# Does Gaussian Approximation Work Well for the Long-Length Polar Code Construction?

**JINCHENG DAI, (Student Member, IEEE), KAI NIU, (Member, IEEE), ZHONGWEI SI, (Member, IEEE), CHAO DONG, (Member, IEEE), and JIARU LIN, (Member, IEEE)**

Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China

Corresponding author: Jincheng Dai (daijincheng@bupt.edu.cn)

**ABSTRACT** Gaussian approximation (GA) is widely used to construct polar codes. However, when the code length is long, the subchannel selection inaccuracy due to the calculation error of conventional approximate GA (AGA), which uses a two-segment approximation function, results in a catastrophic performance loss. In this paper, new principles to design the GA approximation functions for polar codes are proposed. First, we introduce the concepts of polarization violation set (PVS) and polarization reversal set (PRS) to explain the essential reasons that the conventional AGA scheme cannot work well for the long-length polar code construction. In fact, these two sets will lead to the rank error of subsequent subchannels, which means that the orders of subchannels are misaligned, which is a severe problem for polar code construction. Second, we propose a new metric, named cumulative-logarithmic error (CLE), to quantitatively evaluate the remainder approximation error of AGA in the logarithm. We derive the upper bound of CLE to simplify its calculation. Finally, guided by PVS, PRS, and CLE bound analysis, we propose new construction rules based on a multi-segment approximation function, which obviously improve the calculation accuracy of AGA so as to ensure the excellent performance of polar codes especially for the long code lengths. Numerical and simulation results indicate that the proposed AGA schemes are critical to constructing high-performance polar codes.

**INDEX TERMS** Polar codes, Gaussian approximation (GA), polarization violation set (PVS), polarization reversal set (PRS), cumulative-logarithmic error (CLE).

## I. INTRODUCTION

Polar codes proposed by Arıkan [1] have been proved to achieve the capacity of any symmetric binary input symmetric discrete memoryless channels (B-DMCs) under a successive cancellation (SC) decoder as the code length goes to infinity. Recently, polar codes have been identified as one of the channel coding schemes in the 5G wireless communication system due to its excellent performance [2]. To construct polar codes, the channel reliabilities are calculated efficiently using the symmetric capacities of subchannels or the Bhattacharyya parameters for the binary-input erasure channels (BECs). As a heuristic method, Arıkan has suggested to use the recursion which is optimal only for BECs also for other B-DMCs [3]. Mori and Tanaka [4] regarded the construction problem as an instance of density evolution (DE), which theoretically has the highest accuracy. Considering its high computational complexity,

Tal and Vardy devised two approximation methods to simplify the calculation of DE, by which one can get the upper and lower bounds on the error probability of each subchannel. Tal and Vardy's method has almost no performance loss compared with DE [5]–[8]. Afterwards, Gaussian approximation (GA) was proposed to further reduce the computational complexity of DE [9] without much sacrifice in accuracy, which became popular in the construction of polar codes thanks to its good tradeoff between the complexity and performance.

In the GA construction of polar codes, the bit log-likelihood ratio (LLR) of each subchannel is assumed to obey a constraint Gaussian distribution in which the mean is half of the variance. Hence, the iterative evaluation of each subchannel reliability is only involved with the mean update of LLRs. However, the LLR mean updates in check nodes still depend on complex integration. Consequently, for

J. Dai *et al.*: Does GA Work Well for the Long-Length Polar Code Construction?

**IEEE** Access

construction of polar codes, the computational complexity of exact GA (denoted by EGA) grows exponentially with the polarization levels. This makes EGA too complicated to be practically employed. Therefore, in practical implementation, like GA utilized in LDPC codes, the well-known approximate version of GA (denoted by AGA) given by Chung *et al.* based on a two-segment approximation function is used to speed up the calculations [10]–[12].

Initially, the approximation function in conventional AGA chosen by Chung is suitable for LDPC codes. However, in principle, we don't know whether this approximation method can be also good for the polar codes. In fact, the calculation error of AGA versus EGA will be accumulated and amplified in the recursion process of polar codes construction. Consequently, this phenomenon causes the inaccurate subchannel selection and results in a catastrophic block error ratio (BLER) performance loss for the long code lengths. Taking the polar code structure characteristics into consideration, it is lack of the comprehensive framework to design the AGA schemes for the construction of polar codes. In addition, the performance evaluation method for different AGA is also absent.

On the polar code construction, we find that the rank error of the subchannels and the calculation error of each subchannel's reliability are two critical factors to affect the accuracy of AGA approximation function. Here, ranking error means the orders of subchannels are misaligned. For various AGA schemes, the two factors will result in different evaluation error of each subchannel. Followed by above two distortions, we reveal the essential reason that the conventional AGA scheme leads to a catastrophic performance loss. Our ultimate goal is to propose systematic design rules of the AGA approximation function for polar codes, which achieves the excellent performance as well as reduces the GA computational complexity.

Our aim in this paper is to provide the new principles to design the multi-segment GA approximation functions so as to improve the calculation accuracy of AGA and guarantee the excellent performance of polar codes. The main contributions can be summarized in the following three aspects:

- First, we take a closer investigation at the reason behind the poor performance of the long-length polar codes when the conventional versions of AGA (e.g. Chung's scheme) are used. To this end, we introduce the concepts of polarization violation set (PVS) and polarization reversal set (PRS). In the AGA process, when the subchannel's LLR mean belongs to the two sets, it will bring in the rank error and the polarization is 'violated' or 'reverted' among the subsequent subchannels. This phenomenon is not consistent with Arıkan's fundamental polarization relationship. The two sets reveal the essential reason that polar codes constructed by the conventional AGA present poor performance at long code lengths.

- Second, after eliminating PVS and PRS, we further propose a new metric, named cumulative-logarithmic

error (CLE) of channel polarization, to quantitatively evaluate the remainder calculation error between AGA and EGA in the construction of polar codes. We also derived the upper bound of CLE to simplify its calculation. With this bound, the performance of different versions of AGA can be easily evaluated by analytic calculation rather than redundant the Monte-Carlo simulation.

- Finally, guided by PVS, PRS and the CLE bound, we propose new design rules for the improved AGA techniques which is tailored for the polar code construction. In this way, a systematic framework is established to design the high accuracy and low complexity AGA scheme for polar codes at any code length. Followed by the proposed rules, three new AGA schemes are given to guarantee the excellent performance of polar codes.

The remainder of the paper is organized as follows. The preliminaries of polar coding are described in Section II. Then the conventional GA is introduced in Section III. Section IV makes detailed error analysis of GA, in which the concepts of PVS, PRS and CLE are proposed. Then the new design rules of AGA approximation functions are given in Section V, and new AGA schemes with complexity comparison are also given in this part. Different versions of AGA are compared with the help of CLE bound in Section VI, where the simulation results are also analyzed. Finally, Section VII concludes this paper.

## II. PRELIMINARIES
### A. NOTATION CONVENTIONS
In this paper, we use calligraphic characters, such as $\mathcal{X}$, to denote sets. Let $|\mathcal{X}|$ denote cardinality of $\mathcal{X}$. We write lowercase letters (e.g., $x$) to denote scalars. We use notation $v_1^N$ to denote a vector $(v_1, v_2, \cdots, v_N)$ and $v_i^j$ to denote a subvector $(v_i, v_{i+1}, \cdots, v_j)$. The sets of binary and real field are denoted by $\mathbb{B}$ and $\mathbb{R}$, respectively. Specially, let $\mathcal{N}(a, b)$ denote Gaussian distribution, where $a$ and $b$ represent the mean and the variance respectively. For polar coding, only square matrices are involved in this paper, and they are denoted by bold letters. The subscript of a matrix indicates its size, e.g. $\mathbf{F}_N$ represents an $N \times N$ matrix $\mathbf{F}$. The Kronecker product of two matrices $\mathbf{F}$ and $\mathbf{G}$ is expressed as $\mathbf{F} \otimes \mathbf{G}$, and the $n$-fold Kronecker power of $\mathbf{F}$ is denoted by $\mathbf{F}^{\otimes n}$.

Throughout this paper, $\log(\cdot)$ means "logarithm to base 2", and $\ln(\cdot)$ stands for the natural logarithm.

### B. POLAR CODES AND SC DECODING
Let $W : \mathcal{X} \rightarrow \mathcal{Y}$ denote a B-DMC with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. The channel transition probabilities are given by $W(y|x)$, $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Given the code length $N = 2^n, n = 1, 2, \cdots$, the information length $K$, and the code rate $R = K/N$, the polar coding is described as [1]. After the channel combining and splitting operations on $N$ independent duplicates of $W$, we obtain $N$ successive uses of synthesized binary input channels $W_N^{(j)}$, $j = 1, 2, \cdots, N$, with transition probabilities $W_N^{(j)}(y_1^N, u_1^{j-1} | u_j)$. The information bits can be

**IEEE** Access

J. Dai et al.: Does GA Work Well for the Long-Length Polar Code Construction?

assigned to the channels with indices in the information set $\mathcal{A}$, which are the more reliable subchannels. The complementary set $\mathcal{A}^c$ denotes the frozen bit set and the frozen bits $u_{\mathcal{A}^c}$ can be set as the fixed bit values, such as all zeros, for the symmetric channels. To put it in another way [1], polar coding is performed on the constraint $x_1^N = u_1^N \mathbf{G}_N$, where $\mathbf{G}_N$ is the generator matrix and $u_1^N, x_1^N \in \{0, 1\}^N$ are the source and code block respectively. The source block $u_1^N$ consists of information bits $u_{\mathcal{A}}$ and frozen bits $u_{\mathcal{A}^c}$. The generator matrix can be defined as $\mathbf{G}_N = \mathbf{B}_N \mathbf{F}_2^{\otimes n}$, where $\mathbf{B}_N$ is the bit-reversal permutation matrix and $\mathbf{F}_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$.

As mentioned in [1], polar codes can be decoded by successive cancellation (SC) decoding algorithm. Let $\hat{u}_1^N$ denote an estimate of source block $u_1^N$. After receiving $y_1^N$, the bits $\hat{u}_j$ are successively determined with index from 1 to $N$ in the following way:

$$\hat{u}_j = \begin{cases} h_j(y_1^N, \hat{u}_1^{j-1}) & j \in \mathcal{A}, \\ u_j & j \in \mathcal{A}^c, \end{cases} \quad (1)$$

where

$$h_j(y_1^N, \hat{u}_1^{j-1}) = \begin{cases} 0 & \text{if } \dfrac{W_N^{(j)}\left(y_1^N, \hat{u}_1^{j-1} \mid 0\right)}{W_N^{(j)}\left(y_1^N, \hat{u}_1^{j-1} \mid 1\right)} \geq 1, \\ 1 & \text{otherwise.} \end{cases} \quad (2)$$

Given a polar code with code length $N$, information length $K$ and selected channels indices $\mathcal{A}$, the BLER under SC decoding algorithm is upper bounded by

$$P_e(N, K, \mathcal{A}) \leq \sum_{j \in \mathcal{A}} P_e\left(W_N^{(j)}\right), \quad (3)$$

where $P_e(W_N^{(j)})$ is the error probability of the $j$-th subchannel. This BLER upper bound is named as the *SC bound*.

