On Extrinsic Information of Good Codes Operating Over Memoryless Channels with Incremental Noisiness

M. Peleg, *Member, IEEE*, A. Sanderovich, *Student Member, IEEE,* and S. Shamai (Shitz), *Fellow, IEEE*

Abstract: We show that the Extrinsic Information about the coded bits of any *good* (capacity achieving) *code* operating over a binary symmetric channel is zero when the code operates at code rate above capacity and a positive constant otherwise, that is, the Extrinsic Information Transfer (EXIT) chart is a step function of channel quality, for any capacity achieving code. This extends similar previous results proved over the erasure channel and over the AWGN channel, indicating the step function property of the EXIT chart is rather general. It follows that, for a common class of iterative receivers where the error correcting decoder must operate at the first iterations at rate above capacity (such as in turbo equalization, turbo channel estimation, parallel and serial concatenated coding and the like), classical *good codes* which achieve capacity over the relevant channels are not effective and should be replaced by different new ones.

I INTRODUCTION

We examine a certain property of *good codes,* denoting here capacity achieving asymptotically long codes. A codeword comprising transmitted symbols x_i is transmitted over a memoryless channel and the Extrinsic Information (EI) which can be extracted about a particular x_i from the received symbols y_j, excluding the symbol y_i corresponding to x_i, is examined as a function of the channel quality. As shown in [1], [2] and in this work, the average EI about the coded bits of any *good code* operating over a class of Discrete Input Memoryless Channels (DIMC) is zero when the code operates at a code rate R above the capacity C and a positive constant otherwise. That is, the Extrinsic Information Transfer (EXIT) chart is a step function of the channel quality. Consequently any *good code* operating at a rate above channel capacity falls apart into its individual transmitted symbols in the sense that all the information about a coded transmitted symbol is contained in the corresponding received symbol and no information about it can be inferred from the other received symbols.

Let us consider a common class of iterative receivers utilizing preprocessing inside the iterative loop, such as turbo equalization, turbo channel estimation, serial concatenated coding and the like, see Fig. 1.

The authors are with the Department of Electrical Engineering, Technion–Israel Institute of Technology, 32000 Haifa, Israel. M. Peleg is also with Rafael ltd.
The work was supported by the US-Israel Binational Science Foundation (BSF).

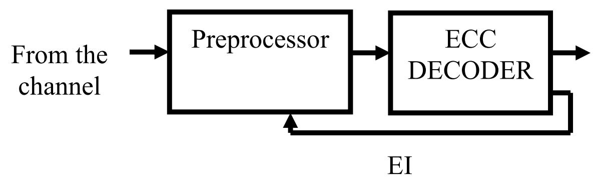

Fig.1: Iterative receiver with a preprocessor inside the decoding loop.

In such a setting the error correcting decoder must operate at the first iteration at a rate above capacity. Clearly, if the iterative feedback to the preprocessor is indeed required for reliable decoding, then classical *good codes* which achieve capacity over the DIMC are not effective here due to their failure to provide any extrinsic information at the first iteration, and so should be replaced by different codes designed specifically for such a system. In fact, serially concatenated turbo codes can also be represented by the structure of Fig.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 presented only with a subset of the channel output symbols. So the search for codes fitting the new iterative systems such as performed in [3], [4], [5] and others is indeed essential if the iterative receiver is to perform better than, say, separate equalization 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 [6] where increasing the constraint length of a convolution component code rendered the iterative feedback of a turbo decoder ineffective.

Interestingly, the proofs of the above step function attribute of EXIT charts over various channels are based on diverse area properties. In the case of the Binary Erasure Channel (BEC), this attribute of *good codes* is a direct consequence of the constant area property of the EXIT chart over erasure channels [7]. A similar result for the Binary Input Additive White Gaussian Noise (AWGN) channel was proved in [1], [2] using the recently discovered link between mutual information and Minimum Mean Square Error (MMSE) [8] and the corresponding constant area under the MMSE versus the Signal to Noise Ratio (SNR) curve.

