

An Information Theoretic Approach to Turbo Codes

Shlomo Shamai (Shitz)¹ and Sergio Verdú²

¹Department of Electrical Engineering, Technion-Israel Inst. of Technology, Haifa 32000, Israel sshlomo@ee.technion.ac.il

²Department of Electrical Engineering, Princeton University, Princeton, New Jersey 08544, U.S.A. verd@Princeton.edu

I. ABSTRACT

In a series of recent papers [5] [2] [9] [4] [6] [10] [11] [3] [7] [8] the concept of so-called parallel-concatenated codes has been proposed and analyzed. Foremost among that class of codes, the Turbo codes proposed in [5] were shown to exhibit a very favorable complexity-performance tradeoff, closely approaching the Shannon theoretic performance limit. Although based on important existing insights on parallel concatenated codes and iterated decoding, the discovery of Turbo codes was largely experimental. A flurry of activity is now under way directed towards the understanding of those codes. A number of coding theoretic explanations of the structure and performance of Turbo codes have recently been put forth recently in the foregoing references and other recent work. It is natural to inquire whether Shannon theory offers any explanation for the unexpectedly excellent performance of Turbo codes. Although, complexity questions are generally outside the purview of Shannon theory, we have found and analyzed a simple information theoretic model which lends itself to an interpretation of the good performance of Turbo codes. Our basic model and analysis were given in [13] and [12]

Consider a communication system where two independent channels operate in parallel. If the inputs to both channels were allowed to be encoded, then Shannon's coding theorem tells us that the source is reliably transmissible provided its entropy rate is below the sum $C_1 + C_2$ of the channel capacities; conversely, if the source entropy rate exceeds $C_1 + C_2$ then reliable transmission is not possible. The new twist in the model of [12] is that the information going through channel 2 is not encoded. A number of practical scenarios fit into this model. Of particular interest to the present paper is the case where a single channel is time-multiplexed into several independent subchannels.

The main result in [12] states

Theorem 1 [12]. *The source can be transmitted reliably provided that its conditional entropy rate given the output of the uncoded channel, $H(\mathbf{X}|\mathbf{Z})$, is below the capacity C_1 of channel 1, and, conversely, it cannot be transmitted reliably if the conditional entropy rate exceeds C_1 .*

This result suggests that we view the information rate of the source as split into two nonoverlapping components, $H(\mathbf{X}) = H(\mathbf{X}|\mathbf{Z}) + I(\mathbf{X}; \mathbf{Z})$. Even though the information quantified by the second term is transmitted uncoded, the source is reproducible with arbitrary reliability at the output. If, furthermore, the source is matched to the uncoded channel in the sense that it maximizes its input/output mutual information, then it is possible to transmit information at rate $C_1 + C_2$ even though no coding is provided for the information going through one of the channels. This implies that the sum of the capacities of K independent parallel binary symmetric channels can be achieved even if only one of them is encoded. This observation is most striking when the encoded BSC has very poor crossover probability.

Parallel-Concatenated codes, and in particular Turbo codes

[5], can be cast within our side-information model by considering a single-channel time-multiplexed into several independent subchannels. For example, one subchannel transmits the uncoded raw data (the Turbo codes are systematic), and two parallel channels are driven by partial encoders which can be viewed as joint source-channel encoders driven by a redundant source. A practically appealing way to ensure that the information encoded by the partial encoders is nonoverlapping is by prepending a sufficiently long interleaver at the input of one of the encoders. This setup is more attractive than simply multiplexing the source because of the complexity reductions of combined source/channel coding with high compression ratios. Good component codes in Parallel-Concatenated schemes are able to trade to some extent the traditional role of reducing the uncertainty of the source given the channel outputs for the easier goal of preserving mutual information. The channel is an additive Gaussian channel with binary inputs, whose capacity is denoted by $C_b(SNR)$, and for rate-1/2 codes, $SNR = E_b/N_0$. The goal in [5] is to obtain a bit-error-rate of 10^{-5} ; rate-distortion theory establishes that the allowable rate is $1 - h(10 \text{sup} - 5) = 0.9998$ times the rate required for arbitrarily reliable communication. $C_b(0.7 \text{dB}) = 0.54$ bits per channel use. Since the source puts out 0.5 bits per channel use, we conclude that it is indeed transmissible through the channel. (We note incidentally that the cutoff rate at this SNR is 0.36 bits per channel use.) The safety margin between 0.5 and 0.54 bits per channel use is equal to $\Gamma_{0.7 \text{dB}} = -0.24 \text{ dB}$, where Γ_{SNR} is defined via:

$$\frac{1}{2} = C_b(\Gamma_{SNR} SNR).$$

We will consider two different ways of time-multiplexing the channel (the first in three branches, and the second in two branches). With each partition we will develop several complementary insights.