## III. GAUSSIAN APPROXIMATION FOR POLAR CODES
In this section, we use the code tree to describe the process of channel polarization. Based on the tree structure, we present and analyze the basic procedure of GA.

### A. CODE TREE
The channel polarization process can be expressed on a code tree. For a polar code with length $N = 2^n$, the corresponding code tree $\mathcal{T}$ is a perfect binary tree.[1] Specifically, $\mathcal{T}$ can be represented as a 2-tuple $(\mathcal{V}, \mathcal{B})$, where $\mathcal{V}$ and $\mathcal{B}$ denote the set of nodes and the set of edges, respectively.

Depth of a node is the length of the path from the root to this node. The set of all the nodes at a given depth $i$ is denoted by $\mathcal{V}_i$, $i = 0, 1, 2, \cdots, n$. The root node has a depth of zero. Let $v_i^{(j)}$, $j = 1, 2, \cdots, 2^i$, denote the $j$-th node from left to right in $\mathcal{V}_i$. As an illustration, Fig. 1 shows a toy example of code tree with $N = 16$, which includes 4 levels. In the nodes
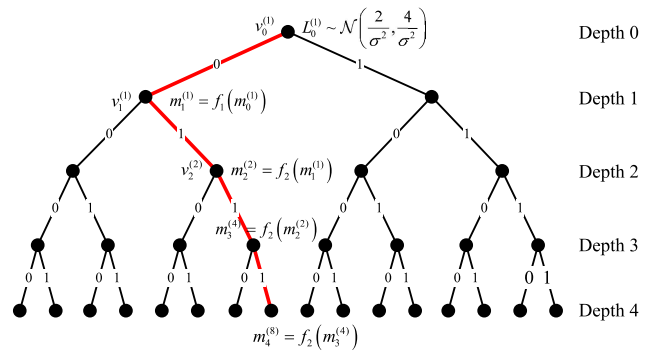
---

[1]A perfect binary tree is a binary tree in which all interior nodes have two children and all leaves have the same depth or same level.



**FIGURE 1.** An example of code tree for $N = 16$, $n = 4$. The red bold edge shows the recursive calculation process of $m_4^{(8)}$.

set $\mathcal{V}_2$, the 2nd node from left to right is denoted by $v_2^{(2)}$. Except for the nodes at the $n$-th depth, each $v_i^{(j)} \in \mathcal{V}_i$ has two descendants in $\mathcal{V}_{i+1}$, and the two corresponding edges are labeled as 0 and 1, respectively. The nodes $v_n^{(j)} \in \mathcal{V}_n$ are called leaf nodes. Let $\mathcal{T}(v_i^{(j)})$ denote a subtree with a root node $v_i^{(j)}$. The depth of this subtree can be defined as $n - i$ which indicates the difference between the depth of the leaf node and that of the root node. In addition, the node $v_i^{(j)}$ has two subtrees, that is, the left subtree $\mathcal{T}_{\text{left}} = \mathcal{T}(v_{i+1}^{(2j-1)})$ and the right subtree $\mathcal{T}_{\text{right}} = \mathcal{T}(v_{i+1}^{(2j)})$.

All the edges in the set $\mathcal{B}$ are partitioned into $n$ levels $\mathcal{B}_l$, $l = 1, 2, \cdots, n$. Each edge in the $l$-th level $\mathcal{B}_l$ is incident to two nodes: one at depth $l - 1$ and the other at depth $l$. An $i$-depth node is corresponding to a path $(b_1, b_2, \cdots, b_i)$ which consists of $i$ edges, with $b_l \in \mathcal{B}_l$, $l = 1, 2, \cdots, i$. A vector $b_1^i = (b_1, b_2, \cdots, b_i)$ is used to depict the above path.

### B. GAUSSIAN APPROXIMATION FOR POLAR CODES
Trifonov [9] suggests a polar code construction method for the binary input AWGN (BI-AWGN) channels based on a Gaussian assumption in every recursion step. For the BI-AWGN channels with noise variance $\sigma^2$, the coded bits are modulated using binary phase shift keying (BPSK). The transition probability $W(y|x)$ is written as

$$W(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-(1-2x))^2}{2\sigma^2}}, \quad (4)$$

where $x \in \mathbb{B}$ and $y \in \mathbb{R}$. The LLR of each received symbol $y$ is denoted by

$$L(y) = \ln \frac{W(y|0)}{W(y|1)} = \frac{2y}{\sigma^2}. \quad (5)$$

Without loss of generality, we assume that all-zero codeword is transmitted. One can check $L(y) \sim \mathcal{N}\left(\frac{2}{\sigma^2}, \frac{4}{\sigma^2}\right)$.

*GA Assumption:* The LLR of each subchannel obeys a constraint Gaussian distribution in which the mean is half of the variance [9]–[12].

According to the GA assumption, the only issue needed to be dealt with is the LLR mean. Therefore, in the construction of polar codes, to obtain the reliability of each subchannel, we

J. Dai *et al.*: Does GA Work Well for the Long-Length Polar Code Construction?

IEEE *Access*

trace their LLR mean. This recursive calculation process can be performed on the code tree. The set of LLRs corresponding to the nodes at depth $i$ is denoted by $\mathcal{L}_i, i = 0, 1, 2, \cdots, n$. Let $L_i^{(j)}, j = 1, 2, \cdots, 2^i$, denote the $j$-th element in $\mathcal{L}_i$. We write $m_i^{(j)}$ as the mean of $L_i^{(j)}$. So the mean of LLR from the channel information can be written as $m_0^{(1)} = \frac{2}{\sigma^2}$, and under the GA assumption we have

$$L_i^{(j)} \sim \mathcal{N}\left(m_i^{(j)}, 2m_i^{(j)}\right). \quad (6)$$

Here, $m_i^{(j)}$ can be computed recursively as

$$m_{i+1}^{(2j-1)} = f_c\left(m_i^{(j)}\right), m_{i+1}^{(2j)} = f_v\left(m_i^{(j)}\right), \quad (7)$$

where the functions $f_c(t)$ and $f_v(t)$ are used for check nodes (left branch) and variable nodes (right branch), respectively. The physical meaning of the function variable $t$ stands for subchannel's LLR mean in GA construction. We have

$$\begin{cases} f_c(t) = \phi^{-1}\left(1 - (1 - \phi(t))^2\right), \\ f_v(t) = 2t. \end{cases} \quad (8)$$

In EGA, $\phi(t)$ is written as

$$\phi(t) = \begin{cases} 1 - \dfrac{1}{\sqrt{4\pi t}} \int_{\mathbb{R}} \tanh\left(\dfrac{z}{2}\right) e^{-\frac{(z-t)^2}{4t}} dz & t > 0, \\ 1 & t = 0, \end{cases} \quad (9)$$

where $\tanh(\cdot)$ denotes hyperbolic tangent function. It is easy to check that $\phi(t)$ is continuous and monotonically decreasing on $[0, +\infty)$, with $\phi(0) = 1$ and $\phi(+\infty) = 0$ [10]. As an illustration, the red bold edge in Fig. 1 depicts the recursive calculation process of $m_4^{(8)}$, whose corresponding path is denoted by $(b_1, b_2, b_3, b_4) = (0, 1, 1, 1)$.

Obviously, the exact calculation of LLR mean in check nodes requires complex integration, which results in a high computational complexity. Therefore, Chung *et al.* give the well known two-segment approximation function of $\phi(t)$, denoted by $\varphi(t)$, for the analysis of LDPC codes in [10],

$$\varphi(t) = \begin{cases} e^{-0.4527t^{0.86}+0.0218} & 0 < t < 10, \\ \sqrt{\dfrac{\pi}{t}} e^{-\frac{t}{4}}\left(1 - \dfrac{10}{7t}\right) & t \geq 10. \end{cases} \quad (10)$$

(10) is also widely used in the construction of polar codes [11]. Its corresponding AGA algorithm is denoted by "Chung".

By the GA assumption of (6), the error probabilities of polarized subchannel $P_e(W_N^{(j)})$ can be written as

$$P_e\left(W_N^{(j)}\right) = Q\left(\frac{m_n^{(j)}}{\sqrt{2m_n^{(j)}}}\right) = Q\left(\sqrt{\frac{m_n^{(j)}}{2}}\right), \quad (11)$$

where $Q(\varsigma) = \frac{1}{\sqrt{2\pi}} \int_{\varsigma}^{+\infty} e^{-\frac{z^2}{2}} dz$. Thus, the SC bound can be written as

$$P_e(N, K, \mathcal{A}) \leq \sum_{j \in \mathcal{A}} Q\left(\sqrt{\frac{m_n^{(j)}}{2}}\right). \quad (12)$$

Since $Q(\varsigma)$ is a monotone decreasing function, the subchannel $W_N^{(j)}$ with a larger mean $m_n^{(j)}$ has higher reliability. The construction of polar codes corresponds to the selection of best $K$ subchannels among $N$ as information set $\mathcal{A}$ in terms of the LLR means $m_n^{(j)}$, where $j = 1, 2, \cdots, N$.

## IV. ERROR ANALYSIS OF GAUSSIAN APPROXIMATION

In this section, we introduce the concepts of polarization violation set (PVS) and polarization reversal set (PRS). By calculating the two sets, we demonstrate the intrinsic reason that polar codes constructed by conventional AGA suffer from catastrophic performance loss at long code lengths. In order to quantitatively evaluate the remainder calculation error between AGA and EGA, we further propose the concept of cumulative-logarithmic error (CLE) of channel polarization and give a bound to simplify its calculation. Based on the CLE bound, we can efficiently evaluate the performance of different approximation functions in AGA.

### A. PVS AND PRS

*Proposition 1:* Under the GA assumption, each subchannel's symmetric capacity monotonically increases with its LLR mean.

