Reliability-based Hybrid ARQ using Convolutional Codes

Abhinav Roongta and John M. Shea
Wireless Information Networking Group
Department of Electrical and Computer Engineering
University of Florida
Gainesville, Florida 32611-6130, USA
Email: abhinavr@dsp.ufl.edu, jshea@ece.ufl.edu

***Abstract*— Reliability-based hybrid ARQ (RB-HARQ) is a recently introduced approach to incremental-redundancy ARQ. In RB-HARQ, the bits that are to be retransmitted are adaptively selected at the receiver based on the estimated bit reliabilities. This technique has the potential to improve performance and minimize retransmission size by targetting those bits that are likely to be in error. However, previous versions of the RB-HARQ algorithm have resulted in large request (NACK) packets that made the technique appropriate only for systems that can tolerate large request packets on the feedback link. In this paper, we show that RB-HARQ is effective with convolutional codes, and we exploit the time-correlation properties of these codes to significantly reduce the size of the retransmission requests.**

I. INTRODUCTION

Incremental-redundancy hybrid-ARQ schemes that use punctured convolutional codes and code combining are considered in [1]. In these and other hybrid ARQ schemes [2], [3], the set of bits that is to be transmitted in response to error detection is a predetermined part of the ARQ algorithm. In [4] a reliability-based hybrid ARQ (RB-HARQ) algorithm was proposed that requests retransmission for those bits that are deemed unreliable at the output of a soft-input, soft-output (SISO) decoder. SISO decoders, which can be used with many modern error-control coding schemes, typically accept estimates of the *a priori* probabilities and output estimates of the *a posteriori* probabilities (APPs) for the message bits. The RB-HARQ technique can be used with codes that use SISO decoders, such as turbo codes and convolutional codes.

The performance of a RB-HARQ technique using turbo codes was shown in [4]. The RB-HARQ technique was shown to offer throughput close to capacity if used with an appropriate initial code rate. However, an asymmetric channel was considered in which the retransmission request packet could be made very large to include the indices of all of the unreliable bits. The size of this retransmission packet can potentially be quite large. For example, if the packet size is approximately 1000 bits, then each bit position can be represented by a 10-bit index. If 30 bits are to be retransmitted, then the retransmission request packet will consist of 300 bits if no source coding is applied.

In this paper, we present results for a RB-HARQ scheme that uses convolutional codes. Not only are convolutional codes currently deployed in many more applications than the turbo codes considered in [4], but error events in convolutional codes are highly correlated in time. The temporal correlation is usually not preserved in turbo codes due to the presence of the interleaver [5] and the fact that many errors correspond to failures of the decoder to converge and are not associated with a specific path in the code trellis. In this paper, we show that the time correlation in the output of the decoder for a convolutional code offers one way to significantly reduce the size of the retransmission request packet.

This paper is organized as follows. In Section II, we present some motivation for the proposed RB-HARQ schemes. The system model is described in Section III. In Section IV we present the results for a RB-HARQ scheme with constant incremental redundancy and large retransmission-request packets. The results for a RB-HARQ scheme with variable incremental redundancy and small retransmission-request packets are given in Section V. Finally, in Section VI we conclude the paper.

II. MOTIVATION

The RB-HARQ technique that we propose is motivated by an understanding of the decoding process and analysis of the error packets. We use the MAP algorithm [6] for the decoding of convolutional codes. For each information bit u_k, the decoder computes the log *a posteriori* probability (log-APP) ratio [7] as follows

$$L(u_k) = \log\left(\frac{P(u_k = +1|\mathbf{y})}{P(u_k = -1|\mathbf{y})}\right), \tag{1}$$

where $\mathbf{y}$ is the received codeword in noise. When the decoder fails to decode a packet correctly, it is because the decoder has failed to find soft-decision log-APP values with the correct signs for some of the information bits in the packet. The bits that have soft-decision log-APP values with incorrect signs result in errors at the decoder output.

Analysis of error packets reveals that the decoder can use the log-APP values to accurately identify the bits that prevent the packet from decoding correctly [4]. We refer to such bits as *weak bits*. To see this, consider a block of 1000 information bits encoded by a rate $1/2$ convolutional code with generator polynomials $1 + D^2$ and $1 + D + D^2$ for transmission over an additive white Gaussian noise (AWGN)