Interesting consequences of the last constant area property are demonstrated in Fig. 2. The MMSE is that of the transmitted codeword **x** estimated from the received codeword **y** [8] and it is a measure of the reliability of decoding. The black curve is that of the *good code*. It is identical for all sub-threshold SNRs (meaning SNRs for which C<R) to that of an uncoded transmission (the blue line), an attribute closely related to the step function property of the EXIT chart [1]. The red line is that of a repetition code, which compensates by suboptimal performance at SNRs above the capacity threshold for the MMSE reduced relatively to the good code at sub-threshold SNRs. The green line is that of a state of art LDPC code from [9] exhibiting performance approaching that of the *good code*; it is approximated by a tight upper bound produced by the belief propagation showed suboptimal in [10]. This is another demonstration of the tradeoff between performance of a code below and above the threshold SNR.

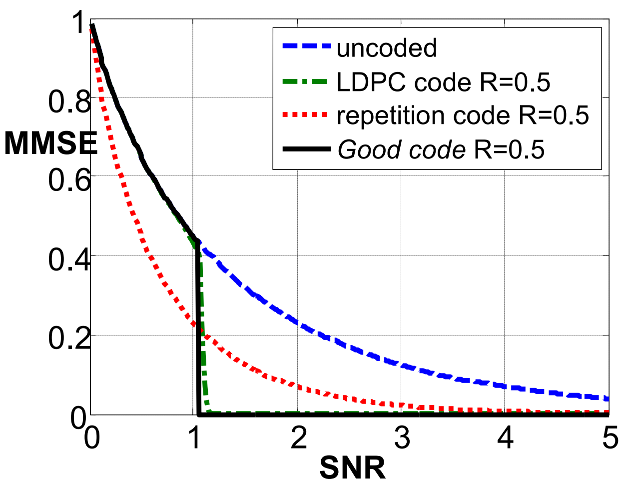

Fig.2: MMSE as a function of SNR for uncoded transmission, for LDPC code of rate 0.5 from table II in [9] (dv=50), for repetition code of rate 0.5 and for a *good code*.

The results require the channel quality to be determined by a single parameter w such as 1/SNR in a manner we denoted 'incremental noisiness', see the next section.

In the following derivation we extend the result to the Binary Symmetric Channel (BSC) using the concept of Generalized Extrinsic Information Transfer (GEXIT) introduced in [10] and exhibiting an inherent area property. The derivation is based on a more general one in [2, part 2] yielding a compact proof for the BSC channel.

GEXIT, area theorems and their applications to important aspects of iteratively decoded systems were studied in [11], [12] and references therein. Related derivatives of mutual information were used in [13] to obtain mutual information in coded systems.

Notations: Mutual information is denoted by I, entropy by H, P and p denote probability and a probability density function respectively and bold letters denote vectors.

II Models and Definitions

A. *Channel model*:

We examine a time invariant BSC characterized by P(y|x) where and x and y are the input and the output bits of the channel respectively. C denotes the channel capacity.

The channel transfer function P(y|x) is determined by a single noisiness parameter denoted by w as to preserve the notation of previous papers, in our case it is chosen identical to the probability of error Pe.

A channel characterized by w^h, for some $w^h > w^l$, can be described as physically degraded with respect to a channel characterized by w^l, where the superscripts h and l denote 'higher' and 'lower' noisiness respectively. The outputs of the two channels are denoted y^h and y^l. By physically degraded we mean that x_i, y_i^l, y_i^h form the Markov chain:

$$x_i - y_i^l - y_i^h \tag{1}$$

It can be seen easily that the BSC conforms to this property.

In order to perform partial derivatives in our derivation, we shall need an extension to a channel characterized for each channel use i by a different w_i, so we use $\mathbf{w}=[w_1,w_2\ldots w_i..w_N]$. Non-bold w will mean that all w_i are equal to w. Derivative with respect to w_i with the other elements of **w** held constant is denoted as $\frac{d()}{dw_i}$, while $\frac{d()}{dw}$ denotes a derivative with respect to the channel quality w common to all the elements of **w**.