It is worth noticing that a BI-AWGN channel's capacity $I(W)$ is written as

$$\begin{aligned} I(W) &= h\left(\sigma^2\right) \\ &\triangleq \frac{1}{2} \sum_{x \in \mathbb{B}} \int_{\mathbb{R}} W(y|x) \log\left(\frac{2W(y|x)}{W(y|0) + W(y|1)}\right) dy, \end{aligned} \quad (13)$$

where the transition probability $W(y|x)$ is given as (4). In terms of the GA principle, each subchannel is approximated by a BI-AWGN channel $W$ with LLR mean $m$. In addition, the variance of corresponding additive white Gaussian noise is $\sigma^2 = \frac{2}{m}$ under GA assumption. Since the function $h(\sigma^2)$ monotonically decreases with $\sigma^2$, the symmetric capacity $I(W)$ monotonically increases with its LLR mean $m$. In addition, we have

$$\lim_{m \to 0} I(W) = 0, \qquad \lim_{m \to +\infty} I(W) = 1. \quad (14)$$

*Proposition 2:* For the size-two channel polarization, suppose $(W, W) \mapsto (W_2^{(1)}, W_2^{(2)})$. Under the GA assumption, the LLR means corresponding to $W$, $W_2^{(1)}$ and $W_2^{(2)}$ are represented as $m$, $m_2^{(1)}$ and $m_2^{(2)}$, respectively. Then, the LLR means should satisfy

$$m_2^{(1)} \leq m \leq m_2^{(2)} \quad (15)$$

with equality if and only if $m = 0$ or $m = +\infty$.

This result follows the [1, Proposition 4]. Combining with *Proposition 1*, the proof of (15) is immediate. It can be seen from (15) that the reliability of the original channel $W$ is redistributed. Based on this interpretation, we may say that after one step polarization, a "bad" channel $W_2^{(1)}$ and a "good" channel $W_2^{(2)}$ have been created.

**IEEE** *Access*

J. Dai *et al.*: Does GA Work Well for the Long-Length Polar Code Construction?

*Proposition 3:* In the AGA construction of polar codes, the approximation function $\Omega(t)$ of $\phi(t)$ should satisfy

$$0 < \Omega(t) < 1, \qquad (16)$$

which is the necessary and sufficient condition for $\Omega(t)$ to satisfy *Proposition 2*.

*Proof:* Different from that in LDPC codes, the approximation function $\Omega(t)$ for polar codes should guarantee that the relationship of (15) holds.[2] Therefore, in the size-two channel polarization, for $t \in (0, +\infty)$, $\Omega(t)$ should satisfy

$$\Omega^{-1}\left(1 - (1 - \Omega(t))^2\right) < t < 2t, \qquad (17)$$

which follows *Proposition 2*. Recall that $\phi(t)$ is continuous and monotonically decreasing on $[0, +\infty)$ [10], we therefore assume its approximated form $\Omega(t)$ monotonically decreases on $(0, +\infty)$. Consequently, the left inequality in (17) can be simplified as

$$1 - (1 - \Omega(t))^2 > \Omega(t) \Rightarrow 0 < \Omega(t) < 1, \qquad (18)$$

In turn, if $\Omega(t)$ satisfies $0 < \Omega(t) < 1$, we have

$$1 - (1 - \Omega(t))^2 > 1 - (1 - \Omega(t))$$
$$\Rightarrow \Omega^{-1}\left(1 - (1 - \Omega(t))^2\right) < \underbrace{\Omega^{-1}\left(1 - (1 - \Omega(t))\right)}_{=t}$$
$$\Rightarrow \Omega^{-1}\left(1 - (1 - \Omega(t))^2\right) < t < 2t \qquad (19)$$

The above analysis indicates that (16) is the necessary and sufficient condition for $\Omega(t)$ to satisfy *Proposition 2*. ∎

If $\Omega(t)$ cannot meet (16), its approximation error with respect to the exact $\phi(t)$ will result in the following two types of reliability rank error:

Type 1) In the size-two channel polarization, if $\Omega(t)$ leads to $m \leq m_2^{(1)} < m_2^{(2)}$, this error indicates that the reliabilities of subchannels are partially violated, which is named as the "polarization violation" phenomenon.

Type 2) Furthermore, when $\Omega(t)$ leads to $m < m_2^{(2)} \leq m_2^{(1)}$. This error indicates the reliabilities of subchannels have been wrongly reversed, which is named as the "polarization reversal" phenomenon.

*Definition 1:* Given the approximation function $\Omega(t)$, the polarization violation set (PVS) $\mathcal{S}_{\mathrm{PVS}}$ is defined as

$$\mathcal{S}_{\mathrm{PVS}} = \left\{t \,\middle|\, t \leq \Omega^{-1}\left(1 - (1 - \Omega(t))^2\right) < 2t\right\}, \qquad (20)$$

where $t \in (0, +\infty)$.

Obviously, in the size-two channel polarization, for any LLR mean $m$ belonging to $\mathcal{S}_{\mathrm{PVS}}$, $\Omega(t)$ will certainly lead to $m \leq m_2^{(1)} < m_2^{(2)}$, which violates the basic order in *Proposition 2*. Therefore, for the AGA algorithm with $\Omega(t)$, if $\mathcal{S}_{\mathrm{PVS}} \neq \varnothing$, any subchannel whose LLR mean belongs to $\mathcal{S}_{\mathrm{PVS}}$ will inaccurately create two "good" channel in the

size-two channel polarization, which will lead to obvious approximation error in the subsequent AGA process.

*Definition 2:* Given the approximation function $\Omega(t)$, the polarization reversal set (PRS) $\mathcal{S}_{\mathrm{PRS}}$ is defined as

$$\mathcal{S}_{\mathrm{PRS}} = \left\{t \,\middle|\, \Omega^{-1}\left(1 - (1 - \Omega(t))^2\right) \geq 2t\right\}, \qquad (21)$$

where $t \in (0, +\infty)$.

Interestingly, in the size-two channel polarization, for any LLR mean $m$ belonging to $\mathcal{S}_{\mathrm{PRS}}$, $\Omega(t)$ will result in $m < m_2^{(2)} \leq m_2^{(1)}$. In other words, the split "good" channel and "bad" channel swap their roles due to the calculation error of $\Omega(t)$, which yields severe error in the size-two polarization. This phenomenon then leads to the substantial error in subchannels' position rank.[3]

*Proposition 4:* The relationship between PVS and PRS can be expressed as

$$\mathcal{S}_{\mathrm{PRS}} \neq \varnothing \Rightarrow \mathcal{S}_{\mathrm{PVS}} \neq \varnothing. \qquad (22)$$

*Proof:* Note that when $t \in (0, +\infty)$, for any $t \in \mathcal{S}_{\mathrm{PRS}}$, it makes $\Omega^{-1}(1 - (1 - \Omega(t))^2) \geq 2t$ by *Definition 2*, which will inevitably result in $\Omega^{-1}(1 - (1 - \Omega(t))^2) \geq t$. Therefore, according to *Definition 1*, the proof of *Proposition 4* is immediate. In other words, $\mathcal{S}_{\mathrm{PRS}} \neq \varnothing$ is the sufficient condition of $\mathcal{S}_{\mathrm{PVS}} \neq \varnothing$. On the contrary, if $\mathcal{S}_{\mathrm{PVS}} = \varnothing$, we can derive $\mathcal{S}_{\mathrm{PRS}} = \varnothing$. ∎

Suppose $\Omega(t)$ monotonically decreases on $(0, +\infty)$, the left inequality in (20) can therefore be simplified as

$$\Omega(t) \geq 1 - (1 - \Omega(t))^2 \Rightarrow \Omega(t) \geq 1. \qquad (23)$$

Analogously, the inequality in (21) will also be simplified as

$$1 - (1 - \Omega(t))^2 \leq \Omega(2t) \Rightarrow 2\Omega(t) - \Omega(t)^2 \leq \Omega(2t). \qquad (24)$$

*Example:* recall that in Chung's conventional AGA scheme, the two-segment approximation function $\varphi(t)$ is a specific form of $\Omega(t)$. Since $\varphi(0) = e^{0.0218} > 1$, $\varphi(t)$ cannot satisfy (16). For $\varphi(t)$, its corresponding PVS and PRS are denoted by $\mathcal{S}_{\mathrm{PVS}} = (a_1, a_2]$ and $\mathcal{S}_{\mathrm{PRS}} = (0, a_1]$, respectively. The boundary points $a_1$ and $a_2$ are given in the following equations

$$\begin{cases} 2\varphi(a_1) - \varphi(a_1)^2 = \varphi(2a_1), \\ \varphi(a_2) = 1, \end{cases} \Rightarrow \begin{cases} a_1 = 0.01476, \\ a_2 = 0.02939, \end{cases} \qquad (25)$$

which follows (23) and (24). Hence for $\varphi(t)$ we have

$$\mathcal{S}_{\mathrm{PVS}} = (0.01476, 0.02939], \mathcal{S}_{\mathrm{PRS}} = (0, 0.01476], \qquad (26)$$

which are denoted in Fig. 2.

*Theorem 1:* For the $N$-channel transform, where $N = 2^n$, $n \geq 1$, suppose that the original channel's LLR mean has two configurations, which are denoted by $\widehat{m}_0^{(1)}$ and $\widecheck{m}_0^{(1)}$. If they

---

[2]In this paper, we analyze the error between AGA and EGA, rather than the error of GA itself.

[3]One should notice that we cannot have $m_2^{(2)} < m$ since $m_2^{(2)} = 2m$. There just exist three orders, which correspond to (15), PVS and PRS.

J. Dai *et al.*: Does GA Work Well for the Long-Length Polar Code Construction?

IEEE *Access*



**FIGURE 2.** Schematic plot of $\varphi(t)$, where $\mathcal{S}_{\text{PVS}} = (0.01476, 0.02939]$ and $\mathcal{S}_{\text{PRS}} = (0, 0.01476]$.

satisfy $\widehat{m}_0^{(1)} \geq \widecheck{m}_0^{(1)}$, then under the GA assumption, for $j = 1, 2, \cdots, 2^n$, we have

$$\widehat{m}_n^{(j)} \geq \widecheck{m}_n^{(j)}. \tag{27}$$

*Proof:* This result will be proved by mathematical induction. Under the GA assumption, recall that $\phi(t)$ is continuous and monotonically decreasing on $[0, +\infty)$ [10], one can easily check that $f_c(t)$ and $f_v(t)$ in (8) monotonously increase on $[0, +\infty)$.

Suppose $N = 2^k$, $k = 1$, if the two LLR mean configurations of the original channel satisfy $\widehat{m}_0^{(1)} \geq \widecheck{m}_0^{(1)}$, then

$$\begin{cases} f_c\left(\widehat{m}_0^{(1)}\right) \geq f_c\left(\widecheck{m}_0^{(1)}\right), \\ f_v\left(\widehat{m}_0^{(1)}\right) \geq f_v\left(\widecheck{m}_0^{(1)}\right), \end{cases} \Rightarrow \begin{cases} \widehat{m}_1^{(1)} \geq \widecheck{m}_1^{(1)}, \\ \widehat{m}_1^{(2)} \geq \widecheck{m}_1^{(2)}. \end{cases} \tag{28}$$

With the increased LLR mean of the original channel $W$, it can be seen that the LLR means of two polarized subchannels will strictly increase.