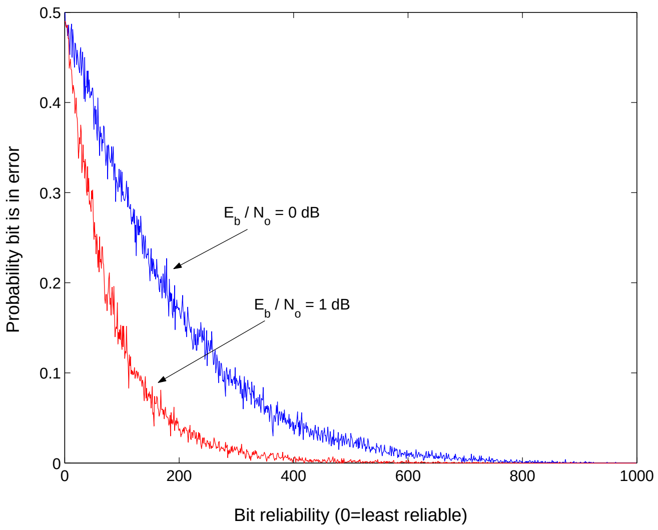

Fig. 1. Probability of bit error by reliability rank for rate 1/2, (5,7) convolutional code.

channel. For each error packet, rank the bits at the output of the convolutional decoder by the magnitude of their soft-decision log-APP values. The bit with the smallest soft-output is considered the least reliable (0), and the bit with the largest soft-output is considered most reliable (999). The probability of error for each bit by rank is shown in Figure 1. These results indicate that the least reliable bits correspond to errors about 50% of the time, while very reliable bits are rarely in error. Thus the bits that have small log-APPs are likely to be the weak bits. The performance of the decoder is likely to improve if additional information about the weak bits can be used to improve their soft-decision estimates. In this paper, we propose a hybrid-ARQ scheme in which additional information for the weak bits is provided by retransmissions.

III. SYSTEM MODEL

Consider the communication system shown in Figure 2. The source radio S and the destination radio D are linked by a data channel through which a packet of information is to be delivered from S to D. The data bits in S are convolutionally encoded and the resulting code bits are modulated using BPSK. The encoded packet is then transmitted over an AWGN channel. The destination radio D attempts to decode the packet and sends a retransmission request through the feedback channel if an error is detected. The retransmission-request packet, sent from D to S, contains a list of the least-reliable information bits based on the magnitudes of the log-APPs. The source S then retransmits the code bits corresponding to those requested information bits. Noisy versions of the retransmitted code bits are received at D and they are added to the previously received values of the same code bits. In our study, we assume perfect error detection and the presence of a highly reliable feedback channel from D to S.

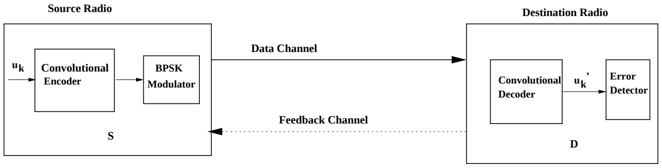

Fig. 2. System Model.

For all of the results presented in this paper , the code used for transmission from S to D is a rate 1/2, constraint length $K = 7$ convolutional code with generator polynomials (in octal) 554 and 744. In this paper, we present the results for a block size of 1000 information bits. In Section IV, we present the results for a constant incremental redundancy RB-HARQ scheme in which we make no attempt to reduce the size of retransmission-request packet. In Section V, we present results for a variable incremental redundancy RB-HARQ scheme with reduced size retransmission-request packet.

IV. RB-HARQ WITH CONSTANT INCREMENTAL REDUNDANCY

We first investigate the performance of the RB-HARQ scheme with no compression of the request packet. We will use the results to determine the relative degradation of the approach to compressing the request packet that is discussed in Section V. These results also apply if the feedback channel has a high capacity so that a large retransmission-request packet can be sent from D to S. The source S initially transmits the packet using a rate 1/2 convolutional code. If D fails to decode the packet correctly, it sends a retransmission-request packet containing a list of the positions of the 50 least-reliable information bits. In [4], S responds to the request packet by retransmitting the information bits. However, the code in [4] is a systematic turbo code, whereas the code we consider in this paper is a nonsystematic convolutional code. Thus, for these results, S retransmits the two code bits corresponding to each of the positions identified by D. To further clarify, D is using the reliability of the information bits to identify weak sections in the code trellis and then requests new code information for those trellis sections. The received code symbols are combined with all previously received copies of those symbols. For BPSK transmission over an AWGN channel, the soft-outputs for the symbols can be added together.

For the results presented in this paper, each packet consists of 1000 information bits. Each retransmission request consists of a list of 50 bit positions, and S transmits 100 code bits in response to each request. This corresponds to 5% incremental redundancy per retransmission. We consider the performance when the request and retransmission process can occur up to three times. Each retransmission effectively reduces the code rate and hence increase the E_b/N_o at the receiver. We account for this additional received energy by defining the *effective* E_b/N_o as the average E_b/N_o at the receiver, taking into account the average number of incremental redundancy transmissions per packet.

