

Lossless Polar Compression of q -ary Sources

Semih Çaycı

Department of Electrical & Electronics Engineering
Bilkent University
Ankara, Turkey
cayci@bilkent.edu.tr

Orhan Arıkan

Department of Electrical & Electronics Engineering
Bilkent University
Ankara, Turkey
oarikan@ee.bilkent.edu.tr

Abstract—In this paper, lossless polar compression of q -ary memoryless sources in the noiseless setting is investigated. Polar compression scheme for binary memoryless sources, introduced by Cronie and Korada, is generalized to sources over prime-size alphabets. In order to reduce the average codeword length, a compression scheme based on successive cancellation list decoding is proposed. Also, a specific configuration for the compression of correlated sources is considered, and it is shown that the introduced polar compression schemes achieve the corner point of the admissible rate region. Based on this result, proposed compression schemes are extended to arbitrary finite source alphabets by using a layered approach.

I. INTRODUCTION

A lossless polar compression method for binary memoryless sources in the noiseless setting is proposed in [1], and it is shown that the entropy bound is achieved for sufficiently large block-lengths in this scheme. In noiseless compression, the encoder has a copy of the codeword received by the decoder, therefore can identify where the decoder encounters errors, potentially increasing the performance of the polar compression at practical block-lengths. This property is exploited in the development of compression schemes based on LDPC codes in [2]. In [3], a lossless polar coding scheme that employs a decoder at the encoder and corrects all decoding errors at the expense of additional overhead prior to transmission is introduced for binary memoryless sources. One of the goals of the present work is to generalize this scheme to q -ary memoryless sources. In addition to adapting the scheme with conventional successive cancellation decoder (SC-D) to q -ary compression, a scheme based on the successive cancellation list decoding (SCL-D) of [4] is proposed to reduce the overhead. The compression idea is then generalized to correlated sources, and arbitrary source alphabets, respectively.

The organization of the paper is as follows. In Section II, basic source polarization concepts that will be referred in further sections are presented. The coding schemes based on SC-D and SCL-D for prime-size alphabets are introduced in Section III. Compression of sources over any arbitrary finite alphabets is discussed in Section IV. Finally, numerical results are presented in Section V.

II. SOURCE POLARIZATION

Let (X, Y) be a pair of random variables over $\mathcal{X} \times \mathcal{Y}$ with a joint distribution $p_{X,Y}(x, y)$, where $\mathcal{X} = \{0, 1, \dots, q-1\}$ for a prime number q , and \mathcal{Y} is a countable set. Following

the notation of [1], (X, Y) is considered as a memoryless source with X to be compressed, and Y to be utilized as side information in the compression of X . For a positive integer n and $N = 2^n$, let $\{(X_i, Y_i)\}_{i=0}^{N-1}$ be independent drawings from the source (X, Y) . By using the following polarization transformation:

$$\mathbf{G}_N = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes n} \mathbf{B}_N, \quad (1)$$

where all operations are performed in $GF(q)$, \otimes^n is the n^{th} Kronecker power and \mathbf{B}_N is the bit-reversal operation, the random vector X_0^{N-1} is transformed into U_0^{N-1} as:

$$U_0^{N-1} = X_0^{N-1} \mathbf{G}_N. \quad (2)$$

The input vector X_0^{N-1} is polarized by this transformation in the following sense:

$$\frac{|\{i : H(U_i|U_0^{i-1}, Y_0^{N-1}) \in [0, \delta)\}|}{2^n} = 1 - H(X|Y), \quad (3)$$

and

$$\frac{|\{i : H(U_i|U_0^{i-1}, Y_0^{N-1}) \in (1 - \delta, 1]\}|}{2^n} = H(X|Y), \quad (4)$$

for any given $\delta > 0$, and $n \rightarrow \infty$ [5]. Here the default base of the entropy function is chosen as q .