Next, suppose $N = 2^k$, if we have $\widehat{m}_0^{(1)} \geq \widecheck{m}_0^{(1)}$, then for $j = 1, 2, \cdots, 2^k$, $\widehat{m}_k^{(j)} \geq \widecheck{m}_k^{(j)}$ holds. Thus, when $N = 2^{k+1}$, one can check

$$\begin{cases} f_c\left(\widehat{m}_k^{(j)}\right) \geq f_c\left(\widecheck{m}_k^{(j)}\right), \\ f_v\left(\widehat{m}_k^{(j)}\right) \geq f_v\left(\widecheck{m}_k^{(j)}\right), \end{cases} \Rightarrow \begin{cases} \widehat{m}_{k+1}^{(2j-1)} \geq \widecheck{m}_{k+1}^{(2j-1)}, \\ \widehat{m}_{k+1}^{(2j)} \geq \widecheck{m}_{k+1}^{(2j)}. \end{cases} \tag{29}$$

In other words, for $j = 1, 2, \cdots, 2^{k+1}$, $\widehat{m}_{k+1}^{(j)} \geq \widecheck{m}_{k+1}^{(j)}$. From above analysis, the proof of (27) is finished. ∎

*Theorem 1* indicates that under the GA assumption, if the LLR mean of the original channel $W$ increases, the polarized subchannel's LLR mean will increase together. Combining PVS and PRS, *Theorem 1* serves to analyze the subchannel's rank error and the poor performance of the long-length polar codes when Chung's conventional AGA is used.

For the AGA construction of polar codes with $\Omega(t)$, suppose $\mathcal{S}_{\text{PRS}} \neq \varnothing$, then for any LLR mean $m_i^{(j)} \in \mathcal{S}_{\text{PRS}}$, we have $m_{i+1}^{(2j-1)} \geq m_{i+1}^{(2j)}$ by the definition of $\mathcal{S}_{\text{PRS}}$ in the recursive calculation of GA. However, $\phi(t)$ in EGA can guarantee

$\mathcal{S}_{\text{PRS}} = \varnothing$ and $\mathcal{S}_{\text{PVS}} = \varnothing$, which claims $m_{i+1}^{(2j-1)} < m_{i+1}^{(2j)}$ for any $m_i^{(j)} > 0$. The above analysis indicates that in the AGA process, as presented in the code tree, due to $\mathcal{S}_{\text{PRS}} \neq \varnothing$, the condition $m_i^{(j)} \in \mathcal{S}_{\text{PRS}}$ will lead to the rank error among the leaf nodes which belong to the left subtree $\mathcal{T}_{\text{left}} = \mathcal{T}(v_{i+1}^{(2j-1)})$ and the right subtree $\mathcal{T}_{\text{right}} = \mathcal{T}(v_{i+1}^{(2j)})$. In other words, we have

$$m_n^{((j-1)2^{n-i}+s)} \geq m_n^{((j-1)2^{n-i}+s+2^{n-i-1})}, \tag{30}$$

where $s = 1, 2, \cdots, 2^{n-i-1}$. This result follows from *Theorem 1* by mechanically applying $m_{i+1}^{(2j-1)} \Rightarrow \widehat{m}_0^{(1)}$ and $m_{i+1}^{(2j)} \Rightarrow \widecheck{m}_0^{(1)}$. Then we have $\widehat{m}_{n-i-1}^{(s)} \geq \widecheck{m}_{n-i-1}^{(s)}$. Note that in these two subtrees, $\widehat{m}_{n-i-1}^{(s)}$ and $\widecheck{m}_{n-i-1}^{(s)}$ correspond to the left and right side in (30), respectively. However, in EGA, the "$\geq$" in (30) should be "$<$". This polarization reversal phenomenon leads to the rank error of subchannel's position, which directly affects the selection of information set $\mathcal{A}$.



**FIGURE 3.** A code tree representation of polarization violation and polarization reversal, where $N = 16$ and $n = 4$. $m_1^{(1)} \in \mathcal{S}_{\text{PRS}}$ and $m_3^{(5)} \in \mathcal{S}_{\text{PVS}}$.

As an example, Fig. 3 shows the polarization violation and polarization reversal in the code tree, where $N = 16$ and $n = 4$. In the AGA computation process, $m_1^{(1)} \in \mathcal{S}_{\text{PRS}}$. Then we have $m_2^{(1)} \geq m_2^{(2)}$, for the leaf nodes with $\mathcal{T}_{\text{left}} = \mathcal{T}(v_2^{(1)})$ and $\mathcal{T}_{\text{right}} = \mathcal{T}(v_2^{(2)})$, following *Theorem 2*, one can check

$$m_4^{(1)} \geq m_4^{(5)}, \quad m_4^{(2)} \geq m_4^{(6)},$$
$$m_4^{(3)} \geq m_4^{(7)}, \quad m_4^{(4)} \geq m_4^{(8)}. \tag{31}$$

Similarly, since $m_3^{(5)} \in \mathcal{S}_{\text{PVS}}$, we have $m_4^{(9)} \geq m_3^{(5)}$, whereas in EGA, it should be $m_4^{(9)} < m_3^{(5)}$, and all the "$\geq$" in (31) should be "$<$". Thus, compared to EGA, the approximation error of AGA results in obvious rank error within the 16 leaf nodes.

The above analysis indicates that in the AGA construction of polar codes, if $\mathcal{S}_{\text{PRS}} \neq \varnothing$, its approximation error leads to polarization reversal. Considering its definition, $\mathcal{S}_{\text{PRS}}$ lies in the vicinity of 0. As stated in [1], when $N$ tends to infinity, the symmetric capacity terms $\{I(W_N^{(j)})\}$ cluster around 0 and 1, and the corresponding LLR means cluster around 0 and $+\infty$. So when the code length becomes longer, there are more LLR

IEEE Access

J. Dai et al.: Does GA Work Well for the Long-Length Polar Code Construction?

means falling in $\mathcal{S}_{\mathrm{PRS}}$ during the AGA recursive computation process. This is the essential reason that polar codes constructed by some AGA suffer from catastrophic performance loss with the long code lengths.

During the recursive process of AGA, assuming the number of code tree nodes belonging to the two sets[4] are denoted by $\mu_{\mathrm{PVS}}$ and $\mu_{\mathrm{PRS}}$, the corresponding ratios with respect to all nodes are written as

$$\theta_{\mathrm{PVS}} = \frac{\mu_{\mathrm{PVS}}}{\sum\limits_{k=0}^{n-1} 2^k}, \quad \theta_{\mathrm{PRS}} = \frac{\mu_{\mathrm{PRS}}}{\sum\limits_{k=0}^{n-1} 2^k}. \tag{32}$$

For Chung's conventional AGA, Table 1 gives the distribution of $\mu_{\mathrm{PVS}}$ and $\mu_{\mathrm{PRS}}$ with different polarization levels $n$. The corresponding $\theta_{\mathrm{PVS}}$ and $\theta_{\mathrm{PRS}}$ are also listed in Table 1, where $E_b/N_0 = 1$dB ($\sigma^2 = 1.1915$). From Table 1, we observe that with the increase of code length, there are more and more nodes falling in PRS, and its corresponding ratio also becomes larger. Therefore, the polar codes constructed by Chung's conventional AGA suffer from catastrophic jitter in performance when their code lengths are long.

**TABLE 1.** The number of nodes and its percentage whose LLR mean belongs to $\mathcal{S}_{\mathrm{PVS}}$ and $\mathcal{S}_{\mathrm{PRS}}$ during the recursive process of Chung's conventional AGA, where $E_b/N_0 = 1$dB.

| $n$ | $\mu_{\mathrm{PVS}}$ | $\theta_{\mathrm{PVS}}$ | $\mu_{\mathrm{PRS}}$ | $\theta_{\mathrm{PRS}}$ |
|---|---|---|---|---|
| 10 | 40 | 3.910% | 33 | 3.226% |
| 11 | 89 | 4.348% | 88 | 4.299% |
| 12 | 191 | 4.664% | 225 | 5.495% |
| 13 | 394 | 4.810% | 549 | 6.702% |
| 14 | 803 | 4.901% | 1297 | 7.917% |
| 15 | 1617 | 4.935% | 3003 | 9.165% |
| 16 | 3280 | 5.005% | 6820 | 10.41% |
| 17 | 6340 | 4.837% | 15240 | 11.63% |
| 18 | 12528 | 4.779% | 33646 | 12.83% |
| 19 | 24550 | 4.683% | 73503 | 14.02% |
| 20 | 48036 | 4.581% | 159132 | 15.18% |

Fig. 4 demonstrates the BLER performance of polar codes constructed by Chung's conventional AGA and Tal&Vardy's method under the BI-AWGN channels. In Fig. 4, polar codes are constructed depending on the signal-to-noise ratio (SNR) one by one, and all the schemes have code rate $R = 1/3$ with SC decoding. The code length $N$ is set to be $2^{12}$, $2^{14}$ and $2^{18}$. We observe that for long code lengths, Chung's scheme obviously presents catastrophic performance loss. It is consistent with the analysis of Table 1.

*Remark 1:* For the AWGN channels, compared with the accurate DE or Tal&Vardy's algorithm, EGA is also found to well approximate the actual polarized subchannels. Note that EGA has strict order preserving property, which follows *Proposition 2*. This order preserving property of EGA indeed

[4]The leaf nodes whose LLR mean falls into the two sets are not counted in, because it will not lead to rank error among the descendants.
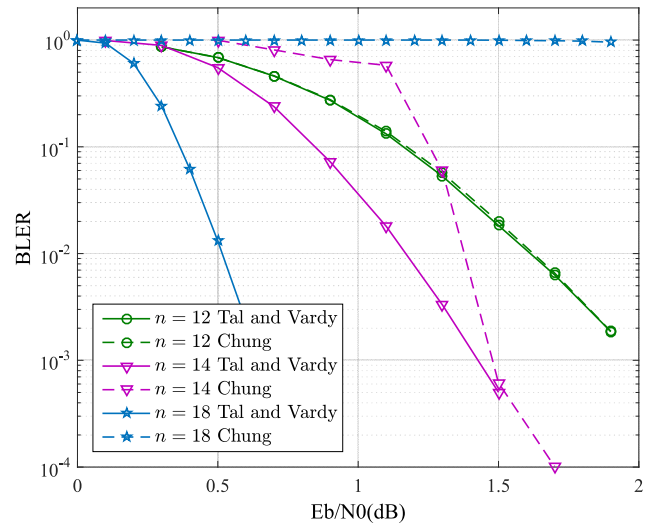


**FIGURE 4.** BLER performance comparison of polar codes with the code length $N = 2^n$ ($n = 12, 14, 18$) and code rate $R = 1/3$ in the AWGN channel.

gives reasonable interpretation about its "good results" versus DE. Therefore, in general, the error between EGA and DE is so small that it can be ignored.

