A Serial-Parallel Concatenated System: Construction and Iterative Decoding with Erasures *

Jan Bajcsy and Hisashi Kobayashi
Department of Electrical Engineering
Princeton University, Princeton, NJ 08544
e-mail: bajcsy@ee.princeton.edu, hisashi@ee.princeton.edu
Fax: (609) 258-3745, Tel: (609) 258-1984

Abstract

We attempt to improve the performance of an existing wireless packet transmission system that uses coding and transmitter diversity. We view it as a concatenated system, system constructed by "concatenation" of encoding blocks and interleavers at the transmitter.

The transmitter structure is improved using a better interleaving scheme, the receiver architecture by applying an iterative decoder with AZD (ambiguity zone detection) structure. This receiver uses generalized erasures, which are resolved in an iterative process. Using this framework, overall encoding and decoding gain of almost 9 dB is achieved with small increase in complexity and decoding delay.

1 Introduction

Some of the best codes known today are based on the basic idea of constructing large codes from simple building blocks via concatenation. Smaller encoders are assembled into an encoding network/cascade at the transmitter and the data are passed through this cascade before being transmitted. This idea was originally used by Elias to obtain product codes, later in Gallager's low density parity check codes and concatenated codes of Forney [8]. It is this approach of combining several old codes into a new, powerful code, that gives rise to the most powerful codes known today, e.g., Turbo codes of Berrou et al. [5]. Moreover, many existing communication systems have such a concatenated structure.

Recently, iterative decoding has been widely studied as an efficient low complexity decoder, since in many cases an overall optimal decoder for a concatenated system may be too complex and an obvious technique based on one-path decoding may be far from being optimal. Decoding of Turbo codes [5], soft decision decoding scheme discussed by Hagenauer et al. in [9] in decoding a concatenated code with Reed-Solomon code and convolutional code, serial concatenation of convolutional codes [4], and an erasure based iterative decoder [1] are just a few of these.

Non-uniform interleavers are one of the key ingredients for good performance of Turbo codes [5] and, more generally, of generalized concatenated codes [1]. Good, relatively short and simple to describe interleavers can significantly improve the overall performance of a concatenated system causing manageable increase in complexity and decoding delay. Recently, we proposed a class of such non-uniform interleavers which offer a significant performance improvement [3].

This paper builds on these ideas to improve the performance of a wireless packet transmission system by redesigning its transmitter and improving the receiver structure.

2 Concatenated Systems

2.1 Encoder Concatenation

We define a *"concatenated system"* as a system that can be built by concatenation (serial, parallel or combination) of encoding blocks that include some redundancy or memory. Such blocks can be formally described as mappings from a message set $\mathcal{M}$ to a codebook $\mathcal{C}$

$$f : \mathcal{M} \rightarrow \mathcal{C} \subseteq \mathcal{X}, \tag{1}$$

where the set $\mathcal{X}$ denotes inputs either to the next encoding block or to the channel. Such an encoding block then maps a message $\mathbf{m}$ into a codeword $\mathbf{c} = f(\mathbf{m})$. The encoding blocks include not only channel encoders, as in the conventional concatenated codes, but also channels with spectral shaping (e.g., partial-response coding) or line coding, channels with ISI (intersymbol interference) and/or multipath delay, modulators with memory

*This work has been supported, in part, by the National Science Foundation, the New Jersey Commission on Science and Technology, Asahi Chemical Co. Ltd, Seiko Communications Systems, and the Ogasawara Foundation for the Promotion of Science and Engineering.

(e.g., trellis coded modulation, continuous phase modulation).

The concatenation may be not only a result of having a concatenated error correcting code at the transmitter side, but also of a cascade of encoders at the transmitter. For instance, a product code may be followed by a run-length limited code and a partial response channel in digital recording applications; time and/or space diversity may be used by the transmitter or receiver of a wireless communication system; a channel with multipath or ISI may follow a coded system in other applications. This can be described using composite mappings, i.e., if overall the encoder f obtained by concatenation, then in serial case

$$f(\mathbf{m}) = f_l(f_{l-1}(...(f_1(\mathbf{m})))), \tag{2}$$

in parallel case

$$f(\mathbf{m}) = [f_1(\mathbf{m}); f_2(\mathbf{m}); ...; f_l(\mathbf{m})] \tag{3}$$

and a combination of these two approaches would denote in a hybrid case. (For notation simplicity, we consider interleavers/permutations that may be inserted between these building blocks as parts of encoder maps.) Product codes and recently introduced Turbo codes [5] can be seen as special cases of the concatenated systems.