B. The good code:

We desire to transmit information **U**. We do so in the standard manner of transmitting a codeword **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 **U** and is equi-probable. The received vector is denoted by **y**. The code is such that it achieves vanishing error probability at channel noisiness level w_0 for which R=C-ε' for any small positive ε' when n is large enough. The Markov chain (1) implies vanishing error probability also for all $w<w_0$.

C. The Extrinsic Information:

We are interested in a symbol x_i, which is a symbol at the i'th position in **x**. We define $\mathbf{x}_i'$ as **x** with x_i excluded and correspondingly $\mathbf{y}'_i$ as **y** with y_i excluded.

We denote by z_i the complete information obtainable from $\mathbf{y}_i'$ about x_i, known as the Extrinsic Information, see for example [6] and [10]. The extrinsic information z_i over the binary channel can be expressed, for example, as the Logarithmic Likelihood Ratio of x_i given $\mathbf{y}_i'$. The average extrinsic information measure is then defined as

$$\text{EI}=\frac{1}{\text{n}}\sum_{\text{i}=1}^{\text{n}}\text{I}(x_i;\mathbf{y}'_i)=\frac{1}{\text{n}}\sum_{\text{i}=1}^{\text{n}}\text{I}(x_i;z_i)\ . \qquad (2)$$

.

When x_i is given, then y_i is independent of $\mathbf{y}'_i$, that is $P(y_i \mid x_i, \mathbf{y}'_i) = P(y_i \mid x_i)$ as shown already in [6]. This extends the Markov chain (1) to

$$\mathbf{y}'^h_i - \mathbf{y}'^l_i - x_i - y^l_i - y^h_i \qquad (3)$$

$$z_i - x_i - y^l_i - y^h_i \qquad (4)$$

Furthermore, due to this Markov chain and the data processing theorem we have $I(\mathbf{y}'^l_i; x_i) \ge I(\mathbf{y}'^h_i; x_i)$, thus EI is a non-increasing function of the noisiness w.

$$EI(w^l) \ge EI(w^h) \qquad (5)$$

III EXIT CHART OF GOOD CODES

When the code rate R is strictly below the capacity C 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

$$R < C \Rightarrow EI = 1 \text{ bits per symbol} \qquad (6)$$

This intuitive attribute of *good codes* is proved in [1] and [2].

The central result of this work is the following proposition and its method of proof:

Proposition: The average EI, eq. (2)*, which can be obtained from all the channel outputs about the coded bits x_i of a good code operating over a BSC, the capacity of which is below the code rate, is zero.*

Proof:

We shall examine a code which is capacity achieving for some noisiness $w=w_0$, that is, it is capable of reliably transmitting information at a rate $R = C(w_0) - \varepsilon'$ over a channel parameterized by $w=w_0$ for any small ε'.

Our proof uses the concept of GEXIT as introduced in [10]. GEXIT is defined [10, eq. (2)] as:

$$GEXIT = \frac{1}{n}\frac{d}{dw}H(\mathbf{x} \mid \mathbf{y}) \qquad (7)$$

which is transformed easily to:

$$GEXIT = \frac{1}{n}\frac{d}{dw}[H(\mathbf{x}) - I(\mathbf{x};\mathbf{y})]$$

$$GEXIT = -\frac{1}{n}\frac{d}{dw}I(\mathbf{x};\mathbf{y}) \qquad (8)$$

GEXIT is an average of $GEXIT_i$, [10, eq. (3)]:

$$GEXIT = \frac{1}{n}\sum_{i=1}^{n} GEXIT_i \qquad (9)$$

where $GEXIT_i = \frac{d}{dw_i}H(\mathbf{x} \mid \mathbf{y})$.