*Remark 2:* Recall that Arıkan in [3] suggested a heuristic BEC approximation method to construct the polar codes for arbitrary binary-input channels, which has also yielded good results in experiments. The above PVS and PRS analyses also give interpretation about this "good results". One can check BEC approximation has strict order preserving property in size-two polarization transform, which shows $I(W_2^{(1)}) < I(W) < I(W_2^{(2)})$. Heuristic BEC approximation will not lead to polarization violation and polarization reversal in this sense. Therefore, this heuristic method will just bring some moderate performance loss rather than catastrophic jitter.

## B. CLE OF CHANNEL POLARIZATION

Polarization violation and polarization reversal reveal the essential reason that AGA cannot work well for long code lengths. Besides these two sets, in this subsection, we further propose a new metric, named cumulative-logarithmic error (CLE) of channel polarization, to quantitatively evaluate the remainder approximation error between AGA and EGA. Then CLE is utilized to guide the design of GA approximation function $\Omega(t)$.

For $\Omega(t)$ in AGA, suppose its $\mathcal{S}_{\mathrm{PVS}} = \varnothing$ and $\mathcal{S}_{\mathrm{PRS}} = \varnothing$, CLE will play a crucial role in evaluating the performance of AGA. We concern about the subchannel's capacity, which is a function of LLR mean under GA assumption. Hence, the difference between $\Omega(t)$ and $\phi(t)$ brings in calculation error in subchannel capacity evaluation. The original absolute error of capacity calculation between AGA and EGA is denoted by $\Delta(t)$, which is a function of LLR mean $t$. Without ambiguity, $\Delta(t)$ will be abbreviated to $\Delta$ in this paper. Assume $\Delta$ occurs after $r$ recursions, denoted by $\Delta_r$. Thus $\Delta_r$ is accumulated as final error after $n - r$ polarization

J. Dai *et al.*: Does GA Work Well for the Long-Length Polar Code Construction?

IEEE *Access*

levels, and this process can be represented on a subtree with a depth $n - r$. To evaluate the calculation error of AGA, we focus on the difference of subchannel's capacity calculated by AGA and EGA in logarithmic domain. The capacities calculated by EGA can form a set $\mathcal{I}$ defined on this code subtree with the following properties:

On this subtree, the set of capacities corresponding to the nodes at a given depth $d$ is denoted by $\mathcal{I}_d, d = r, r+1, \cdots, n$. Let $I_d^{(k)}, k = 1, 2, \cdots, 2^{d-r}$, denote the $k$-th element in $\mathcal{I}_d$. For each $I_d^{(k)} \in \mathcal{I}_d$, $I_d^{(k)}$ takes value on $[0, 1]$. For $d > r$, $I_d^{(k)}$ is a function of path $b_{r+1}^d$, which is actually the binary expansion of $k - 1$ (i.e., $k - 1 = \sum_{i=1}^{d-r} b_{r+i} 2^{d-r-i}$). Therefore, at the root node $v_r^{(j)}$, $\Delta_r$ can be written as

$$\Delta_r = \tilde{I}_r^{(1)} - I_r^{(1)} = h\left(\frac{2}{\tilde{m}_r^{(j)}}\right) - h\left(\frac{2}{m_r^{(j)}}\right), \quad (33)$$

where $\tilde{I}_r^{(1)}$, $I_r^{(1)}$ and $\tilde{m}_r^{(j)}$, $m_r^{(j)}$ stand for the initial symmetric capacities and LLR means calculated by AGA and EGA respectively, and the formula of $h(\cdot)$ is written in (13).

As stated in *Remark 2*, without much sacrifice in accuracy, BEC approximation will act as faithful surrogate for GA in error analysis. According to the iteration structure in channel polarization transform in [1], we can derive as follows

$$\begin{cases} I_{d+1}^{(2k-1)} = \left(I_d^{(k)}\right)^2 & \text{when } b_{d+1} = 0, \\ I_{d+1}^{(2k)} = 2I_d^{(k)} - \left(I_d^{(k)}\right)^2 & \text{when } b_{d+1} = 1. \end{cases} \quad (34)$$

Furthermore, when $b_{d+1} = 1$, we have $I_{d+1}^{(2k)} \leq 2I_d^{(k)}$. Thus, in logarithmic domain, we can get

$$\begin{cases} \log I_{d+1}^{(2k-1)} = 2\log I_d^{(k)} & \text{when } b_{d+1} = 0, \\ \log I_{d+1}^{(2k)} \leq \log I_d^{(k)} + 1 & \text{when } b_{d+1} = 1. \end{cases} \quad (35)$$

Define $\tilde{I}_d^{(k)} = I_d^{(k)} + \Delta_d^{(k)}$, where $\tilde{I}_d^{(k)}$ denotes the capacity corresponding to AGA, $\Delta_d^{(k)}$ represents the absolute error of capacity calculation between AGA and EGA. For $r < d \leq n$, $\tilde{I}_d^{(k)}$ and $I_d^{(k)}$ represent the capacities, which are calculated by BEC approximation, of AGA and EGA respectively. In this paper, We only analyze the error between AGA and EGA, rather than the error of GA itself or the error brought in by heuristic BEC approximation. Therefore, $\Delta_r^{(1)} = \Delta_r$. Let $\rho_d^{(k)} = \Delta_d^{(k)}/I_d^{(k)}$ denote the relative error, and $e_d^{(k)} = \log \tilde{I}_d^{(k)} - \log I_d^{(k)} = \log\left(1 + \rho_d^{(k)}\right)$ denote the capacity calculation error in logarithmic domain. Hence, the partial cumulative-logarithmic error (PCLE) can be written as

$$C_{r:n} = \sum_{k=1}^{2^{n-r}} \left|e_n^{(k)}\right|. \quad (36)$$

The cumulative-logarithmic error (CLE) will be $C = \sum_r C_{r:n}$.

## C. CLE BOUND

The precise calculation of CLE is too complicated to be analyzed using recursive relation (34). In this subsection, we propose an upper bound on CLE to simplify its calculation.

*Definition 3:* For the $k$-th leaf node corresponding to a path $b_{r+1}^n$, we define $|\{r + 1 \leq i \leq n : b_i = 0\}| = \alpha$ and $|\{r + 1 \leq i \leq n : b_i = 1\}| = n - r - \alpha$.

*Theorem 2:* The logarithm error $\left|e_n^{(k)}\right|$ can be bounded by

$$\left|e_n^{(k)}\right| \leq 2^\alpha \left|\log\left(1 + \rho_r^{(1)}\right)\right| = 2^\alpha \left|\log\left(1 + \frac{\Delta_r^{(1)}}{I_r^{(1)}}\right)\right|.$$

*Proof:* Let $\tilde{e}_d^{(k)}$ denote the bound of $e_d^{(k)}$. Then, $\tilde{e}_n^{(k)}$ is determined by specifying $\tilde{e}_r^{(1)} = e_r^{(1)} = \log\left(1 + \rho_r^{(1)}\right)$ and

$$\begin{cases} \tilde{e}_{d+1}^{(2k-1)} = D\left(\tilde{e}_d^{(k)}\right) & \text{when } b_{d+1} = 0, \\ \tilde{e}_{d+1}^{(2k)} = E\left(\tilde{e}_d^{(k)}\right) & \text{when } b_{d+1} = 1, \end{cases} \quad (37)$$

where $E : \mathbb{R} \to \mathbb{R}$, $E(z) = z$ denotes equality, $D : \mathbb{R} \to \mathbb{R}$, $D(z) = 2z$ denotes doubling.

If $\Delta_r^{(1)} \geq 0$, it claims that $\Delta_d^{(k)} \geq 0$ holds in terms of *Theorem 1*. Note that during the iteration, when $b_{d+1} = 0$,

$$0 \leq e_{d+1}^{(2k-1)} = 2e_d^{(k)} = \tilde{e}_{d+1}^{(2k-1)}. \quad (38)$$

When $b_{d+1} = 1$, it can be proved that $e_{d+1}^{(2k)} \leq \tilde{e}_{d+1}^{(2k)}$. According to the first equation of (34), it is easy to get that

$$\begin{aligned} \tilde{I}_{d+1}^{(2k)} &= 2\tilde{I}_d^{(k)} - \left(\tilde{I}_d^{(k)}\right)^2 \\ &= 2\left(I_d^{(k)} + \Delta_d^{(k)}\right) - \left(I_d^{(k)} + \Delta_d^{(k)}\right)^2 \\ &= \underbrace{2I_d^{(k)} - \left(I_d^{(k)}\right)^2}_{=I_{d+1}^{(2k)}} + \underbrace{2\Delta_d^{(k)} - 2I_d^{(k)}\Delta_d^{(k)} - \left(\Delta_d^{(k)}\right)^2}_{=\Delta_{d+1}^{(2k)}}. \end{aligned} \quad (39)$$

Therefore, $e_{d+1}^{(2k)}$ can be written as

$$\begin{aligned} e_{d+1}^{(2k)} &= \log\left(1 + \rho_{d+1}^{(2k)}\right) \\ &= \log\left(1 + \frac{2\Delta_d^{(k)} - 2I_d^{(k)}\Delta_d^{(k)} - \left(\Delta_d^{(k)}\right)^2}{2I_d^{(k)} - \left(I_d^{(k)}\right)^2}\right). \end{aligned} \quad (40)$$

In addition, we have

$$\tilde{e}_{d+1}^{(2k)} = \log\left(1 + \tilde{\rho}_{d+1}^{(2k)}\right) = e_d^{(k)} = \log\left(1 + \frac{\Delta_d^{(k)}}{I_d^{(k)}}\right). \quad (41)$$

Then we can check

$$\tilde{\rho}_{d+1}^{(2k)} - \rho_{d+1}^{(2k)} = \frac{\Delta_d^{(k)}\left(I_d^{(k)} + \Delta_d^{(k)}\right)}{I_d^{(k)}\left(2 - I_d^{(k)}\right)} \geq 0. \quad (42)$$

It can be inferred that $\tilde{e}_d^{(k)} \geq e_d^{(k)} \geq 0$ holds.