For finite-length analysis and code construction, the average minimal error probability, analyzed in [6] and [5] for polar coding, provides a more convenient measure than conditional entropy [3]. The minimal error probability, denoted by $\pi(X|Y = y)$, is the probability of error in the maximum a posteriori estimation of X given an observation $Y = y$:

$$\begin{aligned} \pi(X|Y = y) &= \Pr[X \neq \underset{x \in \mathcal{X}}{\operatorname{argmax}} p_{X|Y}(x|y)|Y = y], \\ &= 1 - \max_{x \in \mathcal{X}} p_{X|Y}(x|y). \end{aligned}$$

Therefore, the average minimal probability of error is as follows:

$$\pi(X|Y) = \sum_{y \in \mathcal{Y}} p_Y(y) \pi(X|Y = y). \quad (5)$$

$\pi(X|Y)$ has a range $[0, \frac{q-1}{q}]$ [6].

Note that the set of average error probabilities, $\{\pi(U_i|U_0^{i-1}, Y_0^{N-1})\}_{i=0}^{N-1}$, for a given source can be computed by using straightforward extensions of greedy code construction algorithms proposed in [7].

III. CODING SCHEME

In order to compress a sequence $\{(X_i, Y_i)\}_{i=0}^{N-1}$, an information set $\mathcal{I}_{X|Y}(N, R)$ consisting of indices i that correspond to NR highest $\pi(U_i|U_0^{i-1}, Y_0^{N-1})$ terms is constructed. Then, a given realization x_0^{N-1} is transformed into u_0^{N-1} by (2) and the compressed word $u_{\mathcal{I}_{X|Y}}$ is formed. For sufficiently large N , this scheme is proved to achieve arbitrarily small probability of error under conventional SC-D with codeword length $NH(X|Y)$ [1]. In [3], for binary memoryless sources, an oracle-based polar compression method that has an improved performance at practical block-lengths is introduced. Here, a similar approach is taken in the design of q -ary compression methods. The methods are based on appending a block to the compressed word $u_{\mathcal{I}_{X|Y}}$ indicating the locations of error that will be encountered in decoding, and correcting them. This block enables zero-error coding at any block-length. Moreover, it is shown that this extra block has a diminishing fraction in the transmitted word, which means that the entropy bound is still achievable asymptotically.

A. Encoding

In noiseless source coding, the encoder has a copy of the codeword received by the decoder. This specific property enables the encoder to run the decoder at the transmitter side and check if a decoding error occurs. In polar compression, this capability can be utilized to prevent any errors by appending a variable length block of error positions and their correct symbols to the codeword; thus fixed-to-variable length, zero-error coding schemes can be designed.

The encoding is specific to the type of decoder. Therefore, we will consider schemes with SC-D and SCL-D separately. First, let us consider the encoding in the case of SC-D, which is a straightforward extension of [3].

For a given source realization x_0^{N-1} , the encoder forms the codeword $u_{\mathcal{I}_{X|Y}}$ and conveys it to the mirror SC-D at the transmitter side. If an error occurs at phase i , the encoder interferes, records the error location together with the respective correct symbol, (i, u_i) , corrects the error and resumes the decoding process. Following this routine, the encoder records the set of all error locations together with the respective correct symbols:

$$\mathcal{T}_{SC} = \{(i, u_i) : u_i \neq \hat{u}_i|u_0^{i-1}, y_0^{N-1}\}.$$

Then, the encoder appends \mathcal{T}_{SC} to the codeword $u_{\mathcal{I}_{X|Y}}$ and transmits $(u_{\mathcal{I}_{X|Y}}, \mathcal{T}_{SC})$ to the receiver side. Having the error locations and their respective correct symbols, the decoder at the receiver side performs decompression with no error. Note that if $q = 2$, there is no need to record u_i since knowing the location of an error is sufficient to correct it through inversion. In the rest of the discussion, a general q will be considered, and the correct symbol value will be included in the oracle.

Given a correctly decoded subsequence \hat{u}_0^{i-1} and observation y_0^{N-1} , the probability of error at phase i of SC-D is $\pi(U_i|\hat{u}_0^{i-1}, y_0^{N-1})$. Thus, the average probability of error at phase i is $\pi(U_i|U_0^{i-1}, Y_0^{N-1})$. If an error occurs at phase i , it

costs an additional overhead of $(\log N + 1)$ symbols. Therefore, the average cost of not including i in the information set is $\pi(U_i|U_0^{i-1}, Y_0^{N-1})[\log N + 1]$ symbols. The cost of including i in the information set is 1 symbol. Combining these results, the expected code rate R is as follows:

$$\mathbb{E}[R] = \frac{1}{N} \{|\mathcal{I}_{X|Y}| + \sum_{i \in \mathcal{I}_{X|Y}^c} \pi(U_i|U_0^{i-1}, Y_0^{N-1})[\log N + 1]\}. \quad (6)$$

