

A Hybrid ARQ Scheme Based on Polar Codes

Kai Chen, *Student Member, IEEE*, Kai Niu, *Member, IEEE*, and Jia-Ru Lin, *Member, IEEE*

Abstract—A hybrid automatic repeat request (HARQ) scheme based on a novel class of rate-compatible polar (RCP) codes are proposed. The RCP codes are constructed by performing punctures and repetitions on the conventional polar codes. Simulation results over binary-input additive white Gaussian noise channels (BAWGNs) show that, using a low-complexity successive cancellation (SC) decoder, the proposed HARQ scheme performs as well as the existing schemes based on turbo codes and low-density parity-check (LDPC) codes. The proposed transmission scheme is only about 1.0-1.5dB away from the channel capacity with the information block length of 1024 bits.

Index Terms—Polar codes, hybrid ARQ, rate-compatible coding, successive cancellation decoding.

I. INTRODUCTION

POLAR codes, recently introduced by Arkan [1], are the first structured codes that provably achieve the symmetric capacity of binary-input memoryless channels (BMCs). Given a BMC W , by performing the channel polarizing transformation over N independent copies of W , we get a second set of synthesized BMCs. As the transformation size N goes infinity, some of the resulting channels tend to completely noised, and the others tend to noise-free, where the fraction of the latter approaches the symmetric capacity of W . By transmitting free bits over the noiseless channels and sending fixed bits over the others, polar codes can achieve the symmetric capacity under a successive cancellation (SC) decoder with both encoding and decoding complexity $O(N \log N)$.

In delay insensitive communications, hybrid automatic repeat request (HARQ) is widely used to obtain capacity-approaching throughput efficiency [2], [3], [4]. This paper focuses on designing an HARQ transmission scheme of polar codes by using punctures and repetitions. The proposed transmitter block diagram is shown in Fig. 1. As the initial transmission, an information block of K bits is fed into a polar encoder. The output codeword of N_0 bits are punctured into N_1 bits and sent over the channel. If the receiver fails to decode the codeword, a NACK (negative acknowledgement) is sent to the transmitter through the feedback channel. And then, $N_2 - N_1$ of the information bits are retransmitted. The receiver tries to perform decoding with all the received N_2 bits. If the decoding is failed again, another $N_3 - N_2$ bits are transmitted. This process continues until the transmitter receives an ACK (acknowledgement), or a maximum number of transmissions T is achieved.

This work was supported in part by the National Basic Research Program of China (973 Program) (No. 2009CB320401), the National Natural Science Foundation of China (No. 61171099), the National Science and Technology Major Project of China (No. 2012ZX03003-007) and Qualcomm Corporation.

The authors are with the Key Laboratory of Universal Wireless Communication, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China. (e-mail: {kaichen, niukai, jrlin}@bupt.edu.cn)

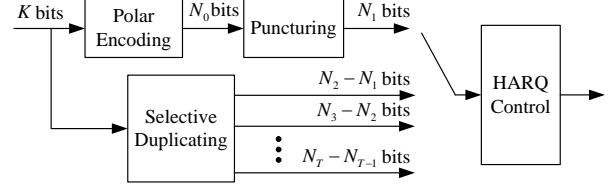


Fig. 1. Transmitter block diagram of the proposed HARQ scheme.

The remainder of the paper is organized as follows. Section II describes the rate-compatible polar (RCP) codes used in the proposed HARQ scheme. Section III proposes the design method of the HARQ scheme based on RCP codes. Section IV provides the simulation comparison of HARQ schemes based on polar codes, turbo codes and LDPC codes. Finally, Section V concludes the paper.

II. RATE-COMPATIBLE POLAR CODES

In this paper, we use calligraphic characters, such as \mathcal{X} to denote a set. Let $|\mathcal{X}|$ denote the number of elements in \mathcal{X} . We write lowercase letters (e.g., x) to denote scalars, bold-face lowercase letters (e.g., \mathbf{x}) to denote vectors, and x_i to denote the i -th element of \mathbf{x} . For any $i \leq j$, $\mathbf{x}_{i:j}$ denotes a subvector of \mathbf{x} , i.e., $\mathbf{x}_{i:j} = (x_i, x_{i+1}, \dots, x_j)$. The bold numbers $\mathbf{0}$ and $\mathbf{1}$ are used to denote the all-zero and all-one vectors, respectively. Throughout this paper, the base of the logarithm is 2.