2.2 Iterative Decoding with Erasures

We introduced a novel receiver structure [1] which combines AZD (ambiguity zone detection) and iterative decoding. The idea of AZD (or sometimes called the null zone detector) was successfully applied to a partial response system [7]. In our iterative decoding, AZD provides flexibility of deferring decision on unreliable digits by labeling them as "ambiguous digits" or "generalized erasures". The decisions about symbols being ambiguous can be made either at the receiver front of a continuous channel, or at constituent decoders of the iterative decoder. The latter is particularly important for codes whose MAP decoders are too complex to be built (e.g., Reed Solomon codes), or in systems, where low decoder complexity is required due to the nature of the application.

The iterative decoder makes step-by-step resolutions of these ambiguous digits by capitalizing on the redundancy introduced by the error correcting code, and modulation/channel with memory. For example, in case of two encoding blocks concatenated in parallel, the receiver obtains from the channel noisy versions of $f_1(\mathbf{m})$ and $f_2(\mathbf{m})$ denoted as $\mathbf{r}$ and $\mathbf{q}$. The iterative decoder can be then written as follows

$$\begin{aligned} \mathbf{a}_n &= \varphi_2(\mathbf{q}, \mathbf{b}_{n-1}) \\ \mathbf{b}_n &= \varphi_1(\mathbf{r}, \mathbf{a}_n), \end{aligned} \tag{4}$$

where the decoders 1 and 2 are denoted as maps φ_1 and φ_2 and the intermediate decisions after the n-th iteration are $\mathbf{a}_n, \mathbf{b}_n \in \mathcal{M}_E$ – the message set extended by messages with erased symbol. Similarly, in serial case the receiver obtains a noisy version of $f_2(f_1(\mathbf{m}))$ denoted as $\mathbf{y}$ and the iterative decoding proceeds as

$$\begin{aligned} \mathbf{a}_n &= \varphi_2(\mathbf{y}, \mathbf{b}_{n-1}) \\ \mathbf{b}_n &= \varphi_1(\mathbf{a}_n). \end{aligned} \tag{5}$$

The iterative decoding technique we introduce is, in concept, similar to iterative decoding procedures used in decoding Turbo codes [5] and soft decision decoding scheme discussed by Hagenauer et al. [9] in decoding a concatenated code with Reed-Solomon code and convolutional code.

3 System Model

This section describes the original wireless packet transmission system as well as the channel coding improvements at the transmitter and receiver end. Packets of 240 coded bits are generated from 160 information bits and duobinary modulation is then used for spectral efficiency purposes. The goal of this system is to achieve packet loss bellow 10^{-3}. The decoding complexity and storage are limited due to size of the receiver and battery power.

3.1 Existing Structure

The current transmitter is depicted as a discrete baseband system in Figure 1. Each packet with 160 information bits is first encoded the by an error correcting code – a shortened Hamming (12, 8) code with generator matrix

$$\mathbf{G} = \begin{bmatrix} 1&0&0&0&0&0&0&0&1&1&0&0 \\ 0&1&0&0&0&0&0&0&1&0&1&0 \\ 0&0&1&0&0&0&0&0&0&1&1&0 \\ 0&0&0&1&0&0&0&0&1&1&1&0 \\ 0&0&0&0&1&0&0&0&1&0&0&1 \\ 0&0&0&0&0&1&0&0&0&1&0&1 \\ 0&0&0&0&0&0&1&0&1&1&0&1 \\ 0&0&0&0&0&0&0&1&0&0&1&1 \end{bmatrix} \tag{6}$$

Resulting 240 bits are permuted by a 20×12 block interleaver and then passed to the duobinary modulator that includes both differential precoding and actual duobinary modulation denoted in discrete time model by

$$g(D) = 1 + D. \tag{7}$$

To achieve sufficiently high packet throughput, space and time diversity are used. Each packet is transmitted

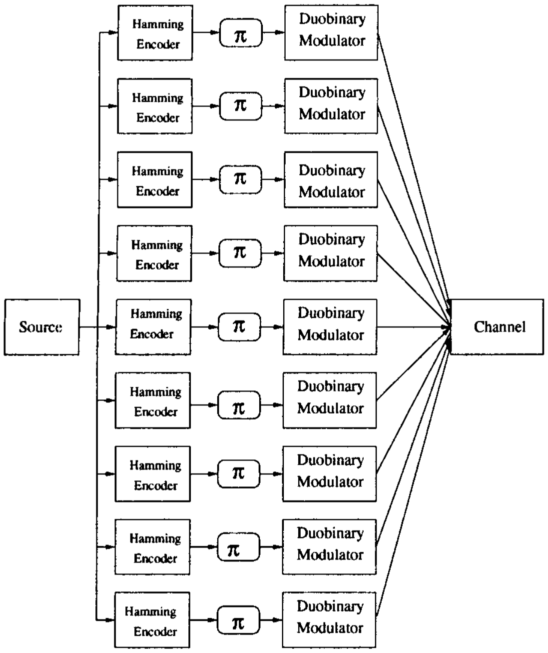

Figure 1: Existing transmitter from Section 4 including time and space diversity.

