International Journal of Innovative Computing, Information and Control
Volume **10**, Number **6**, December **2014**

ICIC International ©2014 ISSN 1349-4198
pp. 2051–2065

CHANNEL POLARIZATION AND RESEARCH, DEVELOPMENT OF POLAR CODES

Guiping Li

School of Engineering and Technology
Xi'an Fanyi University
Taiyigong, Chang'an District, Xi'an 710105, P. R. China
15693685@qq.com

Received December 2013; revised April 2014

Abstract. *Although the modern coding technologies such as Turbo codes and low-density parity-check (LDPC) codes have better correcting error performance, they do not have enough theory basic on encoding and decoding. While polar codes have rich algebra structures and excellent analysis properties because of using sequence mutual information stream rules. In this paper, we summarize fundamentals of polarization and analyze the reason that the polar codes can achieve the optimum performance, and also list some new research direction of polar codes. These analyses and conclusions can provide theory basis and new train of thought on current and proposed work.*

Keywords: Polar codes, Channel polarization, Successive cancellation decoding

1. **Introduction.** Constructing of coding systems which performance are close to Shannon limit with low complexities of encoding and decoding algorithms is the central topic in academic circles of the channel coding for many years. Polar coding is a new technique introduced by Arikan [1] based on a phenomenon called channel polarization. Polar codes are the first known codes that theoretically achieve the symmetric capacities of the binary-input discrete memoryless channels (B-DMCs) with an explicit construction method and low-complexity encoding and decoding algorithms. The well-established LDPC and Turbo codes are also capacity-achieving codes with low-complexity encoding and decoding algorithms, however, there is no rigorous proof of this fact for arbitrary B-DMCs. Despite these nice asymptotical properties of polar codes on complexity and performance, it is not yet clear if polar codes will have an impact on the practice of error-correction coding, where the performance at nonasymptotic regime is important. Another limitation of polar codes to date is that their performance at short to moderate block lengths is disappointing. There are two possible culprits: the codes themselves are inherently weak at these lengths, or the successive decoder employed to decode them is significantly degraded with respect to maximum likelihood decoding performance. These two possibilities are complementary, and so both may occur.

In this paper, we first summary fundamentals of polarization, the construction method of polar codes, the decoding technique using successive cancellation (SC) algorithm and the reason that the polar codes can achieve the optimum performance. Then we review and analyses the existing work on polar codes, and also we present some new research direction of polar codes. These analyses and conclusions will provide theory basis and new train of thought on current and proposed work.

2. **Channel Polarization and Polar Codes.** Let W denote a B-DMC with input alphabet $X = \{0, 1\}$, arbitrary output alphabet Y, and transition probabilities $W(y|x)$, $x \in X$, $y \in Y$. Let W^N denote N independent uses of W.

2.1. **Two important parameters.** In [1], there are two channel parameters of symmetric B-DMCs which are defined as: the mutual information and the Bhattacharyya parameter.

Definition 2.1. *The mutual information of a B-DMC with input alphabet* $X = 0, 1$ *is defined as*

$$I(W) = \sum_{y \in Y} \sum_{x \in X} \frac{1}{2} W(y|x) \log \frac{W(y|x)}{\frac{1}{2}W(y|0) + \frac{1}{2}W(y|1)} \tag{1}$$

Since that the capacity of a symmetric B-DMC equals the mutual information between the input and output of the channel with uniform distribution on the inputs. $I(W)$ is a measure of rate in a channel. It is well-known that reliable communication is possible over a symmetric B-DMC at any rates up to $I(W)$.

Definition 2.2. *The Bhattacharyya parameter of a channel is defined as*

$$Z(W) = \sum_{y \in Y} \sqrt{W(y|0)W(y|1)}, \tag{2}$$

The Bhattacharyya parameter is a measure of the reliability of a channel since $Z(W)$ is an upper bound on the probability of maximum-likelihood (ML) decision error for uncoded transmission over W.

Furthermore, note that the relation between $I(W)$ and $Z(W)$ is quantified as follows.

Proposition 2.1. *For any B-DMC* W, *we have*

$$I(W) \geq \log \frac{2}{1 + Z(W)} \tag{3}$$

$$I(W) \leq \sqrt{1 - Z(W)^2}. \tag{4}$$

From Proposition 2.1, it can be seen that the smaller value of the parameter $Z(W)$, the more reliable of the channel is. Furthermore, these two parameters are used jointly to prove that channel polarization occurs.

2.2. **Channel polarization.** Channel polarization is an operation which produces N channels $\{W_N^{(i)} : 1 \leq i \leq N\}$ from N independent copies of a B-DMC W such that the new parallel channels are polarized in the sense that symmetric capacities approach the poles of capacity limits, i.e., either go to 0 (completely noisy channels) or 1 (perfectly noiseless channels). The channel polarization operation consists of two phases: channel combing and channel splitting.

(1) ***Channel Combining***

In this phases, copies of a B-DMC are combined in a recursive manner in n steps to form a vector channel W_N, where $N = 2^n$. The basic transformation used in channel combining is shown in Figure 1. Two individual channels W are combined to create a new super channel W_2 with inputs u_1, u_2, outputs y_1, y_2, and transition probabilities