The proposed transmission scheme depicted in Fig. 1 is based on a class of RCP codes with punctures and repetitions. The coded bits of a RCP code consist of two parts: the first part come from a punctured polar encoder, and these bits are called *polar bits*; the other bits are the selective copies of the information bits and are called *repetition bits*. Hereafter, the RCP code is represented by a three-tuple (N, K, M) , $K \leq M \leq N$, which means this code has an information block length K , and the resulting N -length codeword consists of M polar bits and $N - M$ repetition bits.

To construct a (N, K, M) RCP code over a BMC W , we first need to construct a punctured polar code with code length M . The punctured polar coding is discussed in [7].

Similar to constructing a conventional polar code, after performing a polarization transform on $N_0 = 2^{\lceil \log M \rceil}$ independent uses of W , we get N_0 successive uses of synthesized binary input channels $W_{N_0}^{(i)}$, $i = 1, 2, \dots, N_0$, where $\lceil \cdot \rceil$ is the top integral function.

Given a symmetric BMC W , let \mathbf{a} denotes the probability density function (pdf) of the log-likelihood ratio (LLR) of the received bit when a bit zero is transmitted. The reliability of W can be measured as the error probability

$$P_e(W) = \int_{-\infty}^0 \mathbf{a}(x) dx \quad (1)$$

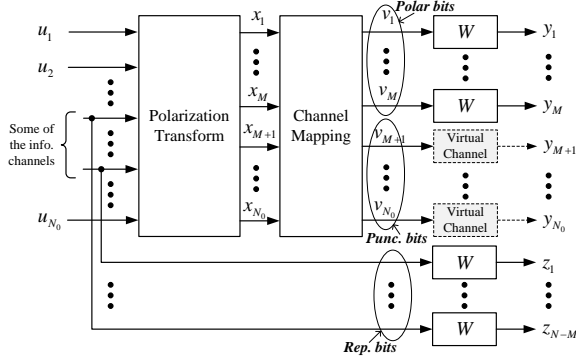


Fig. 2. Channel transform of RCP coding.

Let $\mathbf{a}_{N_0}^{(i)}$, $i = 1, 2, \dots, N_0$ denote the LLR pdfs of the received bit from $W_{N_0}^{(i)}$ when all-zero information bits are transmitted. After calculating $\mathbf{a}_{N_0}^{(i)}$ by density evolution (DE) [5], the reliabilities of $W_{N_0}^{(i)}$ are determined by (1). In transmitting a binary information block of K bits, the K most reliable polarized channels $W_{N_0}^{(i)}$ with indices $i \in \mathcal{A}$ are selected to carry these information bits; and the others with indices $i \in \mathcal{A}^c$ are called frozen channels and are used to transmit a fix sequence, where \mathcal{A}^c is the complementary set of \mathcal{A} . The index set \mathcal{A} and its complement set \mathcal{A}^c are called the information set and the frozen set, respectively.

Different from the conventional polar codes, $N_0 - M$ output bits of the polar encoder should be punctured when constructing a punctured polar code. Therefore, before performing the polarization transform, the underlying channel uses corresponding to these punctured positions should be replaced by virtual channels which have the same input and output alphabets as W but with zero capacities [7]. As the performance of the punctured polar codes relies heavily on the specific puncturing pattern, an efficient near-optimal puncturing scheme is provided in [7].

After the punctured polar code is constructed, some error-prone information bits are duplicated and collected into a sequence of $N - M$ repetition bits. The repetition bits are directly sent into another $N - M$ uses of W , these channel uses are called the *repetition channels*. The channel transform of the RCP coding is shown in Fig.2. With a slight abuse of notation, we write $W_N^{(i)}$ with $i = 1, 2, \dots, N_0$ to denote the resulting (information and frozen) channels, where the subscript N denotes the number of actual channel uses and the superscript i denotes the channel index. Note that N and N_0 do not have to take the same value under this setting.