This analysis can be used in the construction of $\mathcal{I}_{X|Y}$ as well [3]. The objective is to minimize the expected codeword length over all information sets. If the average cost of including an index i in $\mathcal{I}_{X|Y}^c$ is higher than including it in $\mathcal{I}_{X|Y}$, then the symbol is transmitted in $u_{\mathcal{I}_{X|Y}}$. The information set is formed as follows:

$$\mathcal{I}_{X|Y} = \{i : \pi(U_i|U_0^{i-1}, Y_0^{N-1})[\log N + 1] > 1\}. \quad (7)$$

For sufficiently large N , $\mathcal{I}_{X|Y}$ consists of indices such that $\pi(U_i|U_0^{i-1}, Y_0^{N-1}) \in (\frac{q-1}{q} - \epsilon, \frac{q-1}{q}]$, and the length of $\mathcal{I}_{X|Y}$ approaches $NH(X|Y)$. Therefore, the expected rate achieves the entropy bound asymptotically:

$$\mathbb{E}[R] \rightarrow H(X|Y).$$

Hence, this zero-error compression scheme designed for finite block-lengths achieves the theoretical bound asymptotically as well.

If an incorrect decision is made by SC-D at any phase, a block error is flagged, and this causes additional overhead because of the oracle employment. Successive cancellation list decoder is likely to correct an incorrect decision at succeeding phases at the expense of increased complexity. In noiseless source coding, this property of SCL-D can be utilized to reduce codeword length. Consider an SCL-D of list size L at phase $i \notin \mathcal{I}_{X|Y}$. Assume that the correct decoding path $\hat{u}_0^{i-1} = u_0^{i-1}$ is contained among the active paths. At phase i , all symbols in \mathcal{X} is appended to each active path, and all paths are pruned keeping L of the highest probability values. Denoting the set of all active paths at phase i by \mathcal{L}_i , an error is flagged if the correct subsequence u_0^i is not in \mathcal{L}_i . If such an event occurs, the encoder interferes, takes a record of (i, u_i) and appends u_i to each active path as if i is contained in the information set. Eventually, the oracle set is formed as follows:

$$\mathcal{T}_{SCL} = \{(i, u_i) : u_0^i \notin \mathcal{L}_i|u_0^{i-1} \in \mathcal{L}_{i-1}, y_0^{N-1}\}. \quad (8)$$

The employment of this oracle set guarantees the correct decoding path u_0^{N-1} to survive until the end. In the last phase, SCL-D returns the sequence among \mathcal{L}_{N-1} with highest probability. An incorrect sequence is returned if there is a path $\tilde{u}_0^{N-1} \in \mathcal{L}_{N-1}$ with higher probability than u_0^{N-1} . In order to prevent this error, the list index l of the correct sequence can be annexed to the codeword. This increases the codeword length by $\log L$ symbols. On the other hand, the overhead due to the usage of oracle can be decreased by more than this overhead. Thus, SCL-D provides lower rates than SC-D in general.

B. Decoding

For a given source (X, Y) and observation y_0^{N-1} , the probability of observing u_0^{N-1} at the output of the polarization transform is denoted as $P_N(u_0^{N-1}|y_0^{N-1})$, where $P_1(x|y) = p_X(x|y)$. Similarly, the probability of a subsequence u_0^i is denoted as $P_N^{(i)}(u_0^{i-1}, u_i|y_0^{N-1})$.

The SC-D algorithm is summarized in Algorithm 1.

Algorithm 1: SC_Decoder($u_{\mathcal{I}_{X|Y}}, \mathcal{T}_{SC}$)

input : $u_{\mathcal{I}_{X|Y}}$: Codeword, \mathcal{T}_{SC} : Oracle set

output: \hat{x}_0^{N-1} : Reconstructed sequence

```

1 for  $i = 0, 1, \dots, N-1$  do
2   if  $i \in \mathcal{I}_{X|Y}$  or  $(i, u_i) \in \mathcal{T}_{SC}$  then
3      $\hat{u}_i = u_i$ 
4   else
5      $\hat{u}_i = \operatorname{argmax}_{u_i \in \mathcal{X}} P_N^{(i)}(\hat{u}_0^{i-1}, u_i|y_0^{N-1})$ ;
6 Return  $\hat{x}_0^{N-1} = \hat{u}_0^{N-1} \mathbf{G}_N^{-1}$ .
```

The SC-D algorithm can be implemented with $O(N)$ memory and $O(N \log N)$ time complexity [4].

