EUROPEAN TRANSACTIONS ON TELECOMMUNICATIONS
Eur. Trans. Telecomms. 2007; **18**:133–139
Published online 7 August 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/ett.1130

Letter

Information Theory

On extrinsic information of good binary codes operating over Gaussian channels

M. Peleg, A. Sanderovich and S. Shamai (Shitz)*

Department of Electrical Engineering, Technion—Israel Institute of Technology, Technion City, Haifa 32 000, Israel

SUMMARY

We show that the extrinsic information about the coded bits of any good (capacity achieving) binary code operating over a Gaussian channel is zero when the channel capacity is lower than the code rate and unity when capacity exceeds the code rate, that is, the extrinsic information transfer (EXIT) chart is a step function of the signal to noise ratio and independent of the code. It follows that, for a common class of iterative receivers where the error correcting decoder must operate at first iteration at rate above capacity (such as in turbo equalization, iterative channel estimation, parallel and serial concatenated coding and the like), classical good codes which achieve capacity over the Additive White Gaussian Noise Channel are not effective and should be replaced by different new ones. Copyright © 2006 AEIT

1. INTRODUCTION

In this letter we derive the extrinsic information transfer (EXIT) chart of asymptotically long binary codes which achieve a vanishing probability of error over the Additive White Gaussian Noise (AWGN) channel at code rates below the channel capacity. We denote such codes as 'good codes' in the following. The results provide an insight about corresponding iterative receivers designed to approach the channel capacity assuming asymptotically long codewords.

It is well known that when a good error correcting code (ECC) is used to transmit information over a channel the capacity of which is lower than the code rate, then the error rate is high. This scenario actually occurs at the first decoding iteration performed by the new iterative receivers based on the turbo principle, see Figure 1, where some preprocessor such as equalizer, multi-user detector [1], phase estimator or other precedes the decoder and employs the decoder outputs to improve the channel presented to the decoder over the successive iterations. The ECC code may be any code, including an iteratively decodable turbo or LDPC code.

If the whole iterative receiver is designed to approach capacity and if the iterative feedback to the preprocessor is really required then at the first iteration the ECC decoder is presented with a channel the capacity of which is below the code rate while at the next iterations the preprocessor will improve the ECC decoder input and finally enable errorless decoding. To achieve this, the ECC decoder must pass some useful information to the preprocessor at the first iteration, while operating over channel the capacity of which is below the code rate. The relevant information to be passed is the well known extrinsic information (EI) to be defined below.

In fact, serially concatenated turbo codes can also be represented by the structure of Figure 1 where the preprocessor is the decoder of the inner component code. Parallel concatenated turbo codes are decoded by a similar structure and the operation at rate above channel capacity at first iteration is then also clearly required since the component code is

* Correspondence to: S. Shamai (Shitz), Department of Electrical Engineering, Technion—Israel Institute of Technology, Technion City, Haifa 32 000, Israel. E-mail: sshlomo@ee.technion.ac.il

Contract/grant sponsors: REMON Consortium; The EU6th International Framework Programme.

Received 14 April 2005
Revised 14 March 2006
Accepted 10 April 2006

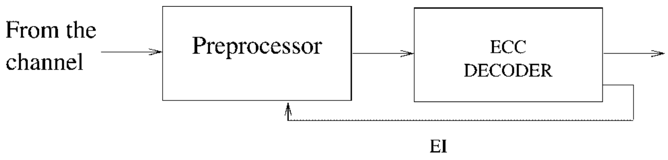

Figure 1. Iterative receiver.

presented only with a subset of the channel output symbols. In the following derivation we will show formally that over AWGN channels, codes considered to be good in the classical sense, provide no EI at all in this setting and thus eliminate any improvement by the iterative feedback in Figure 1. So the search for different codes fitting the new iterative systems, such as performed in References [1–3] and others, is indeed essential if the iterative receiver is to perform better than, say, separate equalisation and decoding. Also, classical good codes cannot perform well as outer codes in a serially concatenated turbo code or component codes in parallel concatenated turbo codes. This was indicated first in Reference [4] where increasing the constraint length of a convolutional component code rendered the iterative feedback of a turbo decoder ineffective. More precisely, we will show that the extrinsic information transfer (EXIT) chart of any good binary code over a memoryless AWGN channel is a step function, the EI being zero for signal to noise ratio (SNR) at which the channel capacity is below the code rate and unity at SNR larger than this. Notations: Mutual information is denoted by I, entropy by H and statistical expectation by E. Probability and probability density functions are denoted by P and p respectively and bold-faced letters denote vectors.