When constructing a RCP code, the mappings between the repetition channels and the information channels are determined one by one. Each time, one of the undetermined repetition channels and the least reliable information channel are tied together. Thus, the reliability of this information channel is improved and updated. This procedure continues until all the repetition channels are determined. The mappings between the information channels and the repetition channels are recorded in a $(N - M)$ -dimensional vector \mathbf{r} , where the k -th repetition channel carries the same bit with the

r_k -th information channel, $k = 1, 2, \dots, N - M$. The detail description is summarized in Algorithm 1, where the operation \otimes denotes the convolution of two LLR pdfs.

Algorithm 1 Determine Repetition Indices

Input: Code length of the RCP code N ;

Code length of the punctured polar code M ;

LLR pdf $\mathbf{a}_{N_0}^{(i)}$ of $W_{N_0}^{(i)}$ for all $i \in \mathcal{A}$;

LLR pdf \mathbf{a} of W when 0 is transmitted;

Output: Repetition vector \mathbf{r} ;

- 1: For every $W_N^{(i)}$, initialize its LLR pdf as $\mathbf{a}_N^{(i)} \leftarrow \mathbf{a}_{N_0}^{(i)}$;
- 2: Calculate $P_e(W_N^{(i)})$ using (1);
- 3: **for** $k \leftarrow 1 : N - M$ **do**
- 4: The index $i \in \mathcal{A}$ with the largest $P_e(W_N^{(i)})$ is found;
- 5: Determine the repetition mapping, $r_k \leftarrow i$;
- 6: Update the LLR pdf of $W_N^{(i)}$ with

$$\mathbf{a}_N^{(i)} \leftarrow \mathbf{a}_N^{(i)} \otimes \mathbf{a} \quad (2)$$

- 7: Re-calculate $P_e(W_N^{(i)})$ using (1);

8: **end for**

9: **return** \mathbf{r} ;

Since the dependencies among the information channels are not affected by adding the repetition channels, RCP codes can be decoded using a successive cancellation (SC) scheme in the same way with the conventional polar codes [1]. The block error rate (BLER) of an (N, K, M) RCP code under SC decoding can be evaluated as

$$P_B(N, K, M) = \sum_{i \in \mathcal{A}} P_e(W_N^{(i)}) \quad (3)$$

III. HARQ SCHEMES BASED ON RCP CODES

The code length of RCP codes can be easily adjusted by adding or reducing the repetition bits. Hence, the design of the HARQ scheme in Fig.1 is to construct a set of (N_t, K, M) RCP codes, where $t = 1, 2, \dots, T$. This section first gives an approximation of the throughput efficiency over BAWGNC, then the construction algorithm of the proposed HARQ scheme is described in detail.

A. An Approximate Bound of Throughput Efficiency

An HARQ scheme based on a set of (N_t, K, M) RCP codes, where $t = 1, 2, \dots, T$, is used to transmit an information block of K bits. After t transmissions, a total of N_t bits are received from the channel. Let E_t with $t = 1, 2, \dots, T$ denote the event that the information block *cannot* be correctly decoded using all the N_t bits received in the first t transmissions, and \bar{E}_t denote the complementary event of E_t . We write $\Pr(E_t)$ to denote the probability of event E_t . Particularly, we write E_0 to denote the event that the information block is not known at the receiver before transmitting any bits. Obviously, $\Pr(E_0) = 1$.

In transmitting information blocks of K -bits, the average numbers of the successfully received information bits $E[K]$ and the total transmitted bits $E[N]$ are as follows:

$$E[K] = K \cdot (1 - \Pr(E_T \cap E_{T-1} \cdots \cap E_0)) \quad (4)$$

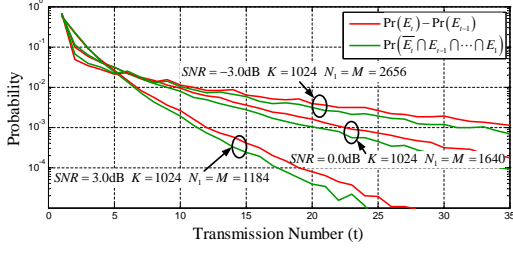


Fig. 3. $\Pr(E_{t-1}) - \Pr(E_t)$ is used to approximate $\Pr(\overline{E}_t \cap \dots \cap E_0)$.

