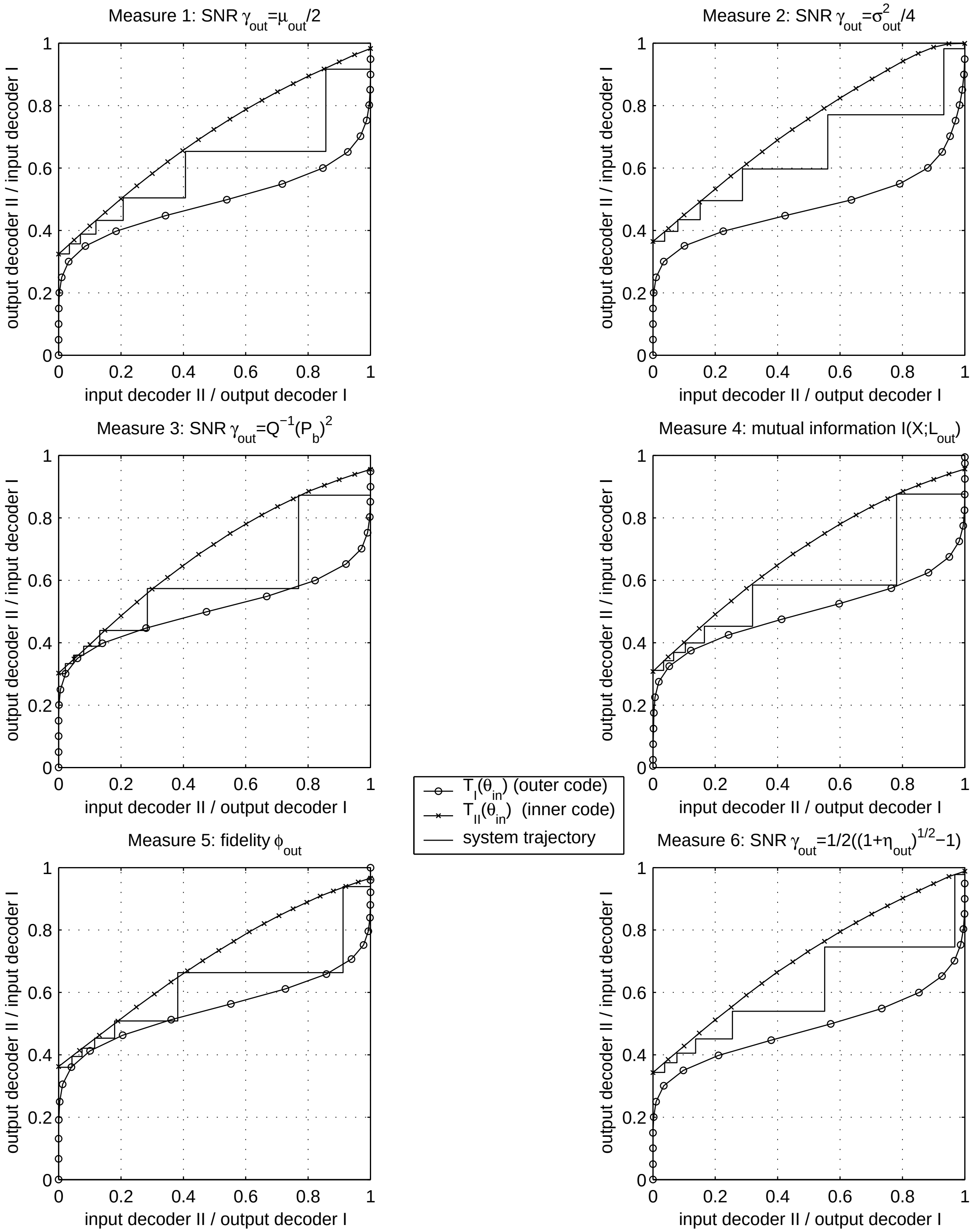


Fig. 4. Measures for tracing decoding convergence of iterative demapping and decoding (turbo BiCM).

[27] S. ten Brink, "Exploiting the chain rule of mutual information for the design of iterative decoding schemes," in *Proc. of the 39th Allerton Conf., Monticello, IL*, Oct 2001.