2. MODELS AND DEFINITIONS

2.1. Channel model

We examine a binary input additive white Gaussian noise (BI-AWGN) channel

$$y = \sqrt{s}(2x - 1) + \eta \qquad (1)$$

where $(2x - 1)$ is the transmitted signal, -1 or 1, x is the corresponding bit at the channel input, 0 or 1 and η is a Gaussian random variable with zero mean and unit variance. The signal power, denoted s, is equal to the SNR. The channel is characterised by the Gaussian probability $p(y|x)$ and C denotes the channel capacity. For some $s^{\mathrm{l}} < s^{\mathrm{h}}$, a channel characterised by $s = s^{\mathrm{l}}$, can be described as physically degraded with respect to a channel characterised by s^{h}, where the superscripts $^{\mathrm{h}}$ and $^{\mathrm{l}}$ denote 'higher' and 'lower' SNR respectively. The outputs of the two channels are denoted y^{l} and y^{h}. By physically degraded we mean that x_i, y^{l} and y^{h} form the following Markov chain:

$$x_i - y_i^{\mathrm{h}} - y_i^{\mathrm{l}} \qquad (2)$$

arising from the possibility to obtain y^{l} from y^{h} by adding independent Gaussian noise and scaling to conform to Equation (1).

2.2. The good code

We desire to transmit information $\mathbf{U}$. We do so in the standard manner of transmitting a codeword $\mathbf{x}$, a vector of n channel symbols x_i belonging to an asymptotically long good code X of rate R. The selection of the transmitted codewords is determined by $\mathbf{U}$ and is equi-probable. The received vector is denoted $\mathbf{y}$. The code is such that it achieves vanishing error probability at channel SNR $s = s^0$ for which $C(s^0) - \varepsilon' < R < C(s^0)$ where ε' is positive and can be made as small as desired by increasing n. A well known result is then

$$\frac{1}{n} I\Big[\mathbf{x}; \mathbf{y}(s^0)\Big] \geq C(s^0) - \varepsilon' \qquad (3)$$

2.3. The extrinsic information

We are interested in a symbol x_i, which is a symbol at the i'th position in $\mathbf{x}$. We define $\mathbf{x}_i'$ as $\mathbf{x}$ with x_i excluded and correspondingly $\mathbf{y}_i'$ as $\mathbf{y}$ with y_i excluded. We denote z_i the complete information obtainable from $\mathbf{y}_i'$ about x_i, known as the extrinsic information, see for example Reference [4]. The extrinsic information z_i can be expressed for example as the logarithmic likelihood ratio (LLR) of x_i or as: $P(x_i = 1|\mathbf{y}_i')$. The average extrinsic information measure is then defined as

$$\mathrm{EI} = \frac{1}{n} \sum_{i=1}^{n} I(x_i; \mathbf{y}_i') = \frac{1}{n} \sum_{i=1}^{n} I(x_i; z_i) \qquad (4)$$

When x_i is given, then y_i is clearly independent of $\mathbf{y}_i'$, that is $p(y_i|x_i, \mathbf{y}_i') = p(y_i|x_i)$ and $p(\mathbf{y}_i'|x_i, y_i) = p(\mathbf{y}_i'|x_i)$. This extends the Markov chain in Equation (2) to

$$\mathbf{y}_i'^{\mathrm{h}} - x_i - y_i^{\mathrm{h}} - y_i^{\mathrm{l}} \qquad (5)$$

where $\mathbf{y}_i'^{\mathrm{h}}$ depends on x_i only due to the dependence of $\mathbf{x}'_i$ on x_i induced by the code. Using the physical degraded channel property in Equation (2) this can be extended to

$$\mathbf{y'}_i^{\text{l}} - \mathbf{y'}_i^{\text{h}} - x_i - y_i^{\text{h}} - y_i^{\text{l}} \tag{6}$$