In the first partition we examine the systematic part is assigned to Channel 3; and the output of each convolutional encoder (consisting of parity check bits only) is assigned to Channel 1 and 2, respectively. In this case, Channel 3 is uncoded. Therefore, according to Theorem 1, the net rate to be reliably transmitted over the combination of both partially coded channels is the conditional entropy of the source given the output of Channel 3:

$$H(\mathbf{X}|\mathbf{Z}) = \frac{1}{2} - \frac{1}{2} C_b(SNR)$$

where we have taken into account that Channel 3 is used once every two time units, and its input/output mutual information is equal to its capacity because it is driven by a Bernoulli(1/2) source. For $SNR = 0.7 \text{dB}$, $H(\mathbf{X}|\mathbf{Z}) = 0.23$ bits per time unit. Since the Turbo code apportions this rate equally to Channels 1 and 2, the effective rate to be transmitted over Channel 1 (or Channel 2) is equal to 0.115 bits per time unit, with both encoders being fed information at the rate of 0.5 bits per time unit. Since each of those channels are used once every four

uses of the original channel, their capacity is equal to 0.135 bits per time unit. A calculation analogous to the foregoing reveals that the safety margin is now 1.2dB. To achieve this, the effective information transmitted through both channels must be nonoverlapping. In the Turbo code, both encoders are identical, but one of them is preceded by an interleaver which makes the information streams at the input of both encoders independent (relative to decoding horizon). The crucial role of the interleaver can also be seen from the classical angle that dictates long constraint lengths to approach the Shannon limit.

Consider now an alternative partitioning which multiplexes the original channel into Channels 1 and 2, such that the systematic part and the output of convolutional encoder A are assigned to Channel 2, while the output of convolutional encoder B (parity check stream) is assigned to Channel 1. Thus Channel 1 [resp. 2] is used once [resp. thrice] every four uses of the original channel. Instead of considering both of the channels as being partially coded, we will return to our original setting where Channel 2 is used uncoded by simply viewing it as the cascade of encoder and the actual binary-input channel. The equivalent uncoded channel is no longer memoryless, but in view of the generality of Theorem 1, we will apply its conclusions to this case. Denote the capacity (in bits per time units) of the equivalent combined Channel 2 by $\frac{3}{4} C_b(SNR)$ (every 4 time units this equivalent channel is used three times). The conditional entropy of the source given the output of Channel 2 is now

$$H(\mathbf{X}|\mathbf{Z}) = \frac{1}{2} - \frac{3}{4} C_b(\Gamma_{SNR}^c SNR)$$

where Γ_{SNR}^c quantifies the degradation in capacity due to the presence of the encoder in the equivalent Channel 2. The capacity in bits per time unit of Channel 1 is equal to $1/4 C_b(SNR)$. The transmissibility condition of Theorem 1 is now:

$$\frac{1}{2} - \frac{3}{4} C_b(\Gamma_{SNR}^c SNR) \leq \frac{1}{4} C_b(SNR)$$

which gives a lower bound on the information degradation that each encoder (from source to parity check) is allowed to incur:

$$\Gamma_{0.7 \text{dB}}^c \geq -0.7 \text{dB}$$

pointing to the fact that the information loss of the component encoders in the Turbo code is remarkably small.

The foregoing observations apply whenever an optimum maximum likelihood decoder (or a near-optimum decoder) is used. The optimum decoder consists of a decoder for Channel 1 which is fed the conditional probability of the sourcewords given the lower channel outputs. The computational complexity of an optimum decoder for Channel 1 which uses nonuniform probabilities for the sourcewords is high. Likewise, once the encoder for the lower channel is introduced, the generation of the conditional probability of the sourcewords given the Channel 2 output words is nontrivial. A practical, but suboptimal, approach adopted in the decoding of Turbo codes is an iterative scheme, whereby the *symbolwise* conditional probabilities are computed in lieu of the desired sourceword conditional probabilities. The rationale for this is the availability of relatively efficient backward-forward dynamic programming algorithms [1] and [14] to carry out that computation. Soft-decoding of Channel 1 using that information leads to the

consideration of the reversal of the roles of both channels and to an iterative scheme where the quality of the estimates of the source symbols is improved.

We summarize some of our main conclusions:

- Based on fundamental Shannon theoretic arguments we conclude that Parallel Concatenated codes are an appealing structure to achieve rates close to capacity.
- Slepian-Wolf source coding, while not necessarily part of actual (suboptimal) codes, provides a key to the understanding of Parallel Concatenated codes. Furthermore, the optimal structure Slepian-Wolf encoder-Channel 1 encoder boils down to a linear encoder for binary-input symmetric memoryless channels.
- Good component codes in a Parallel Concatenated scheme are able to trade (to some extent) the traditional role of reducing the uncertainty of the source given the channel outputs for the easier goal of conserving information (in a mutual information sense). Systematic recursive convolutional codes are particularly appealing in that respect, especially when iterative decoding is used.
- Good complexity-performance tradeoffs are achieved by parallel partial encoders each of which has the task of conveying a fraction of the source information. The key to these gains is that rather than multiplexing the source into independent channels, interleavers achieve the goal of generating nonoverlapping information at the encoder outputs, while preserving a high compression ratio (beneficial for reduced complexity) for each of the individual encoders.