It was shown in [10, eq. (5)] that:

$$GEXIT_i = \frac{d}{dw_i}H(x_i \mid z_i, y_i) \qquad (10)$$

Similarly to (8), $GEXIT_i$ from (10) is

$$GEXIT_i = -\frac{d}{dw_i}I(x_i; z_i, y_i) \qquad (11)$$

Now let us introduce GEXIT0 and $GEXIT0_i$ defined as GEXIT and $GEXIT_i$ respectively but with the transmitted symbols $x0_i$ uncoded and distributed independently, which implies $z_i=0$. Each $x0_i$ is distributed according to the symbol-wise distribution of the corresponding x_i in our *good code* X which may be dependent to a certain limited extent [14] on the symbol index i. We denote this input distribution by the superscript ^{IND} and the corresponding mutual information as $I^{IND}(\mathbf{x}0;\mathbf{y}0)$.

We shall see below that a plot of GEXIT and GEXIT0 versus Pe would resemble the plots of MMSE in Fig. 1 for the coded and uncoded transmissions respectively. In fact MMSE is (up to a constant factor) the GEXIT (7) over the AWGN channel as pointed out in [10] and GEXIT is used in this work to a similar purpose as the MMSE in [1].

Since (11) applies also to $GEXIT0_i$:

$$GEXIT0_i = -\frac{d}{dw_i}I(x0_i; y0_i)\ .$$

Now let us compare $GEXIT_i$ to $GEXIT0_i$. We shall use (11) and the fact that each couple (x_i, y_i) is distributed identically to $(x0_i, y0_i)$:

$$GEXIT0_i - GEXIT_i = -\frac{d}{dw_i}I(x_i; y_i) + \frac{d}{dw_i}I(x_i; z_i, y_i)$$

$$= \frac{d}{dw_i}\left[I(x_i; z_i, y_i) - I(x_i; y_i)\right]$$

By the chain rule of mutual information we have then:

$$GEXIT0_i - GEXIT_i = \frac{d}{dw_i}\left[I(x_i;z_i \mid y_i)\right]$$
$$= \frac{d}{dw_i}\left[I(z_i;x_i,y_i) - I(z_i;y_i)\right]$$

By the Markov chain (4) the first I( ) term is not dependent on y_i and we have:

$$GEXIT0_i - GEXIT_i = -\frac{d}{dw_i} I(z_i;y_i) \qquad (12)$$

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 [15, Theorem 15].

Specifically, it is proved in [1] and in [2] that for $w>w_0$ the mutual information I ($\mathbf{x};\mathbf{y}$) over the channel with the good code is similar to $I^{IND}(\mathbf{x}0;\mathbf{y}0)$. That is for any small $\varepsilon \geq 0$ and sufficiently large n:

$$0 \leq \frac{1}{n} I^{IND}[\mathbf{x}_0;\mathbf{y}_0(w)] - \frac{1}{n} I[\mathbf{x};\mathbf{y}(w)] \ = \gamma(w) \leq \varepsilon^3 = \varepsilon' \qquad (13)$$

where γ is a non-increasing function of w. The substitution $\varepsilon' = \varepsilon^3$ will be required below.

It follows from (13) that for any pair of values w1 < w2 for which R>C we have:

$$0 \leq$$
$$I^{IND}[\mathbf{x}_0;\mathbf{y}_0(w_1)] - I^{IND}[\mathbf{x}_0;\mathbf{y}_0(w_2)] - \{I[\mathbf{x};\mathbf{y}(w_1)] - I[\mathbf{x};\mathbf{y}(w_2)]\}$$
$$= n\gamma(w_1) - n\gamma(w_2) \leq n\varepsilon^3 \ .$$

Applying (8) we get

$$0 \leq \int_{w_1}^{w2} [GEXIT0(w) - GEXIT(w)]dw \leq \varepsilon^3 \ . \qquad (14)$$

So GEXIT is nearly equal to GEXIT0 at w>w0 similarly to the MMSEs with and without coding in Fig. 1. Also, like the MMSE, the GEXIT is zero when R<C, see (8).