Furthermore, due to this Markov chain and the data processing theorem we have $I(\mathbf{y'}_i^{\text{h}}; x_i) \geq I(\mathbf{y'}_i^{\text{l}}; x_i)$, thus EI is a non-decreasing function of s:

$$\text{EI}(s^{\text{h}}) \geq \text{EI}(s^{\text{l}}) \tag{7}$$

3. EXIT CHART OF GOOD CODES

When the capacity C is strictly above the code rate R, we have perfect decoding for asymptotically long good codes, even if the single symbol y_i is removed (erased) before the decoding. Thus, we have for any small positive ε' and large n

$$R < C - \varepsilon' \rightarrow \text{EI} = 1 \tag{8}$$

This intuitive attribute of good codes is verified in Appendix A for finite s. The central result of this letter is the following proposition and its method of proof:

Proposition 3.1. *The average* EI, *Equation* (4), *about the coded bits x_i of a good binary code operating over a BI-AWGN channel, the capacity of which is lower than the code rate, is zero.* (*More precisely smaller than any positive ε_2 for n large enough.*)

Proof. It is well known that, when the code rate R is at or above capacity, good codes mimic closely the channel output statistics of a capacity achieving identically and independently distributed (i.i.d.) input [5, Theorem 15]. Specifically, we prove in Appendix B that for all $s < s^0$, the mutual information $I(\mathbf{x}; \mathbf{y})$ over the channel with the good code is similar to the symbol wise mutual information. That is for any small $\varepsilon' > 0$ and sufficiently large n:

$$0 \leq \frac{1}{n}\sum_{i=1}^{n} I\left[x_i; y_i(s)\right] - \frac{1}{n} I\left[\mathbf{x}; \mathbf{y}(s)\right] = \gamma(s) \leq \varepsilon' \triangleq \varepsilon^3 \tag{9}$$

where $\gamma(s)$ is a non-decreasing function of s. The substitution $\varepsilon' = \varepsilon^3$ will be required below.

We shall need to upper-bound the derivative of Equation (9) with respect to s

$$\text{ID}(s) \triangleq \frac{\mathrm{d}}{\mathrm{d}s}\left\{\frac{1}{n}\sum_{i=1}^{n} I[x_i; y_i(s)] - \frac{1}{n} I[\mathbf{x}; \mathbf{y}(s)]\right\} = \frac{\mathrm{d}}{\mathrm{d}s}\gamma(s) \tag{10}$$

Since γ is non-decreasing, ID is non-negative and its average over an interval of $s^0 - \Delta$ to s^0 cannot exceed ε^3/Δ, otherwise its integral (9) would exceed ε^3. We shall choose $\Delta = \varepsilon$ to limit the average ID to ε^2. Thus, there is some $s = s_t$ in the above interval, ε within s^0, for which ID is bounded by

$$0 \leq \text{ID}(s_t) \leq \varepsilon^2 \tag{11}$$

Due to Equation (7), vanishing EI at s_t implies vanishing EI at all smaller values of s, so it is sufficient to prove vanishing EI at s_t.

In Reference [6] the minimum mean square error (MMSE), when estimating a general input of a Gaussian channel using the channel output, is linked to the derivative with respect to the SNR of the relevant mutual information. This property is useful in analysis of iterative receivers, see for example Reference [7]. Using Reference [6, Theorem 2] while the transmitted signal $\mathbf{Hx}$ in Reference [6] is our $(2\mathbf{x} - 1)$, see Equation (1), and the constant 1 does not influence the estimation errors, we can see that the sum of the estimation errors of all the symbols x_i is

$$\sum_{i=1}^{n} \text{MMSE}(x_i|\mathbf{y}) = \frac{\mathrm{d}}{\mathrm{d}s} 0.5\, I\left[\mathbf{x}; \mathbf{y}(s)\right] \tag{12}$$