ACKNOWLEDGEMENTS

This work was supported in part by a grant from the U.S.-Israel Binational Science Foundation

References

- [1] L.R. Bahl, C.D. Cullum, W.D. Frazer, and F. Jelinek. An efficient algorithm for computing free distance. *IEEE Trans. Inform. Theory*, IT-18:437-439, May 1972.
- [2] G. Battail. Pseudo-random recursive convolutional coding for near-capacity performance. *2nd International Symposium on Communication Theory and Applications*, July 11-16, 1993. B. Honary, D. Darnell and P. Farrell Eds., Communications and Signal Processing Series: Communications Theory and Applications II, pp. 54-65, HW Communications Ltd., 1994.
- [3] S. Benedetto and G. Montorsi. Average performance of parallel concatenated block codes. *Electronics Letters*, 31, No. 3:156-158, February 2, 1995.

- [4] C. Berrou and A. Glavieux. Turbo codes: General principles and applications. *Sixth Terrena Workshop on Digital Communication*, 1993.
- [5] C. Berrou, A. Glavieux, and P. Thitimajshima. Near shannon limit error correcting coding and decoding: Turbo-codes(1). *Proc. International Conference on Communications*, pages 1064-1070, May 23-26, 1993.
- [6] D. Divsalar and F. Pollara. Turbo codes for pcs applications. *ICC '95*, June 18-22, 1995.
- [7] J. Hagenauer. Iterative decoding of block and convolutional codes. *Proceedings of the EIDMA Winter Meeting on Coding Theory, Information Theory and Cryptology*, page 32, December 19-21, 1994.
- [8] J. Hagenauer, P. Robertson, and L. Papke. Iterative (turbo) decoding of systematic convolutional codes with the map and sova algorithms. *ITG-1994 Source and Channel Coding Conference*, October 26-28, 1994.
- [9] J. Lodge, P. Hoher, and J. Hagenauer. The decoding of multidimensional codes using separable map filters. *Queen's University Sixteenth Biennial Symp. on Communications*, pages 343-346, May 1992.
- [10] P. Robertson. Understanding turbo-codes - are they capable of near shannon limit error correction. *Proc. Sixth JCCC Joint Conf. Communications and Coding*, March 1994.
- [11] P. Robertson. Illuminating the structure of code and decoder of parallel concatenated recursive systematic (turbo) codes. *IEEE Int. Conf. on Commun., ICC '94*, pages 1298-1303, May 1-5, 1994.
- [12] S. Shamai (Shitz) and S. Verdú. Capacity of channels with side-information. *European Transactions on Telecommunications*, page 7, 1995.
- [13] S. Shamai (Shitz) and S. Verdú. Capacity of channels with uncoded-message side-information. *Proc 1995 IEEE Int. Symp. on Information Theory*, page 7, 1995.
- [14] S. Verdu and H. V. Poor. Abstract dynamic programming models under commutativity conditions. *SIAM J. Control and Optimization*, 24:990-1006, Jul. 1987.

Tutorial Introduction to Iterative (Turbo)-Decoding

Joachim Hagenauer
 Chair for Communications
 Technical University of Munich
 Arcisstr.21, D-80290 Munich, Germany
 email: hag@LNT.e-technik.tu-muenchen.de

February 20, 1996

Abstract

Information and coding theorists always attempted to come close to the Shannon limit performance with a tolerable complexity. Recently decoding of two and more dimensional product-like codes where proposed with iterative ('turbo') decoding [Ber93] following earlier ideas in [Bat79].

We will give a tutorial introduction into the principle of turbo decoding. A brief review of Shannon's capacity limit shows the goal to be reached. A simple example with single parity check codes shows all the ingredients of turbo decoding and exemplifies the need to use log-likelihood algebra for a brief notation and easy implementation.

We will show that for iterative decoding "soft-in/soft-out" decoders are needed and that they exist for block and convolutional codes: MAP[Bah74], SOVA [Hag89] and LOGMAP [Pap94]. We further have decoders available which use dual codes for high rate block [Hag96] and convolutional codes, modified threshold decoders and suboptimal approximations thereof.

A brief review of the state of the art will complete the talk.

- [Ber93] C. Berrou, A. Glavieux and P. Thitimajshima. "Near Shannon limit error correcting coding and decoding: turbo-codes". *Proc. of ICC '93*, Geneva, pp. 1064-1070, May 1993.
- [Bat79] G.Battail, M.C.Decouvelare, P.Godlewski. "Replication Decoding". *IEEE Trans. Inform. Theory*, Vol. IT-25, pp. 332-345, May 1979.
- [Bah74] L.R. Bahl, J. Cocke, F. Jelinek and J. Raviv. "Optimal decoding of linear codes for minimizing symbol error rate". *IEEE Trans. Inform. Theory*, Vol. IT-20, pp. 284-287, March 1974.