The results in Figure 3 show the probability of bit error for reliability-based hybrid-ARQ with the rate 1/2, constraint-

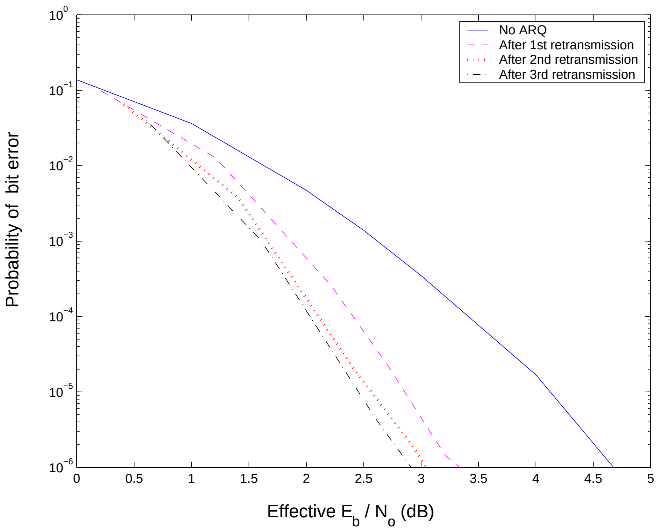

Fig. 3. Probability of bit error vs. Effective E_b/N_o for three retransmissions of 5.0% incremental redundancy each.

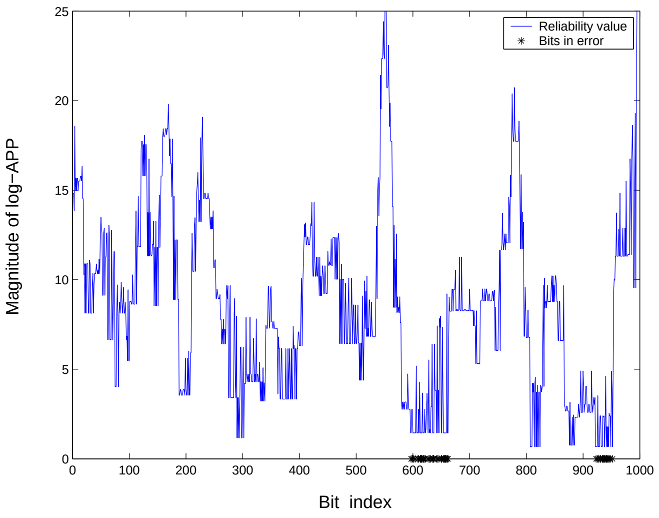

Fig. 4. Reliability values for example packet of 1000 information bits.

length seven convolutional code. In Figure 3 we observe that to achieve a probability of bit error of less than 10^{-6}, a system using RB-HARQ technique with three retransmissions of 5.0% incremental redundancy each requires 1.9 dB lower E_b/N_o than a system with no ARQ. Most of the performance has been gained after only two incremental transmissions, and the third transmission only improves performance by approximately 0.1 dB. Further improvement in performance may be achieved by optimizing the number of bits retransmitted in each iteration.

We note that for the technique presented in this section, the retransmission-request packet can be very large. Consider the following example. For a packet of 1000 information bits, each bit index can be represented by a ten bit binary number. So, without any compression, the retransmission-request packet consisting of 50 least-reliable bit indices, will consist of 500 bits. Such a large retransmission-request packet will generally decrease the overall system throughput. In the next section, we present results for a variable redundancy RB-HARQ scheme which has a much smaller retransmission-request packet.

V. RB-HARQ with Variable Redundancy and Smaller Request Packet

The scheme that we propose in this section is based on two important observations during our simulations. The first observation, as shown in Figure 1, is that in any packet with errors, the bits that are in error have low reliability (magnitude of log-APP) values. The second observation is that in any packet with errors, the error events (the bits that are in error) are correlated in time. The results in Figure 4 illustrate the reliability values for each bit in an example packet that was decoded in error as function of the bit index (position in the packet). The packet size is 1000 information bits, and it was transmitted over an AWGN channel using the rate $1/2$ convolutional code with constraint length $K = 7$. The results in Figure 4 also indicate the bits that were in error. We observe from the figure that bits that are in error have low reliability values and occur in groups (time-correlated). There is one group of error bits around bit index 600 and another group of error bits around bit index 950.