where $\text{MMSE}(x_i|\mathbf{y})$ denotes the MMSE of an individual bit x_i obtained by optimally estimating x_i from $\mathbf{y}$. The factor before the sum in Equation (12) is 0.5 rather than 2 in Reference [6] because the transmitted signal equation (1), is $2x - 1$ rather than x. However, for any x_i, a similar estimation error can be achieved using merely the single received symbol y_i, see Reference [6, Eq. (1)]:

$$\sum_{i=1}^{n} \text{MMSE}(x_i|y_i) = \frac{\mathrm{d}}{\mathrm{d}s} 0.5 \sum_{i=1}^{n} I[x_i; y_i(s)] \tag{13}$$

By Equations (10)–(13) we have then

$$0 \leq \frac{1}{n}\sum_{i=1}^{n} \text{MMSE}(x_i|y_i) - \frac{1}{n}\sum_{i=1}^{n} \text{MMSE}(x_i|\mathbf{y}) \leq 0.5\varepsilon^2$$

$$0 \leq \frac{1}{n}\sum_{i=1}^{n} [\text{MMSE}(x_i|y_i) - \text{MMSE}(x_i|\mathbf{y})] \leq 0.5\varepsilon^2 \tag{14}$$

Clearly each element of the above sum is positive and their average is upper bounded by $0.5\varepsilon^2$. We need to upper bound those elements. The following rather loose but simple bound is sufficient for our purpose. At most $0.5n\varepsilon$ elements may be larger than ε otherwise Equation (14) would be violated. This vanishing proportion of elements can contribute only 0.5ε bits to the average EI Equation (4) because the EI for each bit is bounded by 1, so we can disregard them in our proof of vanishing average EI , Equation (4), and use

$$\text{MMSE}(x_i|y_i) - \text{MMSE}(x_i|\mathbf{y}) \leq \varepsilon \quad (15)$$

Thus, at $s = s_t$, the MMSE estimation error of x_i using $\mathbf{y}$ is nearly the same as if only y_i was used. The MMSE estimate of x_i, valued 0 or 1, is its conditional expectation

$$\hat{x}_i(\mathbf{y}'_i, y_i) = 0 \times P(x_i = 0|\mathbf{y}'_i, y_i) + 1 \times P(x_i = 1|\mathbf{y}'_i, y_i)$$
$$\hat{x}_i(\mathbf{y}'_i, y_i) = P(x_i = 1|\mathbf{y}'_i, y_i) \quad (16)$$

and $\hat{x}_i(y_i) = P(x_i = 1|y_i)$

Equations (15) and (16) imply that y_i is an approximate sufficient statistics, where the full statistics is $\mathbf{y} = (y_i, \mathbf{y}'_i)$. This immediately implies Equation (17) due to continuity as shown in more detail in Appendix C:

$$E[P(x_i = 1|\mathbf{y}'_i, y_i) - P(x_i = 1|y_i)]^2 \leq \varepsilon \quad (17)$$

Thus $P(x_i = 1|\mathbf{y}'_i, y_i) \cong P(x_i = 1|y_i)$, showing that $\mathbf{y}'_i$ cannot provide additional information about x_i when y_i is known.

To establish (for any small positive ε_2)

$$\text{EI} = \frac{1}{n}\sum_{i=1}^{n} I(x_i; \mathbf{y}'_i) \leq \varepsilon_2 \quad (18)$$

we have to show that $\mathbf{y}'_i$ cannot provide information about x_i also when y_i is not known. This is equivalent to $P(x_i = 1|\mathbf{y}'_i) = P(x_i = 1)$. We present in the following the principles leading to Equation (18), while a detailed but tedious proof is presented in Reference [8, Appendix D].