The high-level description of SCL-D is given in Algorithm 2.

Algorithm 2: SCL_Decoder($u_{\mathcal{I}_{X|Y}}, \mathcal{T}_{SCL}, l_0, L$)

input : $u_{\mathcal{I}_{X|Y}}$: Codeword, \mathcal{T}_{SCL} : Oracle set, l_0 : Index of the correct decision path, L : List size

output: \hat{x}_0^{N-1} : Reconstructed sequence

```

1 for  $i = 0, 1, \dots, N-1$  do
2   if  $i \in \mathcal{I}_{X|Y}$  or  $(i, u_i) \in \mathcal{T}_{SCL}$  then
3     Append  $u_i$  to each  $\hat{u}_0^{i-1}[l] \in \mathcal{L}_{i-1}$ , and obtain
        $(\hat{u}_0^{i-1}[l], u_i)$ 
4   else
5     Append all  $\hat{u}_i \in \mathcal{X}$  to each  $\hat{u}_0^{i-1}[l] \in \mathcal{L}_{i-1}$ ;
6     Calculate  $P_N^{(i)}(\hat{u}_0^{i-1}[l], \hat{u}_i|y_0^{N-1})$  for all
        $(\hat{u}_0^{i-1}[l], \hat{u}_i) \in \mathcal{L}_i$ ;
7     Prune all but  $L$  paths with highest probabilities.
8 Return  $\hat{x}_0^{N-1} = \hat{u}_0^{N-1}[l_0] \mathbf{G}_N^{-1}$ .
```

The time complexity of the SCL-D algorithm is $O(LN \log N)$ [4]. Note that SCL-D with list size $L = 1$ corresponds to SC-D. The encoding operation has a computational complexity of $O(N \log N)$. Hence, the overall complexity of the compression schemes is still $O(LN \log N)$ with $L = 1$ for SC-D.

IV. COMPRESSION OF SOURCES OVER ARBITRARY FINITE ALPHABETS

In this section, we generalize the ideas derived in Section III to sources over any arbitrary finite alphabets. In order to realize this, we first consider a specific configuration for the noiseless compression of two correlated sources (X, Y) . In

this scenario, the source output Y_0^{N-1} is available to the X -encoder, the decompressed word \hat{Y}_0^{N-1} is available to the X -decoder, and neither X_0^{N-1} nor \hat{X}_0^{N-1} is used in the compression of Y . The scheme is illustrated in Figure 1.

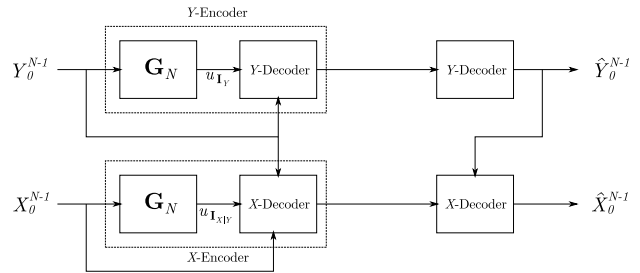


Fig. 1. (0101)-configuration for the compression of (X, Y) .

This configuration is analyzed in [8], where it is called (0101)-configuration. For all $\epsilon_X, \epsilon_Y > 0$, it is possible to achieve rates $R_Y = H(Y) + \epsilon_Y$ and $R_X = H(X|Y) + \epsilon_X$ for Y and X , respectively, and (R_X, R_Y) is referred to as the corner point of the admissible region.