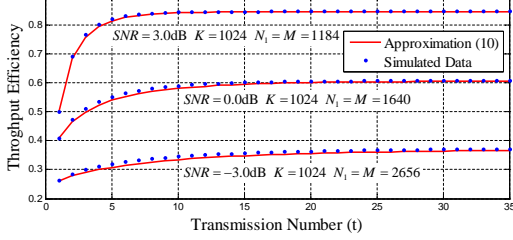


Fig. 4. The approximation (10) is usually a lowerbound.

$$E[N] = \sum_{t=1}^T N_t \cdot \Pr(\overline{E}_t \cap E_{t-1} \cap E_{t-2} \cap \dots \cap E_0) + N_T \cdot \Pr(E_T \cap E_{T-1} \cap \dots \cap E_0) \quad (5)$$

Then, the throughput efficiency can be written as

$$\eta = \frac{E[K]}{E[N]} \quad (6)$$

Obviously, we have

$$\Pr(E_t \cap \dots \cap E_0) \leq \Pr(E_t) \quad (7)$$

$$\Pr(\overline{E}_t \cap \dots \cap E_0) = \Pr(E_{t-1} \cap \dots \cap E_0) - \Pr(E_t \cap \dots \cap E_0) \quad (8)$$

However, (8) is still difficult to evaluate. In this paper, we would like to use the following approximation

$$\Pr(E_{t-1} \cap \dots \cap E_0) - \Pr(E_t \cap \dots \cap E_0) \approx \Pr(E_{t-1}) - \Pr(E_t) \quad (9)$$

Fig. 3 gives the curves of $\Pr(E_{t-1}) - \Pr(E_t)$ and $\Pr(\overline{E}_t \cap \dots \cap E_0)$ under three transmission schemes with symbol signal-to-noise ratio (SNR) of $\{-3.0, 0, 3.0\}$ dB. In each retransmission, 30 repetition bits are transmitted, i.e., $N_t - N_{t-1} = 30$, where $t \geq 2$, and each curve is obtained by simulating at least 10^6 information blocks. As shown in Fig. 3, $\Pr(E_{t-1}) - \Pr(E_t)$ tends to be an upperbound of $\Pr(\overline{E}_t \cap \dots \cap E_0)$.

Therefore, the throughput in (6) can be approximately calculated as

$$\eta \approx \frac{K \cdot (1 - \Pr(E_T))}{\sum_{t=1}^T N_t \cdot (\Pr(E_{t-1}) - \Pr(E_t)) + N_T \cdot \Pr(E_T)} \quad (10)$$

where $\Pr(E_t)$ is in fact the BLER of the (N_t, K, M) RCP code and can be evaluated by (3).

Moreover, since both the substitutions of $\Pr(E_t)$ for $\Pr(E_t \cap \dots \cap E_0)$ in (7) and $\Pr(E_{t-1}) - \Pr(E_t)$ for $\Pr(\overline{E}_t \cap \dots \cap E_0)$ in (9) are upperbounds, the approximation (10) tends to be a lowerbound of η (as shown in Fig. 4).

B. Design An HARQ Scheme

Utilizing (10), an HARQ scheme with information block size K can be designed via a greedy search described in Algorithm 2, where the function $\text{sort}(\mathcal{X})$ puts the elements of \mathcal{X} into a vector \mathbf{x} in ascending order, i.e., for all $1 \leq i \leq j \leq |\mathcal{X}|$, $x_i \in \mathcal{X}$ and $x_i \leq x_j$. The algorithm outputs a vector \mathbf{s} where its elements means that, the RCP codes are constructed with $M = s_1$, and after the t -th transmission attempt, totally s_{t+1} bits should be transmitted, i.e., $N_t = s_{t+1}$.

Algorithm 2 Design an HARQ Scheme Based on RCP codes

Input: Information block length K ;

Maximum number of transmission times T ;

Maximum transmitted bits Q ;