$$W_2(y_1, y_2|u_1, u_2) = W(y_1|x_1)W(y_2|x_2) \tag{5}$$

$$= W(y_1|u_1 \oplus u_2)W(y_2|u_2). \tag{6}$$

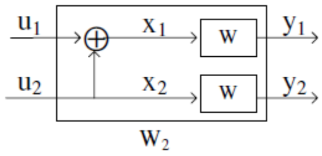

FIGURE 1. Basic channel transformations

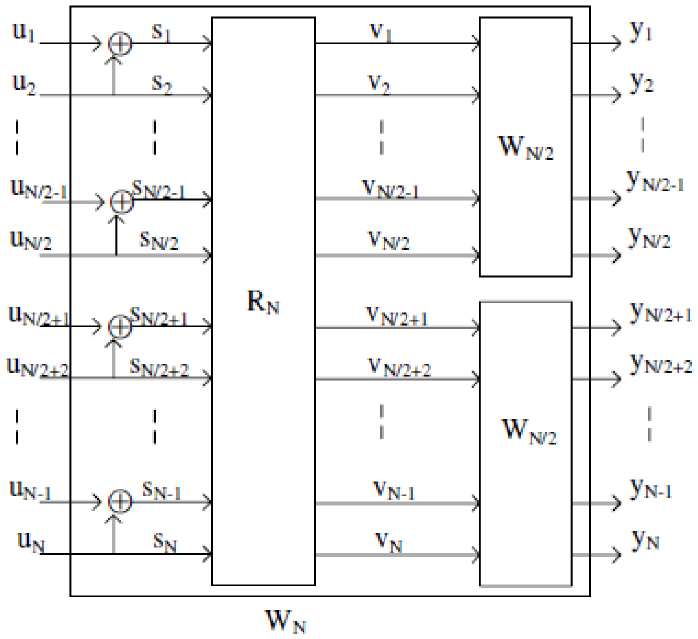

FIGURE 2. Recursive construction of W_N from two copies of $W_{N/2}$

Since the linear transform between (U_1, U_2) and (X_1, X_2) is a one-to-one mapping the following equality holds

$$I(U_1, U_2; Y_1, Y_2) = I(X_1, X_2; Y_1, Y_2) = 2I(W) \tag{7}$$

It is possible to combine two independent copies of W_2 in a similar way to generate a new channel W_4, two copies of W_4 to generate W_8, and so on. For any $N = 2^n$, $n \geq 1$, in general, the channel combining operation can be expressed as a recursive transformation as shown in Figure 2 that yields a channel W_N from two independent copies of $W_{N/2}$. The channel $W_N : \chi^N \to y^N$ is defined as

$$\begin{aligned} W_N(Y^N|U^N) &= W_{N/2}(Y^{N/2}|U_o^N \oplus U_e^N) \\ &= W_{N/2}(Y_{N/2+1}^N|U_e^N)W_{N/2}(y_{N/2}^N|\nu_{N/2}^N) \\ &= W_{N/2}(y_1^{N/2}|u_{1,e}^N \oplus u_{1,o}^N)W_{N/2}(y_{N/2+1}^N|u_{1,o}^N). \end{aligned} \tag{8}$$

The input vector u_1^N to W_N is first transformed into s_1^N with $s_{2i-1} = u_{2i-1} \oplus u_{2i}$ and $s_{2i} = u_{2i}$, $i = 1, \ldots, N/2$, which is further transformed into $v_1^N = (s_{1,o}^N, s_{1,e}^N)$. The vector v_1^N becomes the input to the two copies of $W_{N/2}$. For more details please refer to [1].

(2) ***Channel Splitting***

In the second phase, the vector channel W_N is split back into N channels $W_N^{(i)} : \{0,1\} \to y^N \times \{0,1\}^{i-1}$, $1 \leq i \leq N$. In the basic case of $N = 2$, by using chain rule of mutual information, the left hand side of Equation (7) can be written as

$$I(U_1, U_2; Y_1, Y_2) = I(U_1; Y_1, Y_2) + I(U_1; Y_1, Y_2, U_1) \tag{9}$$

where $I(U_1; Y_1, Y_2)$ can be seen as the mutual information of the channel between U_1 and Y_1, Y_2. $I(U_2; Y_1, Y_2, U_1)$ is the mutual information of the channel between U_2 and the output given that U_1 is known. Let W^- and W^+ denote the two channels respectively. The transition probabilities of W^+ and W^- can be written as

$$W^-(y_1, y_2|u_1) = \frac{1}{2} \sum_{u_2 \in \{0,1\}} W(y_1|u_1 \oplus u_2) W(y_2|u_2) \tag{10}$$

and

$$W^+(y_1, y_2, u_1|u_2) = \frac{1}{2} W(y_1|u_1 \oplus u_2) W(y_2|u_2) \tag{11}$$

In other words, the combined channel W_2 is split into two channels W^+ and W^-. Furthermore, the following two properties [1] for these channels created by the above operations hold

$$I(W^+) + I(W^-) = 2I(W), \tag{12}$$

$$Z(W^-) \le 2Z(W) - Z(W)^2, \quad Z(W^+) = Z(W)^2. \tag{13}$$

In general, we can obtain N channels $W^{s_1, s_2, \ldots, s_n}$, $s_i \in -, +$, $1 \le i \le n$ by splitting the channel W_N. Alternatively, we represent these channels by $W_N^{(i)}$, $1 \le i \le N$. The splitting operation is defined through the chain rule

$$I(U^N; Y^N) = \sum_{i=1}^{N} I(U_i; Y_1^N, U_1^{i-1}). \tag{14}$$

The term $I(U_i; Y_1^N, U_1^{i-1})$ corresponds to the channel between U_i and (Y_1^N, U_1^{i-1}). The transition probabilistic of these channels by the above splitting operation are given by

$$W_N^{(i)}(y_1^N, u_1^{i-1}|u_i) = \frac{1}{2^{N-1}} \sum_{u_{i+1}^N \in (0,1)^{N-i}} W_N(y_1^N|u_1^N). \tag{15}$$

To analyze the behavior of these channels, a random process to represent them is defined in [1]. Let $B_n : n \ge 1$ be a sequence of i.i.d. Bernoulli random variables with probabilities equal to $\frac{1}{2}$, and $F_n, n \ge 1$ be the σ-algebra generated by B^n. Define a tree process as the following

$$W_{n+1} = \begin{cases} W_n^- & \text{if } B_n = 0, \\ W_n^+ & \text{if } B_n = 1. \end{cases} \tag{16}$$

In [1], the behavior and properties of the random processes $I(W_n)$ and $Z(W_n)$ are shown that

A.1 The sequence $I_n, F_n, n \ge 0$ is a bounded martingale.
A.2 The sequence $Z_n, F_n, n \ge 0$ is a bounded super-martingale.
A.3 The sequence I_n converges almost surely to a random variable I_∞ and

$$I_\infty = \begin{cases} 1 & w.p. I(W), \\ 0 & w.p. 1 - I(W). \end{cases} \tag{17}$$

A.4 The sequence Z_n converges almost surely to a random variable Z_∞ and

$$Z_\infty = \begin{cases} 1 & w.p. 1 - I(W), \\ 0 & w.p. I(W). \end{cases} \tag{18}$$

This implies that as the length gets larger almost all the channels $W_N^{(i)}$ get polarized to either clean channels or noisy channels. For obvious reasons, the channels with mutual information close to 1 are good channels and the remaining channels are bad channels. Hence, we can obtain the following proposition.