Since $P(x_i = 1|\mathbf{y}'_i, y_i)$ and $P(x_i = 1|y_i)$ determine the log likelihood ratios (LLR) which are closely coupled to mutual information, Equation (17) leads to

$$I(x_i; \mathbf{y}'_i, y_i) - I(x_i; y_i) < \varepsilon_1 \quad (19)$$

for some ε_1 proportional to ε. Next we shall use the bounds on information obtained by combining the outputs of independent channels [9]. The channel outputs y_i and $\mathbf{y}'_i$ can be considered independent sources of information about x_i by Equation (5). Thus, y_i and $\mathbf{y}'_i$ are outputs of parallel broadcast channels in the sense of Reference [9]. Also the channel (1) x_i to y_i is symmetric and x_i is nearly uniformly distributed over -1 and 1 for all but a vanishing proportion of symbols x_i (otherwise the code cannot achieve capacity, see Reference [10], and further references therein) and those can be disregarded since they cannot influence the average EI Equation (4) because the contribution of one symbol to the average EI is limited to $1/n$.

The bounds [9] hold for x distributed uniformly according to $P(x = 1) = 0.5$, however since the bounds based on mutual information combining are all continuous functions of the parameter $P(x = 1)$ describing the distribution of x, small deviations from $P(x = 1) = 0.5$ inflict small deviation on the outputs. Under those conditions it follows from Reference [9, Theorem 2] that when $I(y_i; x_i)$ and $I(\mathbf{y}'_i; x_i)$ are given, then $I(\mathbf{y}'_i, y_i; x_i)$, the information about x_i obtained by combining both y_i and $\mathbf{y}'_i$, is lower bounded by the one obtained when replacing $\mathbf{y}'_i$ by the output B of a binary symmetric channel (BSC) transmitting x_i, with $I(B; x_i) = I(\mathbf{y}'_i; x_i)$. Straightforward calculation reveals that for the Gaussian channel with input x_i and output y_i and BSC with output B we have $I(B, y_i; x_i) > I(y_i; x_i) + \alpha I(B; x_i)$ for some positive α. Thus $I(\mathbf{y}'_i, y_i; x_i) > I(y_i; x_i) + \alpha I(\mathbf{y}'_i; x_i)$. This together with Equation (19) implies

$$I(x_i; \mathbf{y}'_i) \leq \varepsilon_2 \quad (20)$$

for any small positive ε_2. This establishes Equation (18) and proposition 3.1; a detailed proof is available in Reference [8]. ■

By the data processing theorem for the Markov chain Equation (5) we have $I(x_i; \mathbf{y}'_i) \geq I(y_i; \mathbf{y}'_i)$, thus

$$\frac{1}{n}\sum_{i=1}^{n} I(y_i; \mathbf{y}'_i) \leq \varepsilon_2 \quad (21)$$

when channel capacity is below the code rate. This attribute cannot be derived directly from the results of Reference [5] and does not hold when capacity exceeds slightly the code rate and successful decoding occurs.

Remarks. When capacity exceeds the code rate there is a sharp transition of the EI since $I(\mathbf{x}, \mathbf{y})$ is not determined by Equation (9) any more but reaches a plateau at the code rate and the MMSE, which is proportional to the derivative of $I(\mathbf{x}, \mathbf{y})$ with respect to $s = \text{SNR}$ by Equation (12), goes to zero abruptly. This transition takes place over a small region of s for which the difference Equation (9) is very small but, significantly, not zero. A similar transition of

$$\frac{1}{n}\sum_{i=1}^{n} I(y_i; \mathbf{y}'),$$

see Equation (21), occurs in the same region.

Clearly, the area A under the EI versus $I(x_i; y_i)$ curve,

$$A = \int_0^1 \mathrm{EI}\, d[I(x_i; y_i)],$$

equals $1 - R$, thus the step function EXIT chart derived here for good binary code over the AWGN channel conforms to the 'EXIT chart area property' of outer codes which was proved in Reference [11] for any code over the binary erasure (BEC) channel. This property over the BEC channel together with Equation (8), which is easy to verify also for the BEC channel, implies that Equation (18) holds for good codes over the BEC channel too.