from three different locations, three different times from each location, i.e., each packet is transmitted nine times overall. We denote transmitted packets of duobinary symbols as vectors $\mathbf{x}^r$ of length 240, $(r = 1, 2, ..., 9)$, and $\mathbf{y}^1; \mathbf{y}^2; ...; \mathbf{y}^9$ their received noisy versions. We assume additive white Gaussian noise,that is

$$\mathbf{y}^r = \mathbf{x}^r + \mathbf{z}^r \tag{8}$$

with noise components $\mathbf{z}^r[i]$, i=1, 2, ..., 240, are assumed to be i.i.d. normal random variables $N(0, \sigma^2)$. The channel signal to noise ratio (SNR) is then $\frac{E_s}{\sigma^2}$, where E_s is the average energy of a channel symbol.

The current system receiver, shown in Figure 2(a), is based on bit-by-bit detection of the duobinary signal and a syndrome decoder of the (12, 8) code. Nine independent trials for decoding of each packet are made by the receiver and if any of them is successful, the packet is successfully received.

Successful reception of a decoded message is decided at the sink using an error detection code, that is a part of the source. The error detection code allows almost complete elimination of incorrectly decoded messages.

3.2 Improved Interleaving Scheme

As a concatenated system, the original transmitter has two main weaknesses in terms of used interleaving. First

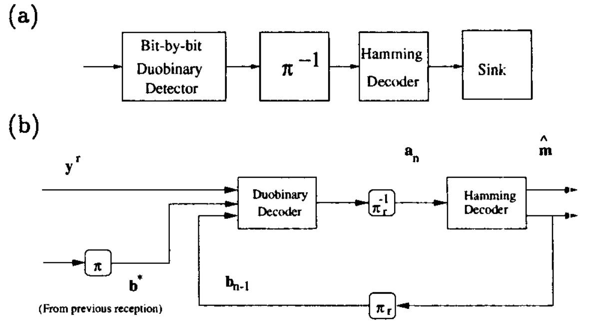

Figure 2: Block diagrams of the (a) original receiver, (b) iterative AZD based decoder utilizing diversity.

of all, the same interleaver is used in every reception, hence the same message is sent nine times. Although this simple diversity scheme is easy to implement, a lot of worthy information is being wasted by sending the same data, using essentially a repetition code like approach. In addition, used interleaver is a uniform interleaver and these are known not to perform well in iterative decoding schemes.

A much better strategy is to send each time differently permuted version of the encoded message and use nonuniform interleavers for good iterative decoding at each reception. The last requirement on the permutations is that they must be easy to describe. This is due to receiver's small size and consequent storage constraint. (For instance, nine random permutations of length 240 would require too much storage for their look-up tables.)

Let $\pi^r(i)$ denotes the position of the i-th symbol after permutation π^r, where $r = 1, 2, ..., 9$ denotes the number of transmission/reception attempt. To satisfy the requirements stated above, we constructed each of permutations $\pi^1, \pi^2, ..., \pi^9$ of length 240 as a combination of 3 permutations as follows

$$\pi^r(i) = \pi_3(\pi_2^r(\pi_1^r(i))) \qquad r = 1, 2, ..., 9. \tag{9}$$

These first two parts of the overall π^r make them different from one transmission to another and can be easily implemented using a 20×12 array and a small look-up table. Each permutation π_1^r reorders bits within each Hamming codeword. It is different for each r and can be described using a permutation of length 12. Permutations π_2^r reorder 20 different Hamming codewords constituting each packet, to avoid bits from same pair of blocks being neighbors repetitively. The reorderings are different for each r.

Finally, for π_3 we decided to use a permutation from the class studied in [3], since they offer good spreading of low weight patterns. The selected π_3 is defined as

$$\pi_3(i) \equiv i + 16\Delta_i \quad (mod\ 240), \tag{10}$$

where

$$\Delta_i = \Delta_j \quad \text{for } i \equiv j \pmod 4 \tag{11}$$

and $\Delta_0 = 0$, $\Delta_1 = 1$, $\Delta_2 = 3$, $\Delta_3 = 5$. This permutation can be easily implemented as using a circular buffer of size 240, where the data are first read in, then shifted cyclically to the right and then read out [3]. Thus we have obtained permutations with low complexity of description and implementation, which achieve good spreading of bits.