Proposition 2.2. (Theorem 1 in [1]). *For any B-DMC W, the channels $W_N^{(i)}$ polarize in the sense that, for any fixed $\delta \in (0,1)$, as N goes to infinity through powers of two, the fraction of indices $i \in \{1,\ldots,N\}$ for which $I(W_N^{(i)}) \in (1-\delta,1)$ goes to $I(W)$ and the fraction for which $I(W_N^{(i)}) \in [0,\delta)$ goes to $1-I(W)$.*

This is illustrated in Figure 3 for a binary erasure channel (BEC) with erasure probability $\epsilon = 0.5$.

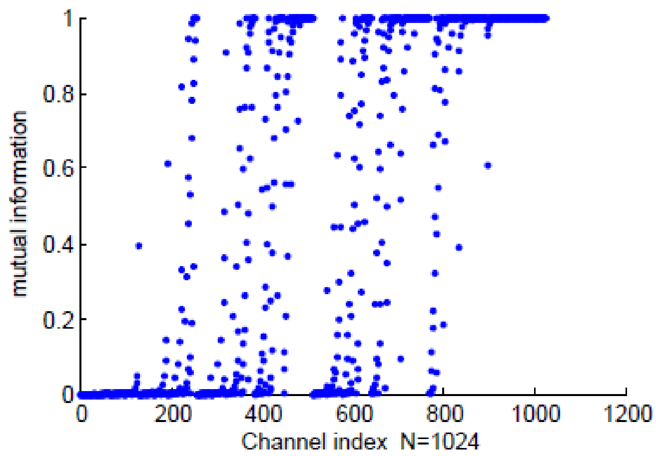

FIGURE 3. The polarization of effect with $N = 1024$ and $\epsilon = 0.5$ over BEC

2.3. **Polar codes.** Using the idea of channel polarization, polar codes are proposed as a new coding technique. Namely, a polar code (N,R) of block length N and rate R send information bits only through the $\lfloor NR \rfloor$ channels for which mutual information $I(W_N^{(i)})$ are near 1.

(1) ***Polar Encoder***

Let G_N denote the generator matrix for the channel combing operation of order N

$$G_N = B_N F^{\otimes n} \tag{19}$$

where

$$F = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \tag{20}$$

and $F^{\otimes n}$ denotes the n-fold Kronecker product of F with itself, B_N is a bit-reversal permutation matrix, and all operations are performed over GF(2). Let the vector x_1^N denote the input sequence of the N independent copies of the initial channel W. The input sequence x_1^N is computed from

$$x_1^N = u_1^N G_N. \tag{21}$$

Depending on the polarized direction of the channels, indices of the data sequence $u_1^N = u_1,\ldots,u_N$ are split into two sets before transmission. The first one includes the indices of the data to be transmitted on the good channels, and is referred as the information set A. The remaining one is the set A^C of which indices are corresponding to the frozen bits to be transmitted on the bad channels. Let $G_N(A)$ denote the matrix constructed by taking the rows of G_N whose indices are in A. Then Equation (21) can also be written as

$$x_1^N = u_A G_N(A) \oplus u_{A^c} G_N(A^c). \tag{22}$$

The above encoding operation can be illustrated in Figure 4.

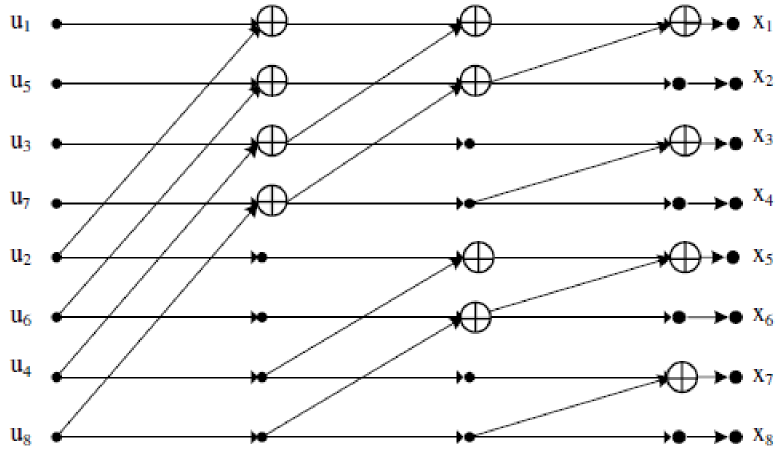

FIGURE 4. Encoding operation of the 8-bit polar code

(2) ***Polar Decoder***

Arikan considered successive cancellation (SC) decoding in order to achieve capacity with low complexity. Let the row vector u_1^N denote the input data sequence. In SC decoding, decoding results for the non-information bits are set to 0. The information bits are decoded sequentially in the ascending order of their indices. More precisely, the SC decoder generates an estimate $\hat{u}_1^N$ of u_1^N by observing the channel output y_1^N. The decoder takes N decisions for each u_i. If u_i is a frozen bit, the decoder will set $\hat{u}_i$ to zero. If u_i is an information bit, the decoder computes the following likelihood ratio (LR) after estimating all the previous bits u_1^{i-1}

$$L_N^{(i)}(y_1^N, \hat{u}_1^{i-1}) = \frac{W_N^{(i)}(y_1^N, \hat{u}_1^{i-1}|0)}{W_N^{(i)}(y_1^N, \hat{u}_1^{i-1}|1)}. \tag{23}$$

and generates its decision as

$$\hat{u}_i = \begin{cases} 0, & \text{if } L_N^{(i)}(y_1^N, \hat{u}_1^{i-1}) \geq 1 \\ 1, & \text{otherwise.} \end{cases} \tag{24}$$

which is then sent to all succeeding decision elements. The SC decoding algorithm successively evaluates the LR value of each bit $\hat{u}_i$. Arikan showed that these LR computations can be efficiently performed in a recursive manner by using a data flow graph which resembles the structure of a fast Fourier transform. That structure, shown in Figure 5, is