**IEEE** *Access*

J. Dai et al.: Does GA Work Well for the Long-Length Polar Code Construction?

Recall *Definition 3*, during the iterative calculation of $\tilde{e}_n^{(k)}$, we count doubling $\alpha$ times and equality $n - r - \alpha$ times. Hence, we have

$$0 \leq e_n^{(k)} \leq \tilde{e}_n^{(k)} = E^{n-r-\alpha} D^\alpha \left( e_r^{(1)} \right) = 2^\alpha e_r^{(1)}. \quad (43)$$

Analogously, if $\Delta_r^{(1)} < 0$, we have $\Delta_d^{(k)} < 0$. From (38) and (42), we can get

$$0 > e_n^{(k)} > \tilde{e}_n^{(k)} = E^{n-r-\alpha} D^\alpha \left( e_r^{(1)} \right) = 2^\alpha e_r^{(1)}. \quad (44)$$

Combing (43) and (44), we prove the theorem. ∎

*Theorem 3:* PCLE $C_{r:n}$ can be upper bounded by

$$C_{r:n} \leq 3^{n-r} \left| \log \left( 1 + \rho_r^{(1)} \right) \right| = 3^{n-r} \left| \log \left( 1 + \frac{\Delta_r^{(1)}}{I_r^{(1)}} \right) \right|.$$

*Proof:* For any $k \in \{1, 2, \cdots, 2^{n-r}\}$, the number of vectors $b_{r+1}^n$ satisfying *Definition 3* is $\binom{n-r}{\alpha}$, where $b_{r+1}^n$ is the binary expansion of $k - 1$. Combined with definition (36) and *Theorem 2*, $C_{r:n}$ satisfies the following constraint

$$C_{r:n} \leq \sum_{k=1}^{2^{n-r}} \left| \tilde{e}_n^{(k)} \right| = \sum_{\alpha=0}^{n-r} \binom{n-r}{\alpha} 2^\alpha \left| e_r^{(1)} \right| = 3^{n-r} \left| e_r^{(1)} \right|. \quad (45)$$

The last equation in (45) uses binomial theorem. Therefore, CLE $C$ can be upper bounded by $\sum_r 3^{n-r} \left| \log_2 \left( 1 + \rho_r^{(1)} \right) \right|$, and the exponent $n - r$ stands for polarization levels. ∎

## V. IMPROVED GAUSSIAN APPROXIMATION

In this section, guided by the previous PVS, PRS and CLE analyses, we propose new rules to design AGA for the construction of polar codes. Then we give three specific forms of the approximation function in AGA, which have advantages in both complexity and performance.

### A. DESIGN RULES OF AGA

For channels other than BEC (e.g. AWGN channel), AGA is widely used to construct polar codes. However, in practical implementation, the accuracy of approximation function $\Omega(t)$ will greatly affect the construction of polar codes especially when the code length is long. According to *Theorem 3*, the initial error will be exponentially amplified with the increase of polarization levels. Note that PCLE bound is mainly affected by two factors: the first term $3^{n-r}$ is relevant to polarization levels, and the second term is dependent on the original relative error $\rho_r^{(1)} = \Delta_r^{(1)}/I_r^{(1)}$. In general, the absolute error $\Delta_r^{(1)}$ is tiny. Therefore, for the good channels whose capacities $I_r^{(1)}$ approach 1, their original relative errors $\rho_r^{(1)}$ are so small that they can be ignored. However, for the bad channels whose capacities $I_r^{(1)}$ approach 0, their original relative errors are not negligible. Subsequently, given a fixed absolute error $\Delta_r^{(1)}$, the more $I_r^{(1)}$ is close to 0, the larger the original relative error $\rho_r^{(1)}$ becomes.

Above analysis indicates that CLE bound is mainly affected by the terms $C_{r:n}$ with the small initial capacity $I_r^{(1)}$, which corresponds to the bad channels. Due to the error, some

frozen subchannels will be wrongly identified as information-carrying ones (role flipping), which results in the performance degradation. Note that the capacity $I_r^{(1)}$ monotonically increases with LLR mean $t$ under the GA assumption. In addition, according to (14), we have

$$\lim_{t \to 0} I_r^{(1)} = 0, \qquad \lim_{t \to +\infty} I_r^{(1)} = 1. \quad (46)$$

Guided by above analyses, the AGA approximation function design is composed of three rules:

Rule 1) PVS and PRS eliminating: $\Omega(t)$ should guarantee $\mathcal{S}_{\text{PVS}} = \varnothing$ and $\mathcal{S}_{\text{PRS}} = \varnothing$. According to *Proposition 4* and its converse-negative proposition, if $\mathcal{S}_{\text{PVS}} = \varnothing$, we have $\mathcal{S}_{\text{PRS}} = \varnothing$. Hence, in order to empty PVS and PRS, we should ensure $0 < \Omega(t) < 1$ for any $t \in (0, +\infty)$.

Rule 2) Low SNR design: When $t$ comes close to 0, we must guarantee $\lim_{t \to 0} \Omega(t) = 1$ to reduce approximation error. Since CLE bound is amplified exponentially with the growth of polarization levels, the only way to reduce CLE bound is to lower the initial relative error $\rho_r^{(1)}$. Therefore, $\Omega(t)$ needs to be divided into more segments when $t$ approaches 0. This rule can reduce the original absolute error $\Delta_r^{(1)}$ in the vicinity of 0 so as to lower $\rho_r^{(1)}$.

Rule 3) High SNR design: When $t$ stays away from 0, thanks to the relatively large $I_r^{(1)}$, CLE bound can tolerate a more obvious absolute error $\Delta_r^{(1)}$. Therefore, $\Omega(t)$ can be selected with some simpler forms to reduce the computational complexity.

Besides, $\Omega(t)$ should keep continuity between the adjacent two segments, which can mitigate the jitter of CLE bound by keeping smooth of initial error. Among these three rules, Rule 1 is the most crucial, which can prevent the corresponding AGA polar code construction scheme from severe performance loss. Then the importance of Rule 2 is second, which reduces the remainder approximation error between AGA and EGA. Rule 3 plays a less important role, which helps to further reduce the computational complexity of AGA.

### B. IMPROVED AGA APPROXIMATION FUNCTION

Recall the analysis in Section IV, Chung's two-segment AGA scheme leads to $\mathcal{S}_{\text{PVS}} \neq \varnothing$ and $\mathcal{S}_{\text{PRS}} \neq \varnothing$, which violates Rule 1. Therefore, guided by above rules, we design the following new two-segment approximation function by employing curve-fitting algorithm with minimum mean square error criterion, denoted by $\Omega_2(t)$,

$$\Omega_2(t) = \begin{cases} e^{0.0116t^2 - 0.4212t} & 0 < t \leq a, \\ e^{-0.2944t - 0.3169} & a < t, \end{cases} \quad (47)$$

where the boundary point $a = 7.0633$. The corresponding new AGA algorithm is denoted by AGA-2. In the construction of polar codes, compared with Chung's scheme, AGA-2 can eliminate rank error and therefore will not lead to catastrophic performance loss at the long code lengths. In addition, since

J. Dai *et al.*: Does GA Work Well for the Long-Length Polar Code Construction?

IEEE*Access*

the inverse function of $\Omega_2(t)$ can be analytically obtained, it has lower calculating complexity. Furthermore, guided by Rule 3, for $f_c(t)$ in (8), when $t$ leaves away from 0, $1 - \Omega_2(t)$ tends to 1, which indicates

$$(1 - \Omega_2(t))^2 \approx 1 - \Omega_2(t) \Rightarrow f_c(t) \approx t. \tag{48}$$

Followed by *Proposition 2*, $f_c(t)$ should satisfy $f_c(t) < t$. Thus, when $t$ stays away from 0, the complex $f_c(t)$ can be further approximated as

$$f_c(t) = t - \varepsilon, \tag{49}$$

where $\varepsilon$ denotes the offset. Then according to (8), when $\Omega_2(t)$ tends to 0, one can check

$$\Omega_2(t - \varepsilon) = 2\Omega_2(t) - \Omega_2(t)^2 \approx 2\Omega_2(t). \tag{50}$$

Therefore, for $\Omega_2(t)$, in terms of (47), when $t \gg 0$ we have

$$e^{-0.2944(t-\varepsilon)-0.3169} = e^{-0.2944t-0.3169+\ln 2}$$

$$\Rightarrow \varepsilon = 2.3544, \tag{51}$$

where $t - \varepsilon$ should locate in the sencond segment, namely $t - \varepsilon > a$, which claims $t > 9.4177$. Following Rule 3, for the entire AGA-2 scheme, its $f_c(t)$ is further simplified as

$$f_c(t) = \begin{cases} \Omega_2^{-1}\left(1 - (1 - \Omega_2(t))^2\right) & 0 \leq t \leq \tau, \\ t - 2.3544 & t > \tau, \end{cases} \tag{52}$$

where the boundary point is $\tau = 9.4177$.

From the subsequent CLE analysis, we find that although AGA-2 can satisfy Rule 1 so as to remove the rank error, it will still bring obvious approximation error in the vicinity of 0. Therefore, when the code length is relatively long, AGA-2 will also bring some moderate performance loss. In order to further improve the GA construction performance, we propose a new piecewise function $\Omega_3(t)$ with three segments by using curve-fitting algorithm, that is

$$\Omega_3(t) = \begin{cases} e^{0.06725t^2 - 0.4908t} & 0 < t \leq a, \\ e^{-0.4527t^{0.86} + 0.0218} & a < t \leq b, \\ e^{-0.2832t - 0.4254} & b < t, \end{cases} \tag{53}$$

where the boundary points $a = 0.6357$ and $b = 9.2254$. Its corresponding AGA algorithm is denoted by AGA-3. It is specially designed for polar codes, which follows the proposed rules. AGA-3 can better satisfy Rule 1 and Rule 2. Especially in the third segment, namely the high SNR region, AGA-3 has lower computational complexity compared to Chung's scheme which obeys Rule 3. In addition, the inverse function of $\Omega_3(t)$ can be easily obtained. Similar with AGA-2, for the whole AGA-3 scheme, its $f_c(t)$ can be further simplified as

$$f_c(t) = \begin{cases} \Omega_3^{-1}\left(1 - (1 - \Omega_3(t))^2\right) & 0 \leq t \leq \tau, \\ t - 2.4476 & t > \tau, \end{cases} \tag{54}$$

where the boundary point is $\tau = 11.673$.

