On Distance Properties of Convolutional Polar Codes

Ruslan Morozov, *Member, IEEE*, Peter Trifonov, *Member, IEEE*

***Abstract*—A lower bound on the minimum distance of convolutional polar codes is provided. The bound is obtained from the minimum weight of the generalized cosets of the codes generated by the bottom rows of the polarizing matrix. Moreover, a construction of convolutional polar subcodes is proposed, which provides improved performance under successive cancellation list decoding. For sufficiently large list size, the decoding complexity of convolutional polar subcodes appears to be lower compared to Arikan polar subcodes with the same performance. The error probability of successive cancellation list decoding of convolutional polar subcodes is lower than that of Arikan polar subcodes with the same list size.**

***Index Terms*—Convolutional polar codes, polar codes, successive cancellation decoding, list decoding, polar subcodes.**

I. Introduction

In this paper we consider codes that were firstly introduced as branching-MERA codes [1] and then as convolutional polar codes (CvPCs) [2] by A. J. Ferris, C. Hirche and D. Poulin. These codes were shown to provide substantially better performance under successive cancellation (SC) decoding compared to classical polar codes [3]. In [2], both open-boundary and periodic-boundary CvPCs are presented, in this paper by CvPCs we always mean open-boundary CvPCs. In [4], the efficient min-sum implementation of SC decoding is presented for CvPCs, which requires one to perform only comparisons and additions and can be easily extended to the case of SC list (SCL) decoding. Other implementations of SCL decoding for CvPCs are presented in [5], [6].

Classical polar codes provide quite poor performance under SCL decoding due to their low minimum distance, which scales as $O(\sqrt{n})$ [7]. Although the minimum distance of a polar code can be found simply, the problem of computing the minimum distance of an arbitrary linear code is NP-complete. However, for moderate-length codes the minimum distance can be obtained by the method presented in [8].

The generator matrix of a CvPC consists of rows of $n \times n$ non-singular matrix $Q^{(n)}$, called the convolutional polarizing transformation (CvPT). In this paper we derive a tight lower bound on the minimum distance of CvPCs, based on computing the minimum weight of a coset, given by the i-th row of the CvPT, of a linear code, generated by the last $n-i-1$ rows of the CvPT. The weight enumerator polynomial of such coset can be expressed as $A_i(x) - A_{i+1}(x)$, where $A_i(x)$ is the weight spectrum of the code generated by the last $n-i$ rows of matrix $Q^{(n)}$. In the case of polar codes, an efficient method for approximate enumerator evaluation is available [9]. However, no methods are known for evaluation of a coset enumerator of convolutional polar codes.

The minimum distance of CvPCs is of the same order as in the case of classical polar codes. However, by generalizing the construction of randomized polar subcodes [10] to the case of CvPC, we obtain convolutional polar subcodes (CvPSs) with a reduced error coefficient, which provide superior performance under SCL decoding, compared to polar subcodes.

The paper is organized as follows. In Section II we introduce a representation of linear block codes, which is natural for the cases of Arikan and convolutional polar codes. The concepts of generalized cosets and recoverable vectors are introduced in Section III and are used to obtain a lower bound on the minimum distance of linear block codes. An efficient algorithm for computing the lower bound in the case of a CvPC is provided in Section IV. This algorithm is aimed to explore some properties of low-weight codewords of a CvPC. These properties are used for a construction of convolutional polar subcodes, which is proposed in Section V. The performance of the proposed code construction is presented in Section VI.

II. Background

A. Notations

The following notations are used throughout the paper. $\mathbb{F}$ denotes the Galois field of two elements. For integer n we denote $[n] = \{0, 1, \ldots n-1\}$. For vector a symbol $a_b^c = (a_b, a_{b+1}, \ldots, a_c)$. For two vectors a and b we denote their concatenation by (a, b). For $m \times n$ matrix A and sets $\mathcal{X} \subseteq [m]$, $\mathcal{Y} \subseteq [n]$, by $A_{\mathcal{X},\mathcal{Y}}$ we denote the submatrix of A with rows with indices from set $\mathcal{X}$ and columns with indices from set $\mathcal{Y}$, indexing of rows and columns starts with zero. Similar notations are applied to vectors as well. If $\mathcal{X} = *$ or $\mathcal{Y} = *$, this means that all rows or all columns of the original matrix are in the submatrix. Furthermore, $A_{\overline{\mathcal{X}},\overline{\mathcal{Y}}}$ denotes submatrix of A consisting of rows and columns with indices that are not in $\mathcal{X}$ and $\mathcal{Y}$, respectively. The vector of i zeroes is denoted by $\mathbf{0}^i$, or just by $\mathbf{0}$, if i is clear from the context.