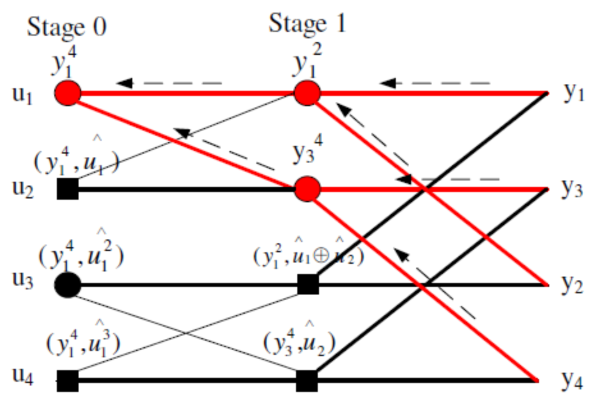

FIGURE 5. SC decoding process of polar codes with length $N = 8$

named a butterfly based decoder. Messages passed in the decoder are LR values. It can be seen that this is a single-pass message passing algorithm, with no revision of estimates. This makes it analytically tractable and vital for theoretical proofs. However, this decoding method also leads to error propagation and possibly cause the mediocre finite-length performance of polar codes.

Let $P_e(N, K, A)$ denote the block error probability when using the (N, K, A) code under SC decoding, averaged over all choices of u_{A^c}. Using SC algorithm, Arikan established the following results on the block error probability of polar codes.

Proposition 2.3. (Proposition 2 in [1]). *For any B-DMC W and block length N the block error probability averaged over a uniform distribution of the frozen symbols satisfies*

$$P(\hat{U} \neq U) \leq \sum_{i \in A} Z(W_N^{(i)}). \tag{25}$$

Proposition 2.3 gives a hint as to how A should be selected. The construction of polar codes are by choosing A in a way that minimizes the above sum. In, principle, $Z(W_N^{(i)})$ can be calculated explicitly. In practice, this calculation has forbiddingly high complexity except for BEC.

2.4. **Encoding and decoding time complexity.** The recursive structure of the channel polarization construction leads to low-complexity encoding and decoding algorithms for polar codes.

Let $X_E(N)$ denote the encoding complexity for block-length N. The recursive channel combining operation shown in Figure 2 implies that

$$X_E(N) = N/2 + 2X_E(N/2). \tag{26}$$

The above recursive relation implies

$$\begin{aligned} X_E(N) &= N/2 + 2X_E(N/2) \\ &= N/2 + 2(N/4 + 2X_E(N/4)) \\ &= N/2 + 2(N/4 + 2(N/8 + 2X_E(N/8))) \\ &= \ldots = N/2 \log N \end{aligned} \tag{27}$$

Therefore, the encoding complexity is $O(N \log N)$. While the complexity of SC decoding is determined by the complexity of computing the LRs. The recursive relations for computing the LRs are

$$L_N^{(2i)}\left(y_1^N, \hat{u}_1^{2i-1}\right) = \frac{1 + L_{N/2}^{(i)}\left(y_1^{N/2-1}, \hat{u}_{1,e}^{2i-1} \oplus \hat{u}_{1,o}^{2i-1}\right) L_{N/2}^{(i)}\left(y_{N/2}^N, \hat{u}_{1,o}^{2i-1}\right)}{L_{N/2}^{(i)}(y_1^{N/2-1}, \hat{u}_{1,e}^{2i-1} \oplus \hat{u}_{1,o}^{2i-1}) + L_{N/2}^{(i)}\left(y_{N/2}^N, \hat{u}_{1,o}^{2i-1}\right)}, \tag{28}$$

$$L_N^{(2i+1)}\left(y_1^N, \hat{u}_1^{2i}\right) = L_{N/2}^{(i)}\left(y_1^{N/2-1}, \hat{u}_{1,e}^{2i-1} \oplus \hat{u}_{1,o}^{2i-1}\right)^{1-2\hat{u}_{2i}} L_{N/2}^{(i)}\left(y_{N/2}^N, \hat{u}_{1,o}^{2i-1}\right). \tag{29}$$

From Equations (28) and (29), it can be seen that each LR value $\{L_N^{(i)}(y_1^N, u_1^{i-1}) : 1 \leq i \leq N\}$ in the pair $(L_N^{(2i)}(y_1^N, \hat{u}_1^{2i-1}), L_N^{(2i+1)}(y_1^N, \hat{u}_1^{2i}))$ is computed with $O(N)$ from the same pair of LRs $(L_{N/2}^{(i)}(y_1^{N/2-1}, \hat{u}_{1,e}^{2i-1} \oplus \hat{u}_{1,o}^{2i-1}), L_{N/2}^{(i)}(y_{N/2}^N, \hat{u}_{1,o}^{2i-1}))$. Thus, the calculation of all N LRs at length N requires exactly N LRs at length $N/2$. If the N LRs at length $N/2$ are split into two classes, namely

$$\begin{gathered} \left\{L_{N/2}^{(i)}\left(y_1^{N/2-1}, \hat{u}_{1,e}^{2i-1} \oplus \hat{u}_{1,o}^{2i-1}\right) : 1 \leq i \leq N/2\right\} \\ \left\{L_{N/2}^{(i)}\left(y_{N/2}^N, \hat{u}_{1,o}^{2i-1}\right) : 1 \leq i \leq N/2\right\} \end{gathered} \tag{30}$$

each class in the above equation generates a set of $N/2$ LRs calculation requests at length $N/4$, for a total of N requests. Using this reasoning inductively across the set of all lengths $\{N, N/2, \ldots, 1\}$, we conclude that the total number of LRs which need to be calculated is $N(1 + \log N)$. So, the complexity of SC decoding can be reduced to $O(N \log N)$.