We denote the difference GEXIT0-GEXIT by DG. From (12), (9) and the Markov chain (4), DG is non-negative:

$$DG(w) \triangleq GEXIT0(w) - GEXIT(w) \geq 0 \qquad (15)$$

Now lets place w_1 at the threshold value $w_1=w_0$. The average of DG over the interval of $w_1 = w_0 \ \ to \ w_2 = w_0 + \Delta$ cannot exceed ε^3/Δ, otherwise its integral (14) would exceed ε^3. We shall choose $\Delta = \varepsilon$ to limit the average DG to ε^2. Since DG is non-negative, there is some $w=w_t$ in the above interval, ε within w_0, for which DG is bounded by

$$0 \leq DG(w_t) \leq \varepsilon^2 \qquad (16)$$

Due to (5), vanishing EI at w_t implies vanishing EI at all larger values of w, so it is sufficient to prove vanishing EI at w_t.

It follows from (9), (12), (15) and (16) that:

$$0 \geq \frac{1}{n}\sum_i \frac{d}{dw_i} I(z_i;y_i) \geq -\varepsilon^2 \qquad (17)$$

By the Markov chain (4) each element of the above sum is negative and their average is lower bounded by $-\varepsilon^2$. This implies that each element is lower bounded by $-\varepsilon$, except at most $n\varepsilon$ elements which may be more negative (with $n\varepsilon$ elements more negative than $-\varepsilon$, (17) will be violated). This vanishing proportion of elements can contribute only ε bits to the average EI, eq. (2), because the EI for each bit is bounded by 1, so we can disregard them in our proof of vanishing average EI and use

$$0 \geq \frac{d}{dw_i} I(z_i;y_i) \geq -\varepsilon \qquad (18)$$

To simplify the proof relative to [2, part 2] we assume in the remaining derivation that

$$P(x_i = 0) = P(x_i = 1) = 0.5 \ \ . \qquad (19)$$

In any case a good code cannot deviate too far from (19), see [14] and references therein. Then

$$\frac{d}{dw_i} I(z_i;y_i) = \frac{d}{dw_i}[H(y_i) - H(y_i \mid z_i)]$$
$$= \frac{-d}{dw_i} H(y_i \mid z_i)$$

We used $\frac{d}{dw_i} H(y_i) = 0$ because

$$P(y_i=1) = P(y_i=0) = 0.5 \qquad (20)$$

due to (19) and to the symmetry of the channel.

Then by (18):

$$0 \leq \frac{d}{dw_i} H(y_i \mid z_i) \leq \varepsilon \ \ . \qquad (21)$$

For the above derivative to vanish, $H(y_i|z_i)$ must not change when w_i, that is Pe of the BSC, increases. Since y_i is binary and the channel is symmetric, this is clearly possible only if

$$P(y_i = 0 \mid z) = P(y_i = 1 \mid z) = 0.5 \ \ . \qquad (22)$$

The last equation, together with (20), implies that y_i and z_i are statistically independent.

This, together with the Markov chain (4), yields independence between z_i and x_i because the only dependence possible between z_i and x_i would be of the form $P(x_i = 0 | z_i) \neq 0.5$ which would be reflected into dependence between y_i and z_i over the BSC with any Pe different from 0.5. The independence between z_i and x_i proves the proposition.

See [2, part 2] for an alternative and a more general proof starting from (18).

Remarks:

1) At decreasing channel noisiness which brings the channel capacity to a value above the code rate there is a sharp transition in the EI curve since I(**x**,**y**) is not determined by (13) any more but reaches a plateau at the code rate and the GEXIT (8), which is proportional to the derivative of I(**x**,**y**) with respect to w, goes to zero abruptly. This transition takes place over a small region of w for which the term (13) is very small but, significantly, not zero. A similar transition of the average mutual information $\frac{1}{n}\sum_{i=1}^{n} I(y_i;\mathbf{y}')$, which is upper bounded by the average EI because of the Markov chain (3), occurs in the same region.

2) Clearly the area A under the EI versus $I(x_i;y_i)$ curve, $A = \int_0^1 EI\, d[I(x_i;y_i)]$, equals (1-R), thus the step function EXIT chart derived here for good code over the BSC conforms to the "EXIT chart area property" of outer codes which was proved in [7] for any code over the binary erasure channel.