Lemma 1. *The SC-D and SCL-D compression schemes achieve the corner point of the admissible region for (0101)-configuration.*

Proof. In order to compress Y , the compression is performed with no side information. The compression rate R_Y asymptotically achieves $H(Y)$. Since this is a zero-error coding scheme, the Y -source output is reconstructed faithfully at the receiver side. In order to compress X , the compression schemes are used with side information Y . Note that Y -source output is available at both transmitter and receiver sides with no error. Thus, X can be compressed at rate $H(X|Y)$ asymptotically, and the corner point of (0101)-configuration is achieved. \square

An extension of this configuration is the noiseless source coding over arbitrary finite alphabets, using a similar approach as in [5]. Let Z be a random variable over a finite alphabet \mathcal{Z} . Z can be decomposed into K symbols using the Chinese remainder theorem as:

$$Z = (Z^{K-1}, Z^{K-2}, \dots, Z^0),$$

where Z^k is over \mathcal{Z}_k , provided that $|\mathcal{Z}_k| = q_k$ and all q_k are pairwise coprime. Note that q_k can be an integer power of a prime, in which a further expansion can be carried out to obtain prime alphabet sizes for compression, and the result can be used to uniquely reconstruct Z . Hence, without loss of generality, it can be assumed in further discussions that all q_k are prime.

At the first step of compressing Z , Z^0 is compressed with no side information, analogous to Y in the previous case, at rate $R_{Z_0} \rightarrow H(Z^0)$. Then, Z^1 is compressed with side information Z^0 at rate $R_{Z_1} \rightarrow H(Z^1|Z^0)$. Now that the source outputs of (Z^1, Z^0) are transmitted, they are utilized as side information and the compression of Z^2 is performed at rate $R_{Z_2} \rightarrow H(Z^2|Z^1, Z^0)$. Following this routine, Z^k can

be compressed at rate $R_{Z^k} \rightarrow H(Z^k|Z^{k-1}, \dots, Z^0)$ for any $k = 0, 1, \dots, K - 1$. After the decompression of Z^{K-1} , Z can be reconstructed faithfully. The total compression in this scheme has the following asymptotical rate:

$$\begin{aligned} R_Z &= \sum_{k=0}^{K-1} R_{Z^k} \\ &\rightarrow \sum_{k=0}^{K-1} H(Z^k|Z^{k-1}, \dots, Z^0) \\ &= H(Z^{K-1}, Z^{K-2}, \dots, Z^0) = H(Z), \end{aligned}$$

which shows that the entropy bound can be achieved by the proposed q -ary polar compression scheme.

V. NUMERICAL RESULTS

In this section, we provide compression rates observed as the average of 10000 Monte-Carlo trials. In Figure 2, the average compression rates for ternary sources with probability mass functions $p_1 = (0.1, 0.275, 0.625)$, $p_2 = (0.07, 0.09, 0.84)$ and $p_3 = (0.9214, 0.0393, 0.0393)$ in the absence of side information are presented at various block-lengths. Base-3 entropy values are marked by lines. The coding scheme based on SC-D provides good performance at practical block-lengths.

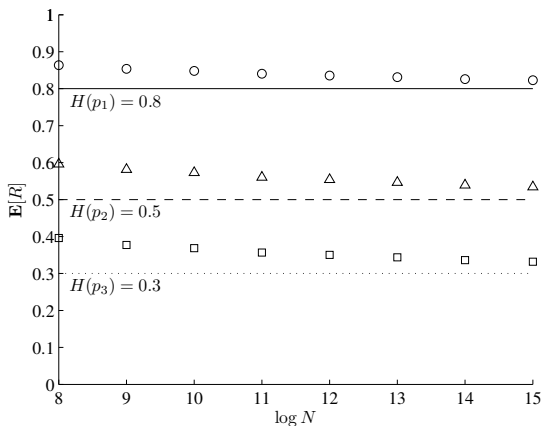


Fig. 2. Average compression rates for ternary sources under SC-D.

In Figure 3, the performance of the SCL-D based scheme is investigated for the ternary source with probability distribution p_2 , and the change in the average compression rate with respect to the list size L is presented. The average code rate decreases with increasing list size.

In Figure 4, the performance of polar compression for a 6-ary source with probability distribution $p_Z = (0.0077, 0.7476, 0.0675, 0.0623, 0.0924, 0.0225)$ under SC-D is presented. The source Z is compressed in two layers, i.e., $Z = (X, Y)$, where Y is a ternary and X is a binary random variable. This example indicates that polar compression framework can be utilized in the compression of sources over arbitrary finite alphabets using the proposed layered approach.