2.5. **Polar codes are capacity achieving.** The single step of channel transform moves both the symmetric capacity and the reliability away from the center in the sense that

$$I\left(W_N^{(2i-1)}\right) \leq I\left(W_{N/2}^{(i)}\right) \leq I\left(W_N^{(2i)}\right) \tag{31}$$

$$Z\left(W_N^{(2i-1)}\right) \geq Z\left(W_{N/2}^{(i)}\right) \geq Z\left(W_N^{(2i)}\right) \tag{32}$$

with equality iff $I(W_{N/2}^{(i)}) = 0$ or $I(W_{N/2}^{(i)}) = 1$, and iff $Z(W_{N/2}^{(i)}) = 0$ or $Z(W_{N/2}^{(i)}) = 1$ respectively. Since both quantities take values in $[0, 1]$, intuitively one expects that, as N is increased, $I(W_N^{(i)})$ and $Z(W_N^{(i)})$ will concentrate around 0 and 1. This behavior is formalized and quantified in Proposition 2.1. However, it is also important to know quickly this occurs with respect to N. The following theorem turns out to be enlightening in this aspect.

Theorem 2.1. (Improved version of Theorem 2 in [1]). *For any B-DMC W with $I(W) > 0$, and any fixed $R < I(W)$ and constant $\beta < \frac{1}{2}$, there exists a sequence of sets $A_N \subset (1, \ldots, N)$, $N \in (1, 2, \ldots, 2^n, \ldots)$. Such that $|A_N| > NR$ and*

$$\sum_{i \in A_N} Z(W_N^{(i)}) = o(2^{-N^\beta}) \tag{33}$$

Due to (25), the LHS of (33) is equal to $P_e(N, K, A)$. This leads to the following theorem.

Theorem 2.2. (Improved version of Theorem 3, [1]). *For polar coding on a B-DMC W at any fixed rate $R < I(W)$ and any fixed $\beta < \frac{1}{2}$*

$$P_e(N, K, A) = o(2^{-N^\beta}) \tag{34}$$

To summarize, polar codes have a structured encoder and a lower-complexity SC decoding algorithm, and the probability of block error under SC decoding is exponentially small in the block length.

2.6. **Simulation results.** In this part, we see the performance of polar codes using SC decoder over different channels. The frozen set for the additive white Gaussian noise (AWGN) are chosen using the estimates for $P_e(W_N^{(i)})$. These estimates are done using the Monte-Carlo method described in [1]. From the simulation result, we can see that the performance of polar codes is not very impressive in short block-lengths. Figures 6 and 7 show the performance of the SC decoder compared to the maximum likelihood decoding (MAP) decoder over the binary erasure channel (BEC) and AWGN channel respectively.

It notices that there exists a considerable gap between the error probability of the SC decoder and the MAP decoder. Therefore, an important question is that how to modify the SC decoder such that this gap decreases while having low complexity.

2.7. **Discussion.** Polar codes are of very high theoretical interest and value, because they are capacity achieving over many channels, as well as provably optimal for numerous other applications, in the sense of optimality that pertains to each case. From an application and implementation perspective, they have a number of desirable properties.

(A.1) Polar codes are a class of codes that *1)* are explicitly defined by a construction

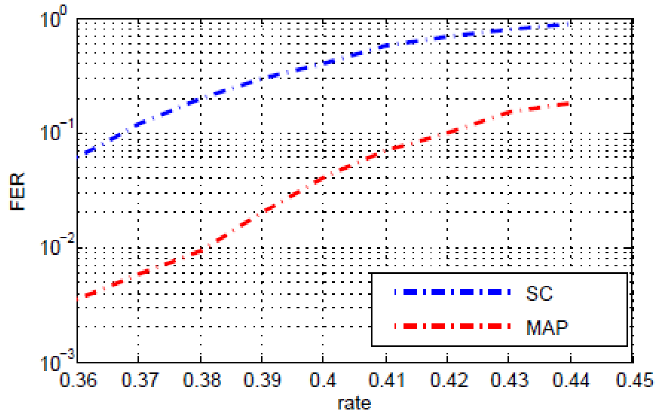

FIGURE 6. Performance of SC and MAP with $N = 1024$ and $\epsilon = 0.5$ over BEC

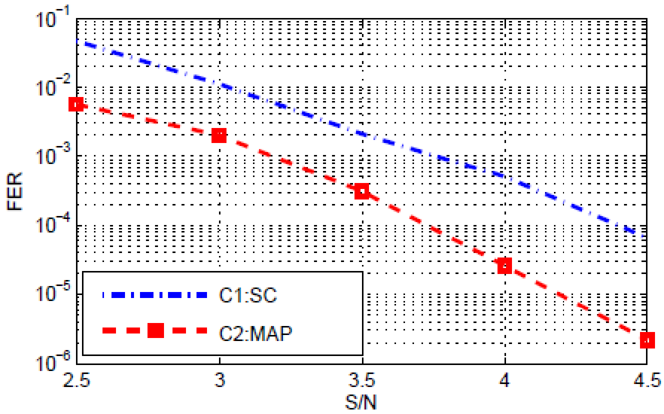

FIGURE 7. Performance of SC and MAP with $N = 256$ over AWGN

rule, *2)* provably achieve channel capacity on any symmetric B-DMCs, and *3)* have low-complexity encoding and decoding algorithms.

(A.2) Both the encoding and the decoding processes are inherently very structured. Structure makes routing and control in hardware implementations much easier.

(A.3) For polar codes, rate adaptation can be carried out trivially and in a very fine-grained way by simply altering the set A on the fly. While for LDPC codes, a different rate usually requires a completely different code.

(A.4) Polar codes have a very low error-floor due to their large stopping distance [2].

Unfortunately, as the above simulation results shown, since sub-optimality of the SC decoding algorithm and low minimum distance of polar codes, the finite length performance of polar codes under SC decoding is not very impressive. This motivates the exploration of more sophisticated decoding algorithms.

3. **Literature Survey on Polar Codes.** Since polar coding is a new technique, there are still many open questions about polar codes. In the following, we report a list of the most recent publications investigating the polar codes and analyzing the current deficiencies of these studies. Based on the above discussion, several interesting open problems and some new research direction of polar codes are presented. These results of researches can give us a new idea of further study polar codes.