There is no doubt that, if the code length becomes extremely long, the three-segment approximation function

in AGA-3 will also bring calculation error, which obey *Theorem 3* and is stated in Rule 2. Hence, we design the following four-segment improved AGA scheme for extremely long polar code, that is

$$\Omega_4(t) = \begin{cases} e^{0.1047t^2 - 0.4992t} & 0 < t \leq a, \\ 0.9981 \cdot e^{0.05315t^2 - 0.4795t} & a < t \leq b, \\ e^{-0.4527t^{0.86} + 0.0218} & b < t \leq c, \\ e^{-0.2832t - 0.4254} & c < t, \end{cases} \tag{55}$$

where the boundary points $a = 0.1910$, $b = 0.7420$ and $c = 9.2254$. The approximation accuracy in the vicinity of 0 is further improved. Its corresponding AGA algorithm is denoted by AGA-4. The inverse function of $\Omega_4(t)$ can also be analytically obtained. Since the last two segments in $\Omega_4(t)$ are the same as that in $\Omega_3(t)$, the $f_c(t)$ function in AGA-4 also has the same form as the equation (54).

The computational complexity of $f_c(t)$ has been remarkably reduced thanks to Rule 3 in the above three proposed AGA schemes. Moreover, Rule 1 and Rule 2 help AGA to achieve excellent performance. Thus, for the general construction of polar codes, the proposed AGA-2 scheme is a good alternate to improve the Chung's conventional AGA scheme. When the code length becomes longer, the proposed AGA-3 scheme will achieve better performance than AGA-2 scheme. If the code length becomes extremely long, AGA-4 will present better performance. When the code length continues to grow, AGA-4 will inevitably bring calculation error, which follows *Theorem 3* and Rule 2. Nevertheless, followed by Rule 1 to 3, we can still design AGA with more complex multi-segment approximation function to keep the calculation accuracy. In this way, we provide a systematic framework to design high accuracy AGA scheme for polar codes at any code length.

As for the specific forms of the new AGA approximation functions, it is heuristically obtained from Chung's original $\varphi(t)$ while taking the Rule 1 to 3 into consideration. The first segment should ensure $\lim_{t \to 0} \Omega(t) = 1$ and use some complex functions to reduce the relative error. When $t$ stays away from 0, $\Omega(t)$ can be selected with some simpler forms to reduce the computational complexity. After choosing the approximation form, its parameters are acquired by curve-fitting. Certainly, one can choose different approximation functions according to the proposed rules. Finally, the performance will be evaluated by calculating its CLE bound.

For other channels $\overline{W}$ rather than the AWGN (e.g. the Rayleigh fading channel), one may expect better performance at the expense of more complexity in the code construction by using DE. However, we can approximate the channel $\overline{W}$ using an AWGN channel $W$ with $\sigma$, where

$$I(\overline{W}) = I(W) = h(\sigma^2). \tag{56}$$

The code construction is then then performed over each of the equivalent AWGN channels in the same GA way as that in the conventional AWGN case. As that will be shown in Section VI, the proposed AGA based construction of polar

IEEE Access

J. Dai *et al.*: Does GA Work Well for the Long-Length Polar Code Construction?

codes works well for other channels as well. This is significant in that it shows the robustness of the AGA construction against uncertainty and variation in channel parameters.

### C. COMPLEXITY COMPARISON

In this part, we compare the order of complexities with four typical polar code construction methods, which are Arıkan's heuristic BEC approximation method, DE algorithm, Tal and Vardy's method and our proposed AGA schemes. Note that the complexity order provided only includes the dominant terms, and the detailed number of calculation depends heavily on the specific hardware implementation, which is beyond the scope of this paper.

One can check the principle of these four methods is to recursively calculate each subchannel's reliability on the code tree. Therefore, their the total number of node visiting complexity can be written as $O(N) + O(N/2) + O(N/4) + \cdots + O(2) + O(1) = O(N)$. It seems that these four methods have the same number of node visiting. Furthermore, we analyze the computational complexity of each visiting.

- For the heuristic BEC approximation, the computational complexity of each visiting operation is $O(1)$, which can be ignored. Hence, it has the lowest complexity among the four schemes.
- DE needs a fast Fourier transform (FFT) and an inverse fast Fourier transform (IFFT) operation when calculating the probability density function (PDF) of LLR of each bit. The corresponding complexity is $O(\xi \log \xi)$, where $\xi$ denotes the number of samples. However, a typical value of $\xi$ is about $10^5$ [13], implying a huge computational burden in practical application. The difficulty is further aggravated by the quantization errors, which will be accumulated over multiple polarization levels.
- Tal and Vardy's method can be viewed as an approximate version of DE [5] and has a complexity of $O(\mu^2 \log \mu)$, where $\mu$ is a fixed, even and positive integer independent of code length $N$. In general, a typical value of $\mu$ is 256 which is much less than $\xi$. Hence, this method has much lower complexity than DE. But when the code length becomes long, Tal and Vardy's method still involves relatively high computational complexity.
- Regarding the proposed AGA method, the complexity of calculating $f_v(t)$ can be ignored since the result can be obtained easily. For $f_c(t)$, when $t > \tau$, its calculating complexity can also be ignored. When $0 \leq t \leq \tau$, since the inverse function in $f_c(t)$ can be analytically obtained, the computational complexity is $O(1)$. From these comparison, it can be concluded that our proposed AGA schemes have similar calculating complexity order with the heuristic BEC approximation, which are much lower than DE or Tal and Vardy's algorithms.

## VI. PERFORMANCE EVALUATION

The precise results and corresponding upper bounds of CLE for various AGAs are depicted in Fig. 5, where the polarization level $n = 8$ and $r = 0$ (CLE $C = C_{0:8}$). It can
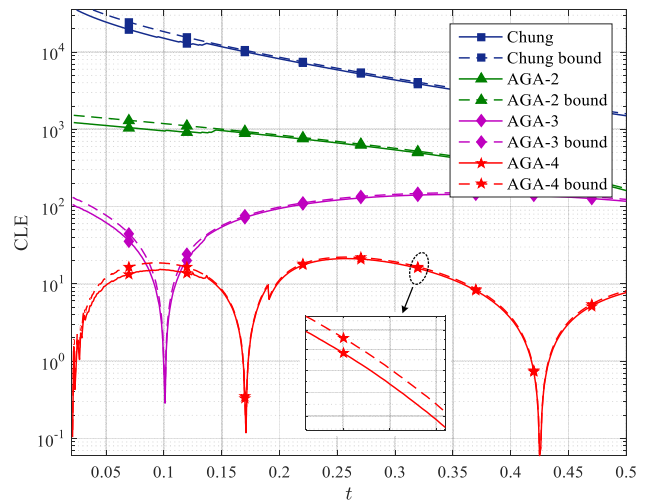


**FIGURE 5.** The precise results and corresponding upper bounds of CLE for various AGAs, where *t* stands for LLR mean and polarization level *n* = 8.

be found that the CLE bound and the exact result coincide well. Therefore, the CLE bound may be used as an effective tool to evaluate the performance of various AGA schemes. From Fig. 5, we can see that the CLE bound of Chung's conventional AGA scheme is obviously higher than that of AGA-2 and AGA-3, which indicates the poor performance of polar codes when Chung's AGA scheme is used. Compared with AGA-2, the CLE bound of AGA-3 also shows some performance gain. Furthermore, AGA-4 can achieve the best performance among these AGA schemes. We notice that there exists a non-monotonic behavior for the CLE bound. This phenomenon is caused by function fitting because there is different relative error $\rho_r^{(1)}$ for different $t$, which results in the jitter of CLE bound.
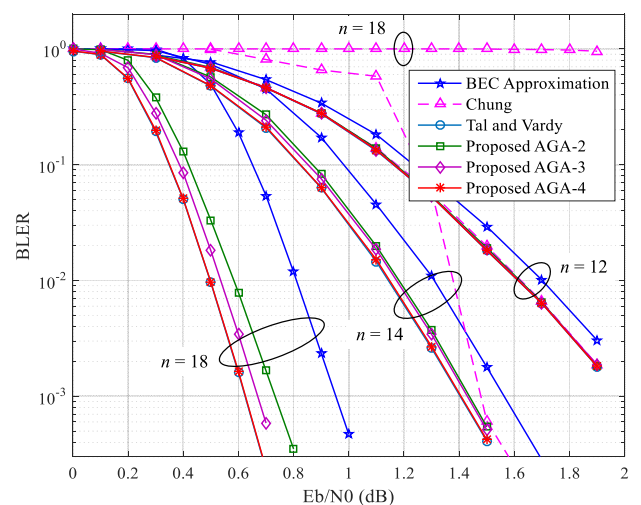


**FIGURE 6.** The SC decoding BLER performance comparison of polar codes with the code length $N = 2^n$ ($n = 12, 14, 18$) and the code rate $R = 1/3$ over the BI-AWGN channel.

Next we compare the BLER performance among different construction schemes under the BI-AWGN channel, which is shown in Fig. 6. All the schemes have code rate $R = 1/3$

J. Dai *et al.*: Does GA Work Well for the Long-Length Polar Code Construction?

IEEE *Access*

with the SC decoding. The code lengths $N$ are set to $2^{12}$, $2^{14}$ and $2^{18}$, respectively. From these results, it can be found that the BEC approximation scheme demonstrates some moderate performance loss compared with other advanced construction methods. For the extremely long code lengths, since DE falls into a huge computational burden in the practical application, we use Tal and Vardy's method as an alternate with almost no performance loss. Therefore, Tal and Vardy's construction possesses the highest accuracy in Fig. 6. However, it still requires high computational complexity at long code lengths. Among the AGA schemes, we can see that Chung's scheme suffers from a dramatic performance loss with the increase of code length since it violates Rule 1. On the contrast, our proposed AGA-2 achieves good performance. Hence, AGA-2 can be employed as a good alternate for Chung's two-segment method at some moderate code lengths. It can also be observed that the proposed AGA-3 scheme achieve better performance which follows Rule 1 and Rule 2. Furthermore, we can see that AGA-4 scheme approaches the performance of Tal and Vardy's method, which can be used for some extremely long code length. In addition, with the increase of code length, the performance gap between AGA-3 and AGA-4 becomes larger. Therefore, in terms of the previous CLE bound analyses and simulation results, it can be predicted that the performance gap will become more and more obvious with the increase of code length.
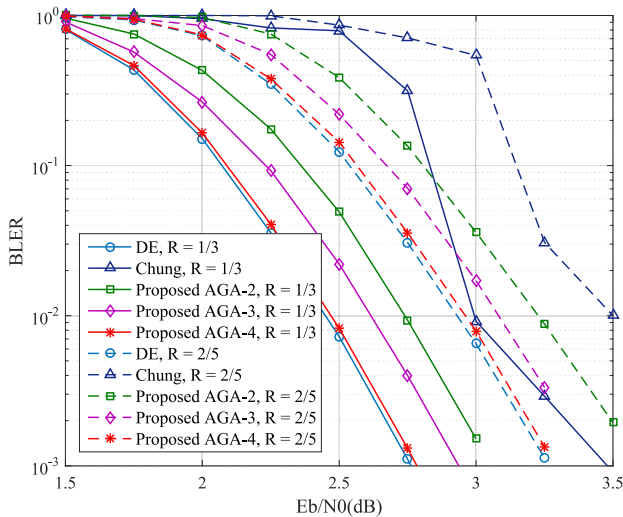


**FIGURE 7.** The SC decoding BLER performance comparison of polar codes with code length $N = 2^{14} = 16384$ and code rates $R = 1/3$ and $R = 2/5$ over the Rayleigh fading channel.