3.3 Iterative Receiver Using Diversity

The iterative receiver is depicted in Figure 2(b) and is based on the idea originally considered in [2]. Two concatenated decoders working in tandem are helping one another to correct more errors and remove erasures as the decoding proceeds during each reception, while utilizing the decisions from previous reception.

This decoder, depicted schematically in Figure 2(b), can be written as follows:

$$\begin{aligned} \mathbf{a}_n &= \varphi_2(\mathbf{y}^r, \mathbf{b}^*, \mathbf{b}_{n-1}) \\ \mathbf{b}_n &= \varphi_1(\mathbf{a}_n) \end{aligned} \tag{12}$$

Both the duobinary decoder φ_2 and Hamming decoder φ_1 are implemented based on their trellis structures – trellises with two and eight states respectively. These decoders are capable of performing decoding with erasures using an erasure outputting algorithm, that can be thought as a low complexity modification of the SOVA algorithm [6] for handling and outputting binary erasures.

Let $l_I(e)$ and $l_O(e)$ denote the input and output labels of the trellis along each edge e. Consider reception r and edge e going from the $(i-1)$-th stage to the i-th trellis stage of the duobinary trellis. At the n-th iteration, the incremental metric at the decoder along this edge e is proportional to

$$\begin{aligned} m(e) = &\ \| \ \mathbf{y}^r[i] - l_O(e)\|^2 \\ &+ c\|\mathbf{b}^*[i] - l_I(e)\|^2 + c\|\mathbf{b}_{n-1}[i] - l_I(e)\|^2 \end{aligned} \tag{13}$$

where c is a scaling factor and the decisions $\mathbf{b}^*[i]$ and $\mathbf{b}_{n-1}[i]$ about the i-th bit come from the previous reception and previous iteration respectively. Similarly, for the Hamming decoder, the incremental metric along the edge f, f going from the stage $(i-1)$ to stage i in the Hamming trellis, is proportional to

$$m(f) = c\|\mathbf{a}_n[i] - l_O(f)\|^2. \tag{14}$$

Note that the initial conditions, i.e., $\mathbf{b}_0$ during the first iteration and $\mathbf{b}^*$ during the first reception, are initialized with all bits being erased.

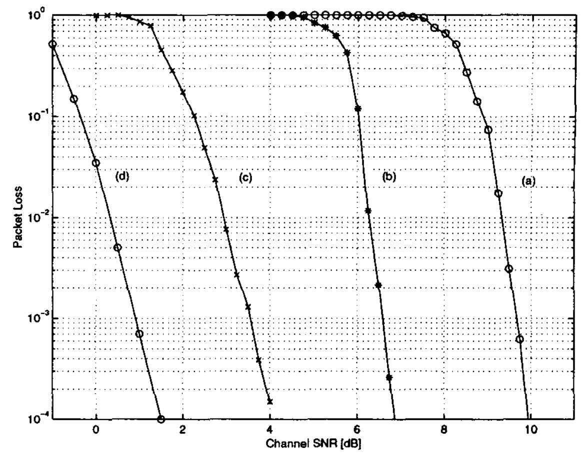

Figure 3: Overall packet loss for four different receivers: (a) Original transmitter and receiver. (b) Original transmitter and iterative AZD based receiver not using diversity. (c) Original transmitter and iterative AZD based receiver using diversity. (d) Improved transmitter and iterative AZD based receiver that uses diversity.

Each decoder serves as ambiguity zone detectors (AZD), discussed originally in [7] in detection of partial response signals. Here, the ambiguous decisions are made in the VA upon mergers of competing paths with cumulative metrics M_1 and M_2. Symbols in which those two paths differ are labeled as erased if the metric difference is bellow a threshold, i.e.,

$$\Delta = |M_1 - M_2| < t, \tag{15}$$

whereas if

$$\Delta = |M_1 - M_2| \geq t, \tag{16}$$

decided symbols and their erasure flags are kept from the winning path.

The stopping rule in step 3, i.e., deciding on correct decision of the packet, is done using the error detecting code that is a part of the source. Hence the receiver brings no time delay in reception of good packets and has almost no significant additional storage needs, since channel data $\mathbf{y}^r$ is not stored from previous receptions.