3.1. **Construction of polar codes.** The construction of polar codes is one of the important concerns in using polar codes in practice. As described in Section 2.3, designing a polar code is equivalent to finding the set of good indices among the N polarized channels. In [1], Arikan proposed a criterion on which i with small $Z(W_n^{(i)})$ are chosen as information variables in order to minimize the upper bound (25). However, unless W is the BEC, the complexity of the evaluation of $Z(W_n^{(i)})$ is exponential in the block-length. In order to avoid the high cost of computation, a Monte-carlo method which estimates $Z(W_n^{(i)})$ by numerical simulation is suggested in [1]. However, polar codes constructed by these methods do not provably achieve symmetric capacity. So computing the exact transition probabilities of these channels seems to be intractable and hence we need some efficient methods to "approximate" these channels.

In [3,4], Mori and Tanaka use density evolution tool to improve the code construction for any B-DMCs. But this method requires large memory and high computation complexity increasing with the code length, the implementation of density evolution is not tractable. Another method using the idea of Gaussian Approximation is proposed in [5] under the AWGN channel. Compared with the method base on density evolution, the computation complexity is reduced. However, it seems that polar codes constructed by this method have quite small minimum distance.

In [6], Tal and Vardy presented a method for constructing polar codes which controls this growth. They approximated each bit-channel with a "better" channel and a "worse" channel while reducing the alphabet size. They constructed a polar code based on the "worse" channel and used the "better" channel to measure the distance from the optimal channel. However, as mentioned in [6], the problem of constructing optimal degrading and upgrading functions is still an open problem. Thus, one can change these functions to produce a new approach. In fact, the real issue is the definition of optimal, but it has not yet been addressed adequately. The development of a clear definition of optimality and a measurement reference to compare different approaches would be very useful in constructing new polar codes.

3.2. **Coding scheme.** To make polar codes more practice, concatenating inner polar codes with outer linear codes is a promising path towards improving their finite length performance while preserving their low decoding complexity.

In [7], it is shown how the classical idea of code concatenation using "short" polar codes as inner codes and a "high-rate" Reed-Solomon code as the outer code results in improved performance with respect to the use of polar codes alone. In particular, code concatenation with a careful choice of parameters boots the rate decay of the error probability to almost exponential in the block-length with essentially no loss in computational complexity. However, this work assumed a conventional method of concatenation, which required the cardinality of the outer Reed-Solomon (RS) code alphabet to be exponential in the block length of the inner polar code, which makes it infeasible for implementation in practical systems.

A scheme for concatenating binary polar codes with interleaved RS codes is considered in [8], which can capture the capacity-achieving property of polar codes, while having a significantly better error-decay rate.

3.3. **Non-binary codes.** The study of polar codes for channels with nonbinary input was undertaken by Şaşoğlu et al. and Mori and Tanaka. In [9], it is shown that given two copies of a q-ary input channel W, where q is prime, it is possible to create two channels W^+ and W^- whose symmetric capacities satisfy $I(W^-) < I(W) < I(W^+)$, where the inequalities are strict expect in trivial cases. This leads to simple proof of channel polarization in the q-ary case.

In [10], the authors calculate by numerical simulation the error probability of non-binary polar codes constructed on the basis of Reed-Solomon matrics. It is confirmed that 4-ary polar codes have significantly better performance than binary polar codes on binary-input AWGN channels. The authors also discuss an interpretation of polar codes in terms of algebraic geometry codes, and further show that polar codes using Hermitian codes have asymptotically good performance.

3.4. **Decoding strategies.** In order to improve the finite-length performance of polar codes, several decoding algorithms have been proposed.

Particularly, the authors in [11] proposed a belief propagation (BP) decoder and compared the performance of polar codes with Reed-Muller codes. The BP decoder has the advantage of good performance and soft outputs, at the expense of high memory and complexity requirements. Therefore, the SC decoder remained an attractive choice for low-cost decoding of polar codes. In [12], an SC list (SCL) decoder is proposed. Empirically, the usage of L concurrent decoding paths yields a significant improvement in the achievable error probability and allows to obtain an error probability comparable to that under MAP decoding with practical values of the list size. Furthermore, they also showed that if concatenated with a very high rate, cyclic redundancy check (CRC) code, polar codes perform comparably to state-of-the-art low density parity check (LDPC) codes. The performances are shown in Figure 8. Despite its good performance, the SCL decoder increased the memory and processing requirement as compared to the SC decoder. Later in [13], the authors showed that polar codes concatenated with CRC-24 codes can reach 0.25 dB away from the information theoretic limit at as low a block length as $N = 2048$. However, they achieved this performance gain using an adaptive SC list (ASCL) decoder of very large list size that may not be suitable for devices which memory is scare. Both SCL and ASCL decoders, however complex, showed that polar codes concatenated with other codes can be very effective.

Except for the above SCL decoding method, concatenated polar codes with Reed-Solomn (RS) codes and LDPC codes are recently proposed respectively in [14] and [7]. In [14], the authors also demonstrated that polar codes concatenated with LDPC codes can reach lower error floors when compared to capacity-achieving LDPC codes, at the cost of higher computation complexity and storage requirement incurred by the BP decoder used

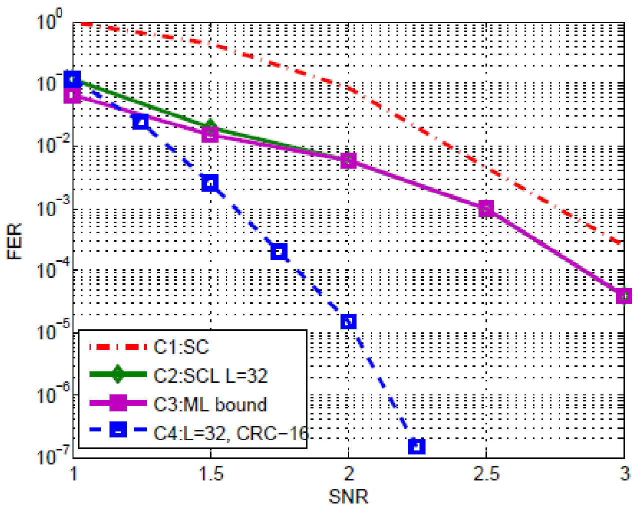

FIGURE 8. Performance of SC and SCL is with different L. The CRC used was 16 bits long.