Output: Transmission scheme vector \mathbf{s} ;

```

1: Initialize  $\mathcal{S}_{opt} \leftarrow \phi$ ,  $\eta_{opt} \leftarrow 0$ ,  $m_{opt} \leftarrow 0$ ;
2: for  $m \leftarrow K : Q$  do
3:   // Compute the BLERs with all possible code length.
4:   Allocate a  $(Q - m + 1)$ -dimensional vector  $\mathbf{e}$ ;
5:   for  $n \leftarrow m : Q$  do
6:     Construct an  $(n, K, m)$  RCP codes;
7:      $e_{n-m+1} \leftarrow P_B(n, K, m)$ ;
8:   end for
9:   // Search for the optimal scheme with  $M = m$ .
10:   $\mathcal{S} \leftarrow \phi$ ,  $\eta \leftarrow 0$ ;
11:  for  $l \leftarrow 1 : T$  do
12:    Allocate a temporary set  $\mathcal{T} \leftarrow \mathcal{S}$ ;
13:    for  $n \leftarrow m : Q$  and  $n \notin \mathcal{S}$  do
14:       $\mathbf{t} \leftarrow \text{sort}(\mathcal{S} \cup \{n\})$ ;
15:      Calculate the throughput of the temporary scheme:
16:       $\lambda \leftarrow \sum_{i=1}^l t_i \cdot (e_{t_{i-1}-m+1} - e_{t_i-m+1})$  (11)
17:       $\rho \leftarrow \frac{K \cdot (1 - e_{t_l-m+1})}{\lambda + t_l \cdot e_{t_l-m+1}}$  (12)
18:      if  $\rho > \eta$  then
19:         $\mathcal{T} \leftarrow \mathcal{S} \cup \{n\}$ ,  $\eta \leftarrow \rho$ ;
20:      end if
21:    end for
22:    Record the locally best scheme,  $\mathcal{S} \leftarrow \mathcal{T}$ ;
23:  end for
24:  if  $\eta > \eta_{opt}$  then
25:     $\mathcal{S}_{opt} \leftarrow \mathcal{S}$ ,  $\eta_{opt} \leftarrow \eta$ ,  $m_{opt} \leftarrow m$ ;
26:  end if
27: end for
28: Allocate a  $(T + 1)$ -dimensional vector  $\mathbf{s}$ ;
29:  $s_1 \leftarrow m_{opt}$ ,  $\mathbf{s}_{2:T+1} \leftarrow \text{sort}(\mathcal{S}_{opt})$ ;
30: return  $\mathbf{s}$ ;

```

In Algorithm 2, the outer loop at line 2 is executed $Q - K + 1$ times. For each m in the loop, the operations in line 5 to 8 are equivalent to construct a punctured polar code with code length m at first, and then construct a set of RCP codes by adding the repetition channels one by one. Each time, only one additional convolution defined in (2) is required. Taking the outer loop into account, a total of $\sum_{m=K}^Q (\overline{m} \log \overline{m} + Q - m)$ convolutions are required, where $\overline{m} = 2^{\lceil \log m \rceil}$. Besides, the

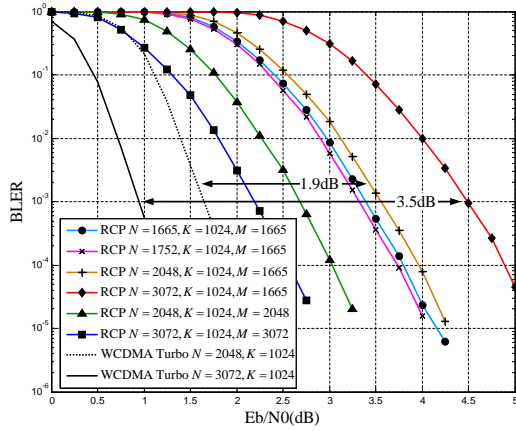


Fig. 5. Performance of RCP codes over BAWGN channels.

loops in line 11 to 21 take time complexity $O(T \cdot (Q - m))$. Since T is usually much smaller than K and Q , the operations in the loop between line 5 and 8 dominate the complexity of the whole algorithm. Therefore, the complexity of Algorithm 2 is upper bounded by $O(Q^2 \log Q)$.

IV. SIMULATION RESULTS

In this section, the performance of the proposed HARQ scheme based on RCP codes is evaluated via simulations over binary-input additive white Gaussian noise channels (BAWGNs). The RCP codes are with information block lengths $K = 1024$ and decoded using SC decoding algorithm. For efficient construction, Gaussian approximation (GA) for DE [6] is used to evaluate the error probabilities in (1).