B. A Representation of a Linear Block Code and Successive Cancellation Decoding

Consider binary linear block code in the form

$$\left\{ u_0^{n-1} G^{(n)} \middle| u_{\mathcal{I}} \in \mathbb{F}^k, u_{\mathcal{F}} = \mathbf{0} \right\}, \mathcal{I} \subseteq [n], |\mathcal{I}| = k, \quad (1)$$

where $G^{(n)}$ is an $n \times n$ non-singular binary matrix, $\mathcal{I}$ is called the information set and $\mathcal{F} = [n] \setminus \mathcal{I}$ is called the frozen set. The generator matrix of such code is $G^{(n)}_{\mathcal{I},*}$. Note that any

The authors are with the Saint Petersburg Polytechnic University, Russia. E-mail: {rmorozov, petert}@dcn.icc.spbstu.ru

This work is partially presented at 2019 IEEE International Symposium on Information Theory

(n,k) linear code with generator matrix G can be expressed as in (1) with $G^{(n)}$, such that $G = G^{(n)}_{\mathcal{I},*}$ for some $\mathcal{I} \subseteq [n]$. For example, classical polar codes [3] have $G^{(n)} = F^{\otimes m}$ for $n = 2^m$, $F = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$.

Such code representation enables one to employ the successive cancellation (SC) decoding method. Consider transmission of codeword $c_0^{n-1} = u_0^{n-1} G^{(n)}$ through a binary-input memoryless channel $\mathcal{W} : \mathbb{F} \to \mathcal{Y}$. Let y_0^{n-1} be the output of this channel. After demodulation, the probabilities $W(c_i|y_i) = \mathcal{W}(y_i|c_i) / \left(\mathcal{W}(y_i|0) + \mathcal{W}(y_i|1)\right)$ for $c_i \in \mathbb{F}$ are provided to the decoding algorithm. Given the prior hard decisions $\hat{u}_0 \dots \hat{u}_{\varphi-1}$, at phase φ the SC decoding algorithm calculates probabilities $W_n^{(\varphi)}(\hat{u}_0^{\varphi-1}, u_\varphi | y_0^{n-1})$, defined as

$$W_n^{(\varphi)}(u_0^{\varphi}|y_0^{n-1}) = \sum_{u_{\varphi+1}^{n-1} \in \mathbb{F}^{n-\varphi-1}} W^n(u_0^{n-1} G^{(n)} | y_0^{n-1}), \quad (2)$$

where $W^n(c_0^{n-1}|y_0^{n-1}) = \prod_{i=0}^{n-1} W(c_i|y_i)$. The channels $W_n^{(\varphi)} : \mathcal{Y} \to \mathbb{F}^{\varphi+1}$ are called *bit subchannels*. Then, the hard decision on u_φ is made by

$$\hat{u}_\varphi = \begin{cases} 0, & \varphi \in \mathcal{F} \\ \arg\max\limits_{u_\varphi \in \mathbb{F}} W_n^{(\varphi)}(\hat{u}_0^{\varphi-1}, u_\varphi | y_0^{n-1}), & \varphi \notin \mathcal{F}. \end{cases}$$

The SC decoding can be defined for any linear code, if an efficient method for computing $W_n^{(\varphi)}(u_0^{\varphi}|y_0^{n-1})$ is available. However, SC decoding can provide reasonable performance only for codes with $G^{(n)}$, such that the capacities of bit subchannels $W_n^{(\varphi)}$ polarize, i.e. converge to 0 or 1 with $n \to \infty$.