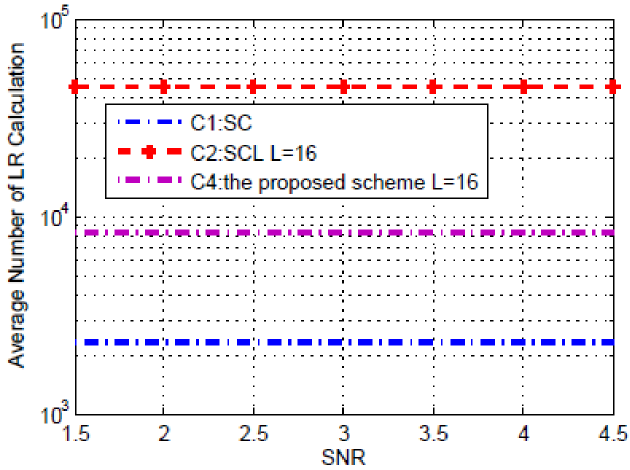

FIGURE 9. Complexities of SC, SCL and our proposed scheme

for polar codes. Therefore, finding an SC decoder with low memory and time-complexity is still an open problem.

In our work, we have mainly studied how to reduce the decoding complexity by using the idea of partial list SC decoding under CRC-aided of polar codes. Figure 9 shows that the complexities of our scheme, SC and SCL. It can be seen that the complexity of our proposed scheme has the lower complexity with almost the same performance.

3.5. **Asymptotic behavior and scaling laws.** The authors of [15] consider the asymptotic behavior of the polarization process for polar codes when the block length tends to infinity. In [16], upper and lower bounds are provided on the escaped rate of the Bhattacharyya process corresponding to polar codes where transmission takes place over the binary erasure channel. In [17], some recent progress in studies on speed of channel polarization are reviewed. Firstly, results on a code rate-dependent upper bound of block error probability of polar codes with SC decoding are reviewed. Then an approach of constructing polar codes for non-binary input alphabets with asymptotic speed of polarization much faster than previous approaches is briefly described. In [18], a rate-dependent upper bound of the best achievable block error probability of polar codes with SC decoding is derived. In [19], the authors make a *Scaling Assumption* that the probability $Q(x)$ that Z_n exceeds x is such that $\lim_{n\to\infty} N^{1/\mu}Q(x)$ exits and equals a function $Q(x)$. Under this assumption, they use simulations to numerically estimate $\mu \approx 3.627$ for the BEC. Using the small x asymptotics of $Q(x)$ suggested by the numerical data, they predict an upper bound on the block length as a function of the gap ε to capacity for the BEC. For general channels, under the heuristic assumption that the densities of log-likelihood ratios behave like Gaussians, an exponent of $\mu \approx 4.001$ is suggested for the Scaling Assumption. However, to the best of our knowledge, it does not appear that one can get a rigorous upper bound on block length N as a function of the gap to capacity via these methods.

3.6. **Source coding.** Just as channel polarization can be used to design capacity-achieving codes, source polarization can be used to design lossless source codes that achieve the entropy bound. In [20], it shows that polar codes with an SC decoder are also optimal for lossy and lossless source coding as well as muti-terminal problems like the Slepian-Wolf, the WynerZiv, and the Gelfand-Pinsker problems. The complexity of the encoding and the decoding algorithm in both cases is also $O(N \log N)$.

In [21], lossless compression with polar codes is considered. A polar encoding algorithm is developed and a method to design the code and compute the average compression rate for finite lengths is given. It is shown that the scheme achieves the optimal compression rate asymptotically. In [22], it is shown that for binary sources, there exists a universal polar code which can compress any source of low enough entropy, without requiring knowledge of the source distribution. While this result does not extend to q-ary sources, it is shown that how it extends to q-ary sources which belong to a restricted family.

3.7. **Wiretap channel.** In the field of wireless physical layer security, polar codes are widely utilized to help obtain the secrecy capacity of wiretap channels and to make sure the information is transmitted both reliably and securely between the transmitter and the legal receiver.

In [23], it is shown that polar codes asymptotically achieve the whole capacity-equivocation region for the wiretap channel when the wiretapper's channel is degraded with respect to the main channel, and the weak secrecy notion is used. The proposed coding scheme also achieves the capacity of the physically degraded receiver-orthogonal relay channel. In [24], The provided scheme achieves the entire rate-equivocation region for the considered model. In [25], polar codes are used to construct a coding scheme that achieves the secrecy capacity of general wiretap channels. The construction works for any instantiation of the wiretap channel model, as originally defined by Wyner, as long as both channels defining the wiretap channel are symmetric and binary-input.

3.8. **The design of efficient hardware architectures.** Despite many desirable properties of polar codes, it requires a very large code block length in order to actually approach the channel capacity ($N > 2^{20}$). Consequently, the practical interest of polar codes highly depends on the possibility to design efficient encoders and decoders for large N values.

In [1], Arikan suggests to use a fast Fourier transform structure to efficiently reuse computations. This first architecture requires $N \log_2(N)$ processing elements (PEs) and as many memory elements (MEs). In [26], a line architecture is implemented. It only uses $N - 1$ PEs and as many MEs without affecting the decoding performance and the throughput. In [27], it is shown that the number of PEs can be further reduced with a negligible impact on throughput. Since SC decoding has a low intrinsic parallelism, complementary works focused on increasing the throughput of SC decoders. In [28,29], lookahead techniques are used to reduce the decoding latency while using limited extra hardware resources. In [30], a simplification of SC decoding is proposed in order to reduce the number of computations without altering error correction performance. Extra latency reduction technique is investigated in [31] where maximum likelihood decoding is used to further speedup the decoding process. However, these low latency decoders have not been implemented yet.