The comprehensive BLER performance comparisons with the SC decoding in terms of different construction schemes under the Rayleigh fading channels are given in Fig. 7. The code length is $N = 2^{14}$, and the code rates $R$ are set to 1/3 and 2/5, respectively. Similar with Fig. 6, we can see that DE algorithm achieves the best performance. Chung's scheme also presents poor performance since it cannot satisfy Rule 1. On the contrary, AGA-4 scheme suffers from an ignorable loss of performance compared with DE algorithm.

Furthermore, it becomes more computationally efficient and implementable in practical use than the former. In addition, AGA-4 can stably achieve $0.15 \sim 0.2$dB gain for different code rates compared with AGA-3 scheme. These results shows the robustness of the AGA construction against uncertainty and variation in channel parameters, which is valuable for polar coding.
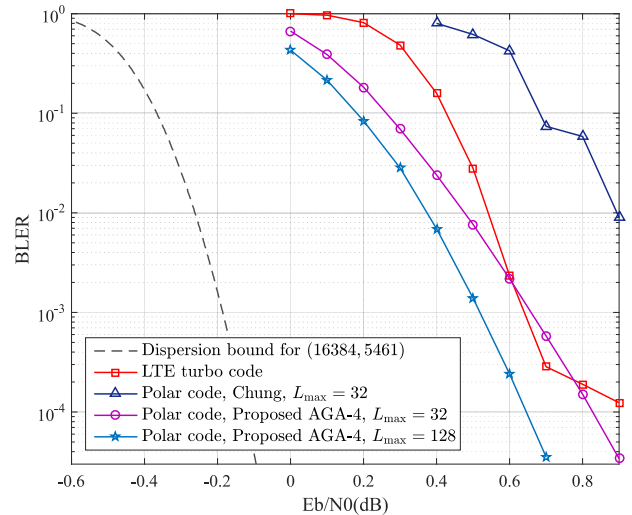


**FIGURE 8.** The Ad-CASCL decoding BLER performance comparison of polar code and LTE turbo code with the code length $N = 2^{14} = 16384$ and the code rates $R = 1/3$. The channel is configured as the BI-AWGN.

In Fig. 8, we provide the BLER performance of polar codes by using the adaptive cyclic redundancy check (CRC) aided SC list decoding (Ad-CASCL) [14], [15]. The maximum list size in the Ad-CASCL decoder is denoted by $L_{max}$. The 16-bit CRC in LTE standard [16] is used. The performance of LTE turbo code is also given as a comparison, where the Log-MAP decoding is applied in the turbo decoder with 8 iterations [17]. We can see that Chung's conventional AGA scheme shows poor performance. It indicates that this traditional two-segment AGA scheme is not suitable for the polar code construction. On the contrary, the polar codes constructed by the proposed AGA-4 scheme perform well. When $L_{max} = 32$, polar code can outperform LTE turbo code in the low SNR region. For the high SNR region, although turbo code sometimes performs better than polar code, it suffers from the error floor. When $L_{max}$ is set to 128, polar code can outperform LTE turbo code for any SNR. Additionally, this polar code constructed by AGA-4 scheme with $L_{max} = 128$ Ad-CASCL decoding can achieve BLER $\leq 10^{-3}$ at $E_b/N_0 = 0.51$dB. We compare this performance to the Shannon limit at the same finite block length, which is provided in [18]. The maximum rate that can be achieved at block length $N$ and BLER $\epsilon$ can be well approximated by

$$R_{max} = C - \sqrt{\frac{V}{N}} Q^{-1}(\epsilon), \qquad (57)$$

where $C$ is the channel capacity and $V$ is a quantity called the channel dispersion that can be computed from the channel

IEEE *Access*

J. Dai *et al.*: Does GA Work Well for the Long-Length Polar Code Construction?

statistics, using the formula:

$$V = \mathrm{Var}\left[\log \frac{W(Y|X)}{W(Y)}\right]. \tag{58}$$

For the BI-AWGN channel, the transition probability $W(y|x)$ is written in (4). The channel dispersion $V$ is written as

$$V = \frac{1}{2}\sum_{x\in\mathbb{B}}\int_{\mathbb{R}} W(y|x)\left[\log\left(\frac{2W(y|x)}{W(y|0)+W(y|1)}\right)\right]^2 dy \\ - I^2(W). \tag{59}$$

By using (57), we can calculate the Shannon limit for the $(N, K) = (16384, 5461)$ code which is named as the "dispersion bound" in Fig. 8. To achieve a rate $R = 5461/16384 = 1/3$, the minimum $E_b/N_0$ required is $-0.186$dB. Hence, polar code constructed by AGA-4 with $L_{\max} = 128$ is 0.696dB from the Shannon limit. When $L_{\max}$ increases, this SNR gap will become smaller.

## VII. CONCLUSION

In this paper, we introduced the concepts of PVS and PRS which explain the essential reason that polar codes constructed by conventional AGA expresses poor performance at long code lengths. Then we proposed a new metric, named CLE, to quantitatively evaluate the remainder error of AGA. We further derived the upper bound of CLE to simplify its calculation. Guided by PVS, PRS and CLE bound analysis, we proposed new rules to design AGA for polar codes. Simulation results show that the performance of all AGA schemes is consistent with CLE analysis. When the polarization levels increase, Chung's conventional AGA scheme suffers from a catastrophic performance jitter. On the contrary, the proposed AGA methods guided by the proposed rules stably guarantee the excellent performance of polar codes for both the AWGN channels and the Rayleigh fading channels.

## 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, Jul. 2009.

[2] *Final Report of 3GPP TSG RAN WG1 #87 v1.0.0*, document R1-1701552, MCC Support, Athens, Greece, Feb. 2017.

[3] E. Arıkan, "A performance comparison of polar codes and Reed–Muller codes," *IEEE Commun. Lett.*, vol. 12, no. 6, pp. 477–479, Jun. 2008.

[4] 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.

[5] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6562–6582, Oct. 2013.

[6] R. Pedarsani, S. H. Hassani, I. Tal, and E. Telatar, "On the construction of polar codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, St. Petersburg, FL, USA, Aug. 2011, pp. 11–15.

[7] I. Tal, "On the construction of polar codes for channels with moderate input alphabet sizes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Hong Kong, Jun. 2015, pp. 1297–1301.

[8] C. Schnelling and A. Schmeink, "Construction of polar codes exploiting channel transformation structure," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2058–2061, Dec. 2015.

[9] P. Trifonov, "Efficient design and decoding of polar codes," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3221–3227, Nov. 2012.

[10] S.-Y. Chung, T. J. Richardson, and R. L. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 657–670, Feb. 2001.

[11] D. Wu, Y. Li, and Y. Sun, "Construction and block error rate analysis of polar codes over AWGN channel based on Gaussian approximation," *IEEE Commun. Lett.*, vol. 18, no. 7, pp. 1099–1102, Jul. 2014.

[12] H. Li and J. Yuan, "A practical construction method for polar codes in AWGN channels," in *Proc. IEEE TENCON Spring Conf.*, Apr. 2013, pp. 223–226.

[13] K. Niu, K. Chen, J. Lin, and Q. T. Zhang, "Polar codes: Primary concepts and practical decoding algorithms," *IEEE Commun. Mag.*, vol. 52, no. 7, pp. 192–203, Jul. 2014.

[14] K. Niu and K. Chen, "CRC-aided decoding of polar codes," *IEEE Commun. Lett.*, vol. 16, no. 10, pp. 1668–1671, Oct. 2012.

[15] B. Li, H. Shen, and D. Tse, "An adaptive successive cancellation list decoder for polar codes with cyclic redundancy check," *IEEE Commun. Lett.*, vol. 16, no. 12, pp. 2044–2047, Dec. 2012.

[16] *Multiplexing and Channel Coding' Release 12*, 3GPP, document TS 36.212, 2014.

[17] S. Lin and D. J. Costello, *Error Control Coding*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2004.

[18] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
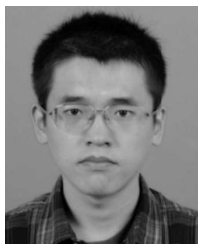
**JINCHENG DAI** (S'17) received the B.S. degree in information engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2014, where he is currently pursuing the Ph.D. degree with the School of Information and Communication Engineering. His research interests are in information theory, coding theory, and multiuser and MIMO signal processing.

**KAI NIU** (M'12) received the B.S. degree in information engineering and the Ph.D. degree in signal and information processing from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 1998 and 2003, respectively. He is currently a Professor with the School of Information and Communication Engineering, BUPT. His research interests include coding theory and its applications, space-time codes, and broadband wireless communication. Since 2008, he has been serving as a Senior Member of the Chinese Institute of Electronics (CIE) and a Committee Member of the Information Theory Chapter of CIE.

**ZHONGWEI SI** (M'16) received the Ph.D. degree from the KTH Royal Institute of Technology, Stockholm, Sweden, in 2013. In 2013, she joined the Beijing University of Posts and Telecommunications, Beijing, China, where she is currently an Associate Professor. Her research interests include wireless communication, information theory, and data mining.

J. Dai *et al.*: Does GA Work Well for the Long-Length Polar Code Construction?

**IEEE** *Access*

**CHAO DONG** (M'16) received the B.S and Ph.D. degrees in signal and information processing from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2007 and 2012, respectively. Since 2014, he has been a Lecturer with the School of Information and Communication Engineering, BUPT. His research interests are MIMO signal processing, multiuser precoding, decision feedback equalizer, and relay signal processing.

**JIARU LIN** (M'13) received the B.S., M.S., and Ph.D. degrees in signal and information processing from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 1982, 1987, and 2001, respectively. From 1991 to 1994, he was with the Swiss Federal Institute of Technology, Zürich, Switzerland. He is currently a Professor and the Ph.D. Supervisor with the School of Information and Communication Engineering, BUPT. His research interests are in wireless communication, cognitive radio networks, and information theory and coding.

● ● ●