C. Convolutional Polar Codes

Convolutional polar codes [2] (CvPCs) are a family of linear block codes, for which $G^{(n)}$, $n = 2^m$, is equal to the matrix of convolutional polarizing transformation (CvPT) $Q^{(n)}$, such that

$$Q^{(n)} = \left(X^{(n)} Q^{(n/2)}, Z^{(n)} Q^{(n/2)} \right), \quad (3)$$

where $Q^{(1)} = (1)$, $X^{(l)}$ and $Z^{(l)}$ are $l \times l/2$ matrices, defined for even l as

$$X^{(l)}_{i,j} = \begin{cases} 1, & \text{if } 2j \le i \le 2j+2 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

$$Z^{(l)}_{i,j} = \begin{cases} 1, & \text{if } 2j < i \le 2j+2 \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

For example, $X^{(4)} = \begin{pmatrix} 1110 \\ 0011 \end{pmatrix}^T$, $Z^{(4)} = \begin{pmatrix} 0110 \\ 0001 \end{pmatrix}^T$. Expansion (3) corresponds to one *layer* of the CvPT. In Fig. 1, the m-th layer of the CvPT is a mapping of vector u_0^{n-1} to vectors $x_0^{n/2-1} = u_0^{n-1} X^{(n)}$ and $z_0^{n/2-1} = u_0^{n-1} Z^{(n)}$.

It is shown in [4] that for $n = 2^m$, $\varphi \in [n]$, the value of $W_n^{(\varphi)}(u_0^{\varphi}|y_0^{n-1})$ for the CvPT can be recursively computed as

$$W_n^{(2\psi)}(u_0^{2\psi}|y) = \sum_w W_{n/2}^{(\psi)} \left((u_0^{2\psi}, w) X^{(2\psi+2)} | y' \right)$$

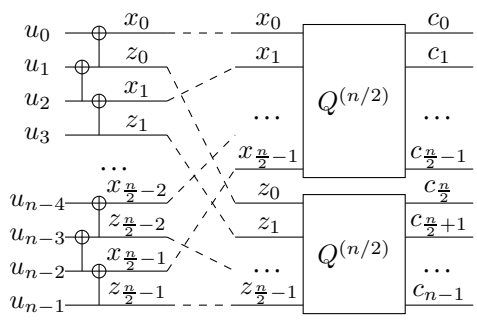


Fig. 1: Convolutional polarizing transformation $Q^{(n)}$.

$$\times W_{n/2}^{(\psi)} \left((u_0^{2\psi}, w) Z^{(2\psi+2)} | y'' \right) \quad (6)$$

$$W_n^{(2\psi+1)}(u_0^{2\psi+1}|y) = \sum_{u_{2\psi+2}, w} W_{n/2}^{(\psi+1)} \left((u_0^{2\psi+2}, w) X^{(2\psi+4)} | y' \right)$$

$$\times W_{n/2}^{(\psi+1)} \left((u_0^{2\psi+2}, w) Z^{(2\psi+4)} | y'' \right) \quad (7)$$

$$W_n^{(n-1)}(u_0^{n-1}|y) = W_{n/2}^{(\frac{n}{2}-1)} \left(u_0^{n-1} X^{(n)} | y' \right)$$

$$\times W_{n/2}^{(\frac{n}{2}-1)} \left(u_0^{n-1} Z^{(n)} | y'' \right) \quad (8)$$

for $0 \le \psi < n/2 - 1$, where $y = y_0^{n-1}$, and $y' = y_0^{n/2-1}$, $y'' = y_{n/2}^{n-1}$ are subvectors of y. These formulae are the same as in [4] under permutation of the output vector y by the bit-reversal permutation, which is omitted from the definition (3) of the CvPT for the sake of simplicity.

III. A Lower Bound on The Minimum Distance of Linear Codes

A. Basic Definitions

Let $\mathbb{S}_n$ be the set of all linear subspaces of $\mathbb{F}^n$.