Based on the two observations made above, we modify the RB-HARQ technique proposed in the previous section. The system model remains the same as in Figure 2. Whenever the destination radio D fails to decode a packet correctly, it calculates a threshold γ based on the reliability values of the bits in that packet. Then it performs an *elimination* operation in which all the reliabilities greater than γ are made zero. Following the *elimination* operation, D performs a *smoothing* operation as follows:

$$L(u_k) = \frac{\sum_{m=-2}^{2} L(u_{k+m})}{5}. \tag{2}$$

In our study, the threshold γ is calculated as follows:

$$\gamma = \alpha + 0.1 \cdot \mu, \tag{3}$$

where α is the minimum reliability value and the μ is the average reliability value of the packet in consideration. We were guided by the following considerations while selecting the threshold in (3):

(i) The threshold calculation should be computationally simple.
(ii) The threshold should be large enough so that it is greater than the least reliability value because the bits with low reliabilities are the ones that are likely to be in error.
(iii) The threshold should be small enough that the size of retransmission request packet is small and the number of bits retransmitted is not very large.

The *smoothing* operation was performed using a rectangular window of length 5 as described by (2). Figure 5 shows

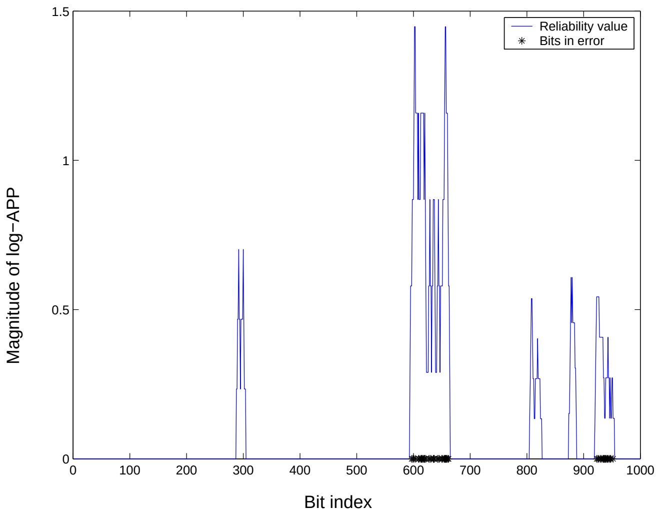

Fig. 5. Reliability values, after elimination and smoothing for example packet of 1000 information bits.

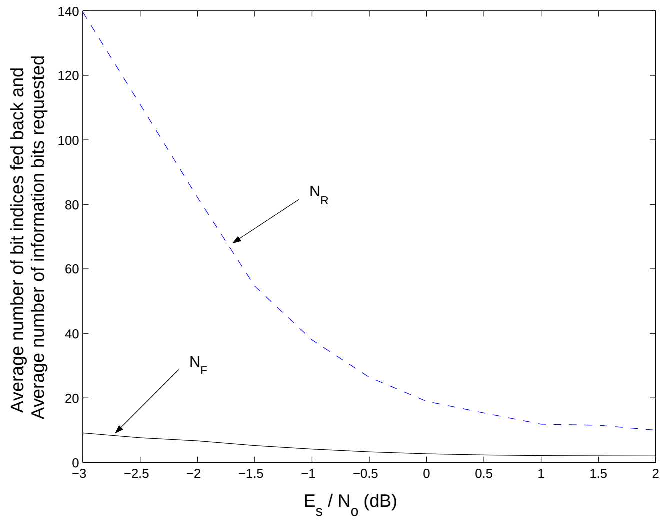

Fig. 6. Average number of bit indices fedback (N_F) and average number of information bits requested for retransmission (N_R) vs. E_s/N_o for RB-HARQ scheme.

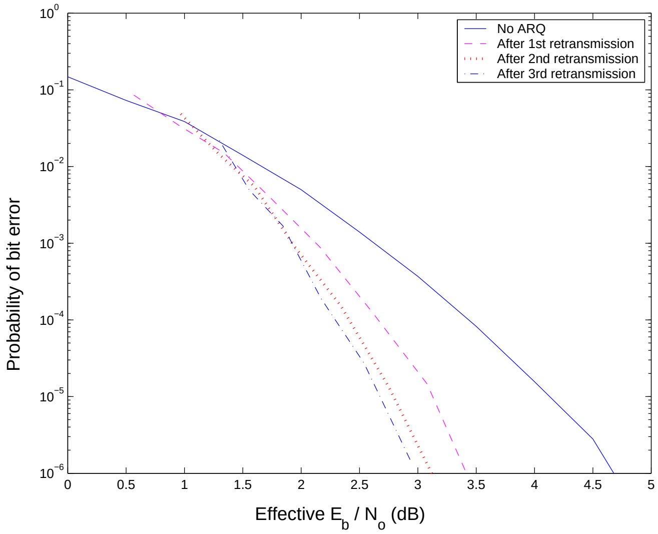

Fig. 7. Probability of bit error vs. Effective E_b/N_o for RB-HARQ scheme with variable redundancy and reduced retransmission-request packet