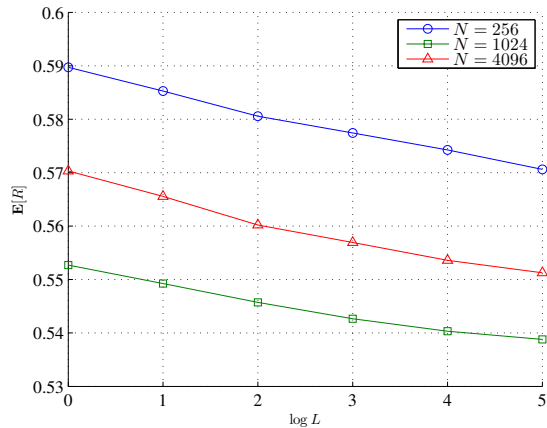


Fig. 3. Average compression rates for a source with probability distribution $p_2 = (0.07, 0.09, 0.84)$ under SCL-D with various list sizes L .

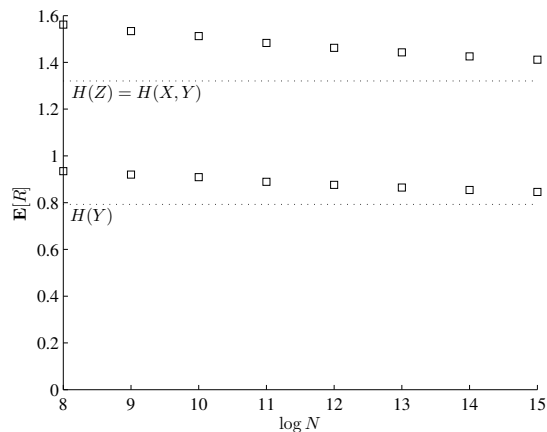


Fig. 4. Average compression rates for a 6-ary source $Z = (X, Y)$. Base-2 entropy values are marked by dotted lines.

VI. CONCLUSION

A lossless polar compression scheme for q -ary sources that has a good performance at finite block-lengths and achieves the entropy bound asymptotically is proposed. To improve the performance, an SCL-D based scheme is proposed, and it is shown numerically that lower compression rates at an acceptable computational load can be achieved by this scheme. Based on the compression scheme for correlated sources, a layered approach for the compression of sources over arbitrary finite alphabets is developed. In all cases, simulation results show that polar compression achieves rates close to the entropy bound with low complexity encoding and decoding algorithms.

ACKNOWLEDGEMENT

We would like to thank Prof. Erdal Arıkan for his support and motivation on this research. This work was supported by The Scientific and Technological Research Council of Turkey (TÜBİTAK) under contract no. 110E243.

REFERENCES

- [1] E. Arıkan, "Source polarization", *2010 IEEE International Symposium on Information Theory Proceedings (ISIT)*, pp.899-903, 13-18 June 2010.
- [2] G. Caire, S. Shamai, and S. Verdú, "Noiseless data compression with low density parity check codes," in *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, P. Gupta and G. Kramer, Eds., pp. vol. 66, pp. 263-284. American Mathematical Society, 2004.
- [3] H. S. Cronie, S. B. Korada, "Lossless source coding with polar codes", *2010 IEEE International Symposium on Information Theory Proceedings (ISIT)*, pp.904-908, 13-18 June 2010.
- [4] I. Tal, A. Vardy, "List decoding of polar codes", *2011 IEEE International Symposium on Information Theory Proceedings (ISIT)*, pp.1-5, July 31 2011-Aug. 5 2011.
- [5] E. Şaşıođlu, "Polar Coding Theorems for Discrete Systems", Ph.D. dissertation, Computer, Communication and Information Sciences, EPFL, Lausanne, Switzerland, 2011.
- [6] M. Feder, N. Merhav, "Relations between entropy and error probability", *IEEE Transactions on Information Theory*, vol.40, no.1, pp.259-266, Jan 1994.
- [7] I. Tal, A. Vardy, "How to Construct Polar Codes", <http://arxiv.org/abs/1105.6164>
- [8] D. Slepian, J. K. Wolf, "Noiseless coding of correlated information sources", *IEEE Transactions on information Theory*, vol.19, no.4, pp. 471- 480, Jul 1973.