Denote $a_0^{l-1} \cdot b_0^{l-1} = \sum_{i=0}^{l-1} a_i b_i$, where $a_i, b_i \in \mathbb{F}$. For vectors $b^{(0)}, \dots, b^{(l-1)} \in \mathbb{F}^t$, denote by $\langle b^{(0)}, \dots, b^{(l-1)} \rangle$ the linear subspace of $\mathbb{F}^t$ with basis vectors $b^{(i)}$, i.e.

$$\left\langle b^{(0)}, \dots, b^{(l-1)} \right\rangle = \left\{ \sum_{i=0}^{l-1} a_i b^{(i)} \mid a_0^{l-1} \in \mathbb{F}^l \right\}.$$

A sum over an empty set is assumed to be equal to zero, which implies $\langle\rangle = \{\mathbf{0}^t\}$, where t is clear from the context. By abuse of notation, we write $x_0 x_1 \dots x_{t-1}$ for $x_i \in \mathbb{F}$ to denote a vector $(x_0, x_1, \dots, x_{t-1}) \in \mathbb{F}^t$.

Example 1. It can be seen that $\mathbb{S}_2 = \{\langle\rangle, \langle 10 \rangle, \langle 01 \rangle, \langle 11 \rangle, \langle 10, 01 \rangle\}$, and $|\mathbb{S}_3| = 16$.

B. Outline of the Approach

Consider a code $\mathcal{P}$ in the form (1) with $\mathcal{F} = [\varphi]$, i.e. the set of vectors $(\mathbf{0}^\varphi, u_\varphi^{n-1}) G^{(n)}$. Code $\mathcal{P}$ can be split in two sets corresponding to each value of u_φ. Namely, $\mathcal{P} = \mathcal{P}_0 \cup \mathcal{P}_1$, where $\mathcal{P}_a$ consists of all codewords of the form $(\mathbf{0}^\varphi, a, u_{\varphi+1}^{n-1}) G^{(n)}$. These subsets are equal to the subsets, which probabilities are computed at the φ-th phase of the SC decoding algorithm by (2), provided that the estimated symbols $\hat{u}_0^{\varphi-1}$ are zero. Since we are interested in distance properties of the code, we can assume that $\hat{u}_0^{\varphi-1} = \mathbf{0}^\varphi$.

Let $d_n^{(\varphi)}$ be the distance between $\mathcal{P}_0$ and $\mathcal{P}_1$, i.e. $d_n^{(\varphi)} = \min_{\dot{c}\in\mathcal{P}_0,\ddot{c}\in\mathcal{P}_1} \mathbf{wt}(\dot{c}+\ddot{c})$. Consider $\dot{c}$ and $\ddot{c}$, for such the minimum is achieved, i.e., $\dot{c} = (\mathbf{0}^{\varphi}, 0, \dot{u}_{\varphi+1}^{n-1})G^{(n)}$, $\ddot{c} = (\mathbf{0}^{\varphi}, 1, \ddot{u}_{\varphi+1}^{n-1})G^{(n)}$, such that $d_n^{(\varphi)} = \mathbf{wt}(\dot{c}+\ddot{c}) = \mathbf{wt}(\tilde{c})$. Note that $\tilde{c} = (\mathbf{0}^{\varphi}, 1, \dot{u}_{\varphi+1}^{n-1} + \ddot{u}_{\varphi+1}^{n-1})G^{(n)}$ corresponds to value $u_\varphi = 1$, so $\tilde{c} \in \mathcal{P}_1$. Hence, $d_n^{(\varphi)}$ is equal to the weight of a minimum-weight codeword from $\mathcal{P}_1$. In general, we can say that if $\hat{u}_0^{\varphi-1} = u_0^{\varphi-1}$, i.e., all previous symbols are estimated correctly, then the probability of erroneous estimation of u_φ in the case of transmission over sufficiently good binary memoryless channel is mainly defined by $d_n^{(\varphi)} = \min_{c\in\mathcal{P}_1} \mathbf{wt}(c)$.

In section III-C we consider the partition of $\mathcal{P}$ in two sets $\mathcal{P}'_0$ and $\mathcal{P}'_1$ not by the value of u_φ, but by the value of some linear combination $p_0^{j-1} \cdot u_\varphi^{\varphi+j-1}$ of symbols $u_\varphi^{\varphi+j-1}$. Thus, set $\mathcal{P}'_a$, $a \in \mathbb{F}$ consists of all codewords $(\mathbf{0}^{\varphi}, u_\varphi^{n-1})G^{(n)}$ satisfying $p_0^{j-1} \cdot u_\varphi^{\varphi+j-1} = a$.