the reliability values, after the *elimination* and *smoothing* operations were performed, for the packet with errors shown in Figure 4. We observe in Figure 5 that there are 5 windows (groups) of non-zero reliabilities in the entire packet. The destination radio, D, sends the first bit index and the last bit index, of each window, to S. The source S then retransmits the code bits, corresponding to all the information bits, in each of the window. Thus, the number of bit indices sent back from D to S is fewer than the number of information bits that are actually requested for retransmission.

We define N_F to be the average number of bit indices per retransmission-request packet sent from D to S. We also define N_R to be the average number of information bits requested for retransmission for every packet in error. The results in Figure 6 show the above two quantities (N_F and N_R) at various values of the channel symbol energy-to-noise density ratio (E_s/N_o). We observe that a large reduction in the size of retransmission-request packet has been obtained. For example, at -3 db the the average number of bit indices per retransmission-request packet (N_F) is 9.1 whereas the average number of information bits requested per packet with errors (N_R) is 139.5. At 2 dB, N_F is 2.0 and N_R is 11.0. We note that in the RB-HARQ technique presented in the previous section, all the bit indices had to be fed back ($N_F = N_R$) to the source radio. Thus we have obtained more than 80 percent reduction in the size of retransmission-request packet. The scheme presented in this section has variable redundancy compared to fixed redundancy in the previous section. By doing this we are able to take advantage of better channel conditions. As E_b/N_o improves, we request fewer information bits and hence, fewer code bits are retransmitted. Hence, the redundancy decreases with increasing SNR, which leads to higher throughput.

The results in Figure 7 show the probability of bit error for a system that uses RB-HARQ technique with variable redundancy and small retransmission-request packets. Figure 7 shows that to achieve a probability of bit error of less than 10^{-6}, a system using the above ARQ technique requires 1.7 dB lower E_b/N_o than a system with no ARQ. We note that this improvement in the system performance has been obtained by using a simple heuristic for calculation of threshold γ. System performance can be further improved by optimizing the threshold, packet size and the window length used for *smoothing* operation. The RB-HARQ scheme in this section performs about 0.2 dB worse than the scheme in the previous section, but reduces the size of the retransmission-request packet by at least 80 percent at all signal to noise ratio (SNR) values.

VI. CONCLUSION

In this paper, we develop reliability-based hybrid ARQ schemes for convolutional codes. We show that the schemes provide significant performance gain for systems that have a reliable feedback channel. The proposed ARQ scheme achieves a gain of approximately 2 dB over a system with no ARQ when a large retransmission-request packet is used. We show that the error events in a convolutional code are highly correlated in time. We modify the proposed scheme by implementing a simple heuristic that leads to more than 80 percent reduction in the size of retransmission-request packet. The heuristic utilizes the high time-correlation of the error events in the convolutional code to reduce the size of retransmission-request packet. The amount of redundancy is very small at moderate to high signal to noise ratio thus leading to a wide applicability of the proposed scheme to future communication systems.

ACKNOWLEDGMENT

This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Office of Naval Research under Grant N00014-00-1-0565, by the Office of Naval Research under Grant N00014-02-1-0554, and by the National Science Foundation under Grant ANI-0220287.

REFERENCES

[1] J. Hagenauer, “Rate compatible punctured convolutional codes (RCPC-codes) and their application,” *IEEE Trans. Commun.*, vol. 36, pp. 389–400, Apr. 1988.

[2] S. Lin and D. Costello, Jr., *Error Control Coding:Fundamentals and Applications*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1983.

[3] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1995.

[4] J. M. Shea, “Reliability-based hybrid ARQ,” *IEE Electronics Letters*, vol. 38, pp. 644–645, June 2002.

[5] C. Berrou, A. Galvieux, and P. Thitimajshima, “Near Shannon limit error-correcting coding and decoding,” in *Proc. 1993 IEEE Int. Conf. Commun.*, vol. 2, (Geneva, Switzerland), pp. 1064–1070, 1993.

[6] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, “Optimal decoding of linear codes for minimizing symbol error rates,” *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 284–287, Mar. 1974.

[7] W. E. Ryan, “Concatenated convolutional codes and iterative decoding.” Web Page: `http://www.ece.arizona.edu/ ryan`.