4. **Conclusions.** Polar codes have received a rapidly growing attention in the literature because of these nice asymptotical properties on complexity and performance. In this paper, we describe this technique, and briefly discuss its recent publications to provide a perspective on current and proposed work, then analyze the existing disadvantages of the current work and also propose some several future study problems.

Acknowledgment. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers. This research is supported by the Shaanxi Science Technology Research Project (Grant No. 2013JK2018).

REFERENCES

[1] E. Arikan, Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels, *IEEE Transactions on Information Theory*, vol.55, no.7, pp.3051-3073, 2009.

[2] A. Eslami and H. Pishro-Nik, On bit error rate performance of polar codes in finite regime, *Proc. of the 48th Annual Allerton Conf. Communication, Control, and Computing (Allerton)*, pp.188-194, 2010.

[3] R. Mori and T. Tanaka, Performance and construction of polar codes on symmetric binary-input memoryless channels, *IEEE International Symposium on Information Theory Proceedings*, Seoul, South Korea, pp.1496-1500, 2009.

[4] R. Mori and T. Tanaka, Performance of polar codes with the construction using density evolution, *IEEE Communications Letters*, vol.13, no.7, 2009.

[5] P. Trifonov, Efficient design and decoding of polar codes, *IEEE Transactions on Communications*, vol.60, no.11, 2012.

[6] I. Tal and A. Vardy, How to construct polar codes, *IEEE Trans. on Inform. Theory*, http://arxiv.org/abs/1105.6164, 2013.

[7] M. Bakshi, S. Jaggi and M. Effros, Concatenated polar codes, *IEEE International Symposium on Information Theory Proceedings*, pp.918-922, 2010.

[8] H. Mahdavifar, M. El-Khamy, J. Lee and I. Kang, On the construction and decoding of concatenated polar codes, *Proc. of the 37th IEEE Conf. on Decision & Control*, Tampa, FL, USA, pp.4551-4552, 1998.

[9] E. Sasoglu, An entropy inequality for q-ary random variables and its application to channel polarization, *IEEE International Symposium on Information Theory Proceedings*, pp.1360-1363, 2010.

[10] R. Mori and T. Tanaka, Non-binary polar codes using reed-solomon codes and algebraic geometry codes, *IEEE Symposium on Computational Intelligence and Games*, pp.1-5, 2010.

[11] E. Arikan, A performance comparsion of polar codes and reed-muller codes, *Communications Letters, IEEE*, vol.12, no.6, pp.447-449, 2008.

[12] I. Tal and A. Vardy, List decoding of polar codes, *IEEE International Symposium on Information Theory Proceedings*, pp.1-5, 2011.

[13] B. Li, H. Shen and D. Tse, An adaptive successive cancellation list decoder for polar codes with cyclic redundancy check, *Communication Letters, IEEE*, vol.16, no.12, pp.2044-2047, 2012.

[14] A. Eslami and H. Pishro-Nik, On finite-length performance of polar codes: Stopping sets, error floor, and concatenated design, *CoRR*, vol.abs/1211.2187, 2012.

[15] S. Hassani and R. Urbanke, On the scaling of polar codes: I. the behavior of polarized channels, *IEEE International Symposium on Information Theory Proceedings*, pp.874-878, 2010.

[16] S. Hassani, K. Alishahi and R. Urbanke, On the scaling of polar codes: II. The behavior of un-polarized channels, *IEEE International Symposium on Information Theory Proceedings*, pp.879-883, 2010.

[17] T. Tanaka, On speed of channel polarization, *IEEE Symposium on Computational Intelligence and Games*, pp.1-5, 2010.

[18] T. Tanaka and R. Mori, Refined rate of channel polarization, *IEEE International Symposium on Information Theory Proceedings*, pp.889-893, 2010.

[19] S. Korada, A. Montanari, E. Telatar and R. Urbanke, An empirical scaling law for polar codes, *IEEE International Symposium on Information Theory Proceedings*, pp.884-888, 2010.

[20] S. B. Korada and R. Urbanke, Polar codes are optimal for lossy source coding, *IEEE Transactions on Information Theory*, vol.56, no.4, pp.1751-1768, 2009.

[21] H. Cronie and S. Korada, Lossless source coding with polar codes, *IEEE International Symposium on Information Theory Proceedings*, pp.904-908, 2010.

[22] E. Abbe, Universal source polarization and sparse recovery, *IEEE Symposium on Computational Intelligence and Games*, pp.1-5, 2010.

[23] M. Andersson, V. Rathi, R. Thobaben, J. Kliewer and M. Skoglund, Nested polar codes for wiretap and relay channels, *Communications Letters, IEEE*, vol.14, no.8, pp.752-754, 2010.

[24] E. Hof and S. Shamai, Secrecy-achieving polar-coding, *IEEE Trans. on Information Theory Workshop*, pp.1-5, 2010.

[25] H. Mahdavifar and A. Vardy, Achieving the secrecy capacity of wiretap channels using polar codes, *IEEE International Symposium on Information Theory Proceedings*, pp.913-917, 2010.

[26] C. Leroux, A. J. Raymond, G. Sarkis, I. Tal, A. Vardy and W. J. Gross, Hardware implementation of successive-cancellation decoders for polar codes, *Journ. of Sig. Proc. Syst.*, 2012.
[27] C. Leroux, A. Raymond, G. Sarkis and W. Gross, A semiparallel successive-cancellation decoder for polar codes, *IEEE Transactions on Signal Processing*, 2012.
[28] C. Zhang, B. Yuan and K. K. Parhi, Low-latency SC decoder architectures for polar codes, *arXiv:1111.0705*, 2011.
[29] C. Zhang and K. Parhi, Low-latency sequential and overlapped architectures for successive cancellation polar decoder, *IEEE Transactions on Signal Processing*, 2013.
[30] A. Alamdar-Yazdi and F. R. Kschischang, A simplified successive-cancellation decoder for polar codes, *IEEE Communications Letters*, 2011.
[31] G. Sarkis and W. J. Gross, Increasing the throughput of polar decoders, *IEEE Communications Letters*, 2013.