4. CONCLUSIONS

The EXIT chart of any good binary code operating over a Gaussian channel is a step function, zero at channel capacities below the code rate and unity at capacities above the code rate. Thus codes which are good over the AWGN channels are very inefficient when used in an iterative receiver of the type presented in Figure 1 which includes turbo-equalization, iterative multi-user receivers and serially concatenated codes as special cases. Interestingly, the step function EXIT chart derived here for the AWGN channel conforms to the EXIT chart area property derived in Reference [11] for the erasure channel. Furthermore, good code operating at rate above channel capacity falls apart into its individual transmitted symbols in the sense that all the information about a coded bit x_i is contained in the corresponding received symbol y_i and no information about x_i can be inferred from the other received symbols, neither alone, see Equation (18) and neither as supplement to y_i, see Equation (17).

It is of interest if the main result of this letter, namely Equation (18), holds for more general memoryless channels. Based on [11], the results hold for the BEC channel as explained above, Reference [8] outlines an extension to a wider class of M-ary input memoryless channels using the concept of GEXIT [12], and the arguments presented in Reference [1] suggest extension to any memoryless channel for binary random codes.

APPENDIX

This appendix verifies Equation (8) for finite s. When $R < C - \varepsilon'$, the symbol x_i is decoded with zero error probability. Furthermore, by the Markov chain Equation (5)

$$P(\mathbf{y}'_i, y_i|x_i) = P(\mathbf{y}'_i|x_i)P(y_i|x_i)$$

Thus

$$P(x_i|\mathbf{y}'_i, y_i) = P(\mathbf{y}'_i|x_i)P(y_i|x_i)P(x_i)\,\frac{1}{P(\mathbf{y}'_i, y_i)} \quad \text{(A.1)}$$

Let us denote the actually transmitted x_i by x_t. Perfect decoding of x_i implies $P(x_i|\mathbf{y}'_i, y_i) = 0$ for $x_i \neq x_t$ for all y_i and $\mathbf{y}'_i$ possible when $x_i = x_t$. For the channel (1) all y_i have non-zero probability for both possible x_t, that is $P(y_i|x_i) > 0$. Then since any of the terms on the right hand side of Equation (A.1) except of the first one is not zero for all possible $\mathbf{y}'_i$ and x_i, the first term must be $P(\mathbf{y}'_i|x_i) = 0$, for $x_i \neq x_t$, which ensures perfect decoding of x_i from $\mathbf{y}'_i$, implying Equation (8).

APPENDIX

Proof of Equation (9). In this appendix we shall use two types channel inputs. One of them will be a codeword $\mathbf{x}$ chosen randomly and uniformly from the good code (GC) X approaching capacity within ε' at channel SNR s^0. All the properties related to this input will be denoted by the superscript $^{\mathrm{GC}}$, such as I^{GC}. The other type of input will be a vector $\mathbf{x}$ of symbols x_i chosen independently and according to the symbol-vise distribution of our good code X which may be dependent to a certain extent on the symbol index i. We denote this input distribution by the superscript $^{\mathrm{IND}}$ and the corresponding mutual information as $I^{\mathrm{IND}}(\mathbf{x}; \mathbf{y})$. The symbol-wise mutual information $I(x_i; y_i)$ is identical for both the distributions for each i. Thus, the first term in Equation (9) equals $\frac{1}{n}\, I^{\mathrm{IND}}(\mathbf{x}; \mathbf{y})$ and the second equals $\frac{1}{n}\, I^{\mathrm{GC}}(\mathbf{x}; \mathbf{y})$.

In the rest of this appendix we shall denote by y^0 the output of a channel with SNR equal to s^0 and by y^{l} and y^{h} the output of a channel parameterised by some $s^{\mathrm{l}} < s^{\mathrm{h}} < s^0$. For both the $\mathbf{x}^{\mathrm{GC}}$ and $\mathbf{x}^{\mathrm{IND}}$ types of channel inputs

$$I(\mathbf{x}; \mathbf{y}) = H(\mathbf{y}) - H(\mathbf{y}|\mathbf{x}) \quad \text{(B.1)}$$