4 Performance Improvement

The simulation results on the additive white Gaussian noise channel are shown in Figure 3 for four different systems. Curve (a) shows the performance for the original transmitter and receiver; curve (d) is the performance curve for the improved transmitter and receiver from

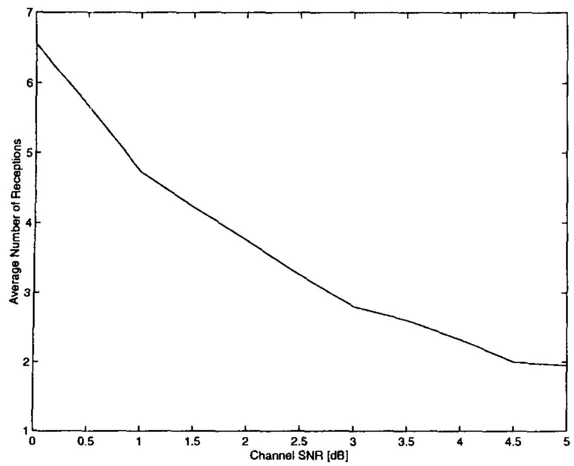

Figure 4: Average number of needed packet receptions for successfully decoded packets in the constructed transmitter and receiver system.

Section 3. Improvement of almost 9 dB is achieved for required *overall* packet loss of 10^{-3}.

To see which part of this improvement is due to the improved encoding and which is due to improved decoding, two other curves are shown for comparison. Curve (b) shows the performance of the original transmitter, with the iterative receiver starting afresh at each reception. (I.e., iterative receiver not utilizing the diversity.) Curve (c) shows the original transmitter, as performing with the iterative receiver from previous section.

Finally, for the improved transmitter and its receiver, it is also interesting to observe, how the average number of receptions needed to decode a packet successfully increases with lowered channel SNR (Figure 4). At higher SNR's, the decoder is able to receives the message almost immediately. At lower SNR's, the decoder has to wait for later receptions of a packet to recover it sufficiently.

5 Conclusion

This paper first discusses concatenated systems, i.e., digital communications or recording systems in which data are encoded by a cascade of encoders at the transmitter. The connecting pattern of the encoders can be in series, in parallel and a combination thereof. In fact, recently invented Turbo codes [5] can be viewed as a parallel version of such concatenated system. A receiver structure [1] which combines AZD (ambiguity zone detection) and iterative decoding, is then reviewed for these systems.

This framework is then applied to improve an existing wireless packet transmission system that uses coding and diversity. As the simulation results in Sections 4 show, improved transmitter and iterative receiver bring almost 9 dB gain in system's performance with only a small complexity increase.

Further improvements in the performance of the wireless packet transmission system are possible by further modifying the transmitter side. The current error correcting code is a shortened Hamming code. Although it is easy to implement at both encoding and decoding sides, the use of a more powerful code at the transmitter may bring further performance improvement.

References

[1] J. Bajcsy and H. Kobayashi, "Error control of generalized concatenated systems," Proceedings of the 1997 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM'97), Victoria, B.C., Canada, August 1997, pp. 749-752.

[2] J. Bajcsy and H. Kobayashi, "Iterative decoding with erasures in a concatenated system with diversity". *Proc. of ISCC'98*, pp. 100-104, Athens, Greece, June 30-July 2, 1998.

[3] J. Bajcsy and H. Kobayashi, "Good interleavers for concatenated finite shift register codes" *Proc. of ISITA '98*, pp. 327-330, Mexico City, Mexico, October 1998.

[4] S. Benedetto, D. Divsalar, G. Montorsi, F. Pollara, "Continuous MAP algorithms and their applications to decode parallel and serial code concatenations," *Proc. ESA International Workshop on DSP Techniques Applied to Space Communications*, Barcelona, Spain, Sept. 1993, pp. 8.10-8.24.

[5] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error correcting coding and decoding: Turbo codes (I)," *Proc. ICC '93*, Geneva, Switzerland, May 23-26, 1993, pp. 1064-1070.

[6] J. Hagenauer, P. Hoeher, "A Viterbi algorithm with soft-decision outputs and its applications," *GLOBECOM 1989*, Dallas, Texas, *Conference Record*, pp. 1680-1686.

[7] H. Kobayashi and D. T. Tang, "On decoding of correlative level coding systems with ambiguity zone", *IEEE Trans. Communications*, Vol. COM-19, Aug. 1971, pp. 467-477.

[8] W. W. Peterson and E. J. Weldon, *"Error-correcting Codes,"* MIT Press, Cambridge, 1972.

[9] S. B. Wicker and V. K. Bhargava (eds.), *Reed-Solomon Codes and their Applications*, IEEE Press, 1994.