IV CONCLUSIONS:

We proved that the EXIT chart of any *good* (capacity achieving) *code* operating over a BSC is a step function of the channel noisiness, zero when the code operates at code rate above capacity and 1 at C>R . This is an extension of previous results over the binary erasure channel and the binary input AWGN channel. Each proof relies on some area property. Thus codes good over memoryless 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 conforms to the EXIT chart area property derived in [7] for the erasure channel.

Furthermore, vanishing EI, as defined in (2), implies that a 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 $\mathbf{y'}_i$.

REFERENCES:

[1] M. Peleg, A. Sanderovich, and S. Shamai (Shitz): " On Extrinsic Information of Good Binary Codes Operating Over Gaussian Channels", to be published in the *European Transaction on Telecommunications (ETT),*see http://www3.interscience.wiley.com/cgi-bin/abstract/112748555/ABSTRACT

[2] 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.

[3] A. Sanderovich, M. Peleg, and S. Shamai (Shitz): " LDPC Coded MIMO Multiple Access with Iterative Joint Decoding *", IEEE Transactions on Information Theory*, Volume: 51, Issue: 4, pp.1437-1450, April 2005

[4] S. ten Brink: "Designing Iterative Decoding Schemes with the Extrinsic Information Transfer Chart", *AEU International Journal of Electronics and Communications*, vol. 54 no. 6, Dec. 2000, pp. 389-398.

[5] M. Peleg and S. Shamai (Shitz): "Efficient Communication over the Discrete-Time Memoryless Rayleigh Fading Channel with Turbo Coding/Decoding", *European Transaction on Telecommunications (ETT),* vol. 11, Sept/Oct 2000, pp. 475-485.

[6] C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding," *Proceedings of Int. Communications Conf. (ICC'93)* Geneva, Switzerland, May 1993, pp. 1064-1070.

[7] A. Ashikhmin, G. Kramer, S. ten Brink: "Code Rate and the Area under Extrinsic Information Transfer Curves", *Proceedings of the International Symposium on Information Theory,* 2002, p.115, IEEE

[8] D. Guo, S. Shamai (Shitz) and S. Verdu, ``Mutual Information and Minimum Mean-Square Error in Gaussian Channels", *IEEE Transactions on Information Theory,* Volume: 51, Issue: 4, pp. 1261-1282 April 2005

[9] T. J. Richardson, M. A. Shokrollahi and R. L. Urbanke, "Design of Capacity-Approaching Irregular Low-Density Parity-Check Codes", *IEEE Trans. on Information Theory,* Vol. 47, No. 2, Feb. 2001, pp. 619-637.

[10] C. Measson, R. Urbanke, A. Montanari, and T. Richardson: "Life Above Threshold: From List Decoding to Area Theorem and MSE", *IEEE Information Theory Workshop* (ITW), Oct. 24-29 2004, Texas.

[11] C. Measson, A. Montanari, T. Richardson and R. Urbanke, "The Generalized Area Theorem and Some of its Consequences", in *ArXiv* , http://tw.arxiv.org/PS_cache/cs/pdf/0511/0511039.pdf

[12] C. Measson, A. Montanari and R. Urbanke, "Why We Can Not Surpass Capacity: The Matching Condition*," the 43rd Allerton Conference on Communications, Control and Computing*, Allerton House, Monticello, Illinois, USA, September 28-30, 2005.

[13] D. P. Palomar and S. Verdu, "Computation of Mutual Information for Memoryless Channels with Arbitrary Inputs," submitted to *the IEEE Trans. Information Theory*.

[14] S. Shamai (Shitz) and S. Verdu, "The Empirical Distribution of Good Codes*", IEEE Trans. on Information Theory*, Vol. 43, No. 3, pp. 836-846, May 1997.

[15] T. S. Han and S. Verdu, "Approximation theory of output statistics," *IEEE Transactions on Information Theory*, Volume: 39, Issue: 3, May 1993, pp. 752 – 772.