Since $H(\mathbf{y}|\mathbf{x})$ is invariant with respect to the type of channel input ($^{\mathrm{GC}}$ or $^{\mathrm{IND}}$) over the memoryless channel, the difference $\mathrm{DI} = I^{\mathrm{IND}}(\mathbf{x}; \mathbf{y}) - I^{\mathrm{GC}}(\mathbf{x}; \mathbf{y})$ is determined wholly by

$$\mathrm{DI} = H^{\mathrm{IND}}(\mathbf{y}) - H^{\mathrm{GC}}(\mathbf{y}) \quad \text{(B.2)}$$

By the chain rule of entropy we have for both the types of channel inputs:

$$H(\mathbf{y}^{\mathrm{h}}, \mathbf{y}^{\mathrm{l}}) = H(\mathbf{y}^{\mathrm{l}}) + H(\mathbf{y}^{\mathrm{h}}|\mathbf{y}^{\mathrm{l}}) = H(\mathbf{y}^{\mathrm{h}}) + H(\mathbf{y}^{\mathrm{l}}|\mathbf{y}^{\mathrm{h}})$$

$$H(\mathbf{y}^{\rm l}) = H(\mathbf{y}^{\rm h}) + H(\mathbf{y}^{\rm l}|\mathbf{y}^{\rm h}) - H(\mathbf{y}^{\rm h}|\mathbf{y}^{\rm l}) \quad \text{(B.3)}$$

Let us compare $H(\mathbf{y}^{\rm l})$ for the two types of channel inputs taking into account that $H(\mathbf{y}^{\rm l}|\mathbf{y}^{\rm h})$ does not depend on the channel input type due to the Markov, Equation (2), and the memoryless properties of the channel:

$$H(\mathbf{y}^{\rm l})^{\rm IND} - H(\mathbf{y}^{\rm l})^{\rm GC} = H(\mathbf{y}^{\rm h})^{\rm IND} - H(\mathbf{y}^{\rm h})^{\rm GC} - H(\mathbf{y}^{\rm h}|\mathbf{y}^{\rm l})^{\rm IND} + H(\mathbf{y}^{\rm h}|\mathbf{y}^{\rm l})^{\rm GC} \quad \text{(B.4)}$$

Let us define $\beta = H(\mathbf{y}^{\rm h}|\mathbf{y}^{\rm l})^{\rm IND} - H(\mathbf{y}^{\rm h}|\mathbf{y}^{\rm l})^{\rm GC}$. Then from Equations (B.2) and (B.4)

$$I(\mathbf{x};\mathbf{y}^{\rm l})^{\rm IND} - I(\mathbf{x};\mathbf{y}^{\rm l})^{\rm GC} = I(\mathbf{x};\mathbf{y}^{\rm h})^{\rm IND} - I(\mathbf{x};\mathbf{y}^{\rm h})^{\rm GC} - \beta \quad \text{(B.5)}$$

The difference β is positive or zero since the $^{\rm IND}$ and the $^{\rm GC}$ distributions induce the same symbol-wise distributions $p(y_i^{\rm l}, y_i^{\rm h})$ while only the $^{\rm GC}$ induces dependence between different symbols. So $\mathrm{DI} = I(\mathbf{x};\mathbf{y})^{\rm IND} - I(\mathbf{x};\mathbf{y})^{\rm GC}$ is a non-decreasing function of s. Furthermore, DI is always positive or zero since the symbol-wise distributions of the two input types are identical while only the $^{\rm GC}$ input induces dependency between the input symbols. Thus $\mathrm{DI} = I(\mathbf{x};\mathbf{y})^{\rm IND} - I(\mathbf{x};\mathbf{y})^{\rm GC}$ is a positive non-decreasing function of s. At $s = s^0$, DI is small as desired since $\frac{1}{n} I(\mathbf{x};\mathbf{y})^{\rm GC}$ approaches capacity within ε' at channel SNR of s^0 while $I(\mathbf{x};\mathbf{y})^{\rm IND}$ cannot exceed it. This proves Equation (9) including $\gamma(s)$ being non-decreasing function of s.