Intuitively, RCP codes are not good codes. This can be confirmed by the simulation results over BAWGNs shown in Fig. 5. In comparison, the BLER performance of the turbo code used in WCDMA systems [8] is also given. We can see that, there is a gap of 2.0 – 3.5dB between the simulated RCP codes and the turbo codes at the BLER around 1.0×10^{-3} , and the longer the repetition sequence is the worse the performance will be. An extreme case is that when the code length N goes much larger than M , the codes look more like the naive repeat codes which do not have any coding gain.

However, when applying the RCP codes to HARQ schemes, the results are quite interesting. Fig. 6 shows the throughput efficiency of the proposed HARQ scheme over BAWGNs. In comparison, the throughput efficiency curves of the HARQ schemes based on the rate-compatible punctured turbo codes (RCPT) [3] and the rate-compatible irregular repeat-accumulate (RCIRA) codes [4] (as a representative LDPC codes) are provided. As the figure shows, the RCP coded schemes can work as well as those with turbo codes or LDPC codes and are only about 1.0 – 1.5dB from the capacity. In the low SNR regime, RCP coded schemes are a little worse than those of turbo and LDPC, the performance loss are as small as only about 0.5dB. As the SNR goes higher, the performance gap between RCP and turbo/LDPC coded schemes becomes smaller. When the SNR is above 4.0dB, the proposed scheme achieves better throughput efficiency than that of turbo/LDPC.

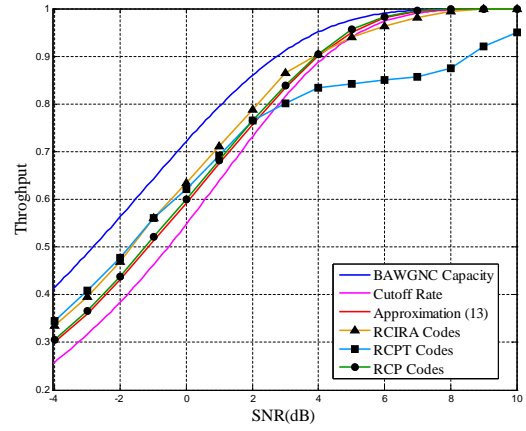


Fig. 6. Throughput of HARQ transmission scheme employing RCP codes with $K = 1024$ in BAWGN channels, where results for RCIRA codes are from [4] ($K = 512$) and RCPT codes are from [3] ($K = 1024$).

V. CONCLUSIONS

An HARQ scheme based on RCP codes is proposed. Simulation results over BAWGNs show that, the proposed transmission scheme under the SC decoding can perform as well as the existing schemes based on turbo codes and LDPC codes, and are only about 1.0-1.5dB away from the channel capacity with information block length of 1024 bits.

REFERENCES

- [1] E. Arıkan, “Channel polarization: a method for constructing capacity achieving codes for symmetric binary-input memoryless channels,” *IEEE Trans. Inf. Theory*, Vol. 55, No. 7, pp. 3051-3073, July 2009.
- [2] J. Hagenauer, “Rate-compatible punctured convolutional codes (RCPC codes) and their applications,” *IEEE Trans. Commun.*, vol. 36, no. 4, pp. 389-400, 1988.
- [3] D. N. Rowitch and L. B. Milstein, “On the performance of hybrid FEC/ARQ systems using rate compatible punctured turbo (RCPT) codes,” *IEEE Trans. Commun.*, vol. 48, no. 6, pp. 948-959, 2000.
- [4] G. Yue, X. Wang, and M. Madihian, “Design of rate-compatible irregular repeat accumulate codes,” *IEEE Trans. Commun.*, vol. 55, no. 6, pp. 1153-1163, 2007.
- [5] R. Mori and T. Tanaka, “Performance of polar codes with the construction using density evolution,” *IEEE Commun. Lett.*, vol. 13, no. 7, pp. 519-521, Jul. 2009.
- [6] P. Trifonov, “Efficient design and decoding of polar codes,” *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3221-3227, Nov. 2012.
- [7] K. Niu, K. Chen, and J. R. Lin, “Beyond turbo codes: rate-compatible punctured polar codes,” accepted for presentation in *Proc. of IEEE ICC’13*, Budapest, Hungary, June 2013.
- [8] 3GPP TS 25.212: “Multiplexing and channel coding (FDD),” Release 9, 2009.