APPENDIX

Proof of Equation (17). Denote the MMSE estimators Equation (16) of x_i by

$$A = P(x_i = 1|y_i) = \hat{x}_i(y_i)$$
$$B = P(x_i = 1|\mathbf{y}_i', y_i) = \hat{x}_i(\mathbf{y}_i', y_i) \quad \text{(C.1)}$$

Then, by Equation (15):

$$\begin{aligned} \varepsilon &\geq E[(x-A)^2 - (x-B)^2] \\ &= E[A^2 - B^2 - 2x(A-B)] \\ &= E[(A-B)(A+B) - 2x(A-B)] \\ &= E[(A-B)(A-x+B-x)] \\ \varepsilon &\geq E[(A-B)(A-x+B-x)] \end{aligned} \quad \text{(C.2)}$$

It is well known that the error of an MMSE estimator is not correlated neither to the estimate itself and neither to any function of the information which was used to form the estimate. Now, since B is the MMSE estimate using the full information $\mathbf{y}$ which can also produce A, we have $0 = E[A(B-x)] = E[B(B-x)]$. Then $2(B-x)$ can be subtracted from the term inside the right parenthesis in Equation (C.2) yielding

$$\begin{aligned} \varepsilon &\geq E[(A-B)(A-x-(B-x))] \\ \varepsilon &\geq E[(A-B)(A-B)] = E(A-B)^2 \end{aligned} \quad \text{(C.3)}$$

This, together with the definitions Equation (C.1) implies Equation (17).

ACKNOWLEDGEMENT

The authors thank Dr Yossef Steinberg for his helpful suggestions. We also thank the anonymous reviewer for a concise and helpful review. The work was supported the EU6th International Framework Programme via the NEWCOM Network of Excellence.

REFERENCES

1. Sanderovich A, Peleg M, Shamai (Shitz) S. LDPC coded MIMO multiple access with iterative joint decoding. *IEEE Transactins on Information Theory* 2005; **51**(4):1437–1450.
2. ten Brink S. Designing iterative decoding schemes with the extrinsic information transfer chart. *AEU International Journal of Electronics and Communications* 2000; **54**(6):389–398.
3. Peleg M, Shamai (Shitz) S. Efficient communication over the discrete-time memoryless Rayleigh fading channel with turbo coding/decoding. *European Transactions on Telecommunications (ETT)* 2000; **11**:475–485.
4. Berrou C, Glavieux A, Thitimajshima P. Near Shannon limit error-correcting coding and decoding. In *Proceedings of International Communications Conference (ICC'93)*, May 1993, pp. 1064–1070, Geneva, Switzerland.
5. Han TS, Verdú S. Approximation theory of output statistics. *IEEE Transactions on Information Theory* 1993; **39**(3):752–772.
6. Guo D, Shamai (Shitz) S, Verdú S. Mutual information and minimum mean-square error in Gaussian channel. *IEEE Transactions on Information Theory* 2005; **51**(4):1261–1282.
7. Bhattad K, Narayanan K. An MSE based tranfer chart to analyze iterative decoding schemes. In *42nd Allerton Conference on Communication, Control and Computing*, October 2004, pp. 1724–1732, Monticello, IL.
8. Peleg M, Sanderovich A, Shamai (Shitz) S. On extrinsic information of good codes operating over discrete memoryless channels. web accessible via arXive, http://arxiv.org/abs/cs.IT/0504028 and http://arxiv.org/ftp/cs/papers/0504/0504028.pdf.
9. Sutskover I, Shamai (Shitz) S, Ziv J. Extremes of information combining. *IEEE Transactions on Information Theory* 2005; **51**(4):1313–1325.
10. Shamai (Shitz) S, Verdú S. The empirical distribution of good codes. *IEEE Transactions on Information Theory* 1997; **43**(3):836–846.

11. Ashikhmin A, Kramer A, ten Brink S. Code rate and the area under extrinsic information transfer curves. In *Proceedings of the International Symposium on Information Theory (IEEE)*, 2002, p. 115, IEEE.
12. Meassone C, Urbanke R, Montanari A, Richardson T. Life above threshold: From list decoding to area theorem and MSE. In *IEEE Information Theory Workshop (ITW)*, October 24–29 2004, Texas.