In section III-D we consider transmission of codewords through binary erasure channel (BEC) $W : \mathbb{F} \to \mathbb{F} \cup \{\epsilon\}$, defined as $W(x|x) = 1 - p_\epsilon$, $W(\epsilon|x) = p_\epsilon$, where p_ϵ is the erasure probability. We consider the mapping of the set of erased symbols $\mathcal{E} \subseteq [n]$ to the set of all linear combinations of symbols $u_\varphi^{\varphi+j-1}$, which can be recovered by the receiver by given $c_{\overline{\mathcal{E}}} = (c_i)_{i\notin\mathcal{E}}$. Thus, we consider a set $s \subseteq \mathbb{F}^j$ of all vectors $p_0^{j-1} \in \mathbb{F}^j$, such that the value of corresponding linear combination $p_0^{j-1} \cdot u_\varphi^{\varphi+j-1}$ can be recovered by receiver after erasure configuration $\mathcal{E}$. It is shown that $s \in \mathbb{S}_j$, i.e. s is a linear subspace of $\mathbb{F}^j$.

In section III-E, we prove that the minimum weight of vector from $\mathcal{P}'_1$ (i.e., the distance between $\mathcal{P}'_0$ and $\mathcal{P}'_1$) is equal to the minimum number of erasures, s. t. corresponding subspace $s \in \mathbb{S}_j$ of coefficients of recoverable linear combinations does *not* include the linear combination with coefficients p_0^{j-1}.

These results are combined to derive the algorithm for computing $d_n^{(\varphi)}$ in the case of CvPCs, which leads to the lower bound on minimum distance of CvPCs and the construction of CvPSs. Furthermore, we believe that the introduced concepts and their properties can be used for other $G^{(n)}$ that have recursive structure.

C. The Minimum Weight of Cosets and the Minimum Distance

Definition 1. Given an $n \times n$ non-singular matrix $G^{(n)}$, for a vector $p \in \mathbb{F}^j$ define a generalized coset $\mathcal{C}_n^{(\varphi)}(p)$ as

$$\mathcal{C}_n^{(\varphi)}(p) = \left\{ u_0^{n-1}G^{(n)} | u_0^{\varphi-1} = \mathbf{0} \ \wedge \ p \cdot u_\varphi^{\varphi+j-1} = 1 \right\} \quad (9)$$

Remark 1. In the case of $j > n - \varphi$, we assume in (9) that $u_l = 0$ for $l \geq n$.

We define *the weight of the φ-th bit subchannel* $W_n^{(\varphi)}$ as

$$d_n^{(\varphi)} = \min_{c\in\mathcal{C}_n^{(\varphi)}(1)} \mathbf{wt}(c).$$

Observe that for all $j > 0$ one has $\mathcal{C}_n^{(\varphi)}(p) = \mathcal{C}_n^{(\varphi)}(p, \mathbf{0}^j)$, which implies $d_n^{(\varphi)} = \min_{c\in\mathcal{C}_n^{(\varphi)}(1,\mathbf{0}^j)} \mathbf{wt}(c)$.

Lemma 1. *If a linear code with minimum distance d is generated by rows of $G^{(n)}$ with indices from $\mathcal{I} \subseteq [n]$, then*

$$d \geq \min_{\varphi\in\mathcal{I}} d_n^{(\varphi)}. \quad (10)$$

Proof. Consider the minimum-weight codeword $c_0^{n-1} = u_0^{n-1}G^{(n)}$, $\mathbf{wt}(c_0^{n-1}) = d$. Let ψ be the first position of non-zero element in u_0^{n-1}. Thus, $\psi \in \mathcal{I}$, $u_\psi = 1$, $u_0^{\psi-1} = \mathbf{0}$, which implies $c_0^{n-1} \in \mathcal{C}_n^{(\psi)}(1)$ and $d = \mathbf{wt}(c_0^{n-1}) \geq d_n^{(\psi)} \geq \min_{\varphi\in\mathcal{I}} d_n^{(\varphi)}$. □

This bound is valid for any linear block code represented in the form of (1). However, the evaluation of $d_n^{(\varphi)}$ is not a simple problem for an arbitrary $G^{(n)}$.

D. Recoverable and erased vectors

Consider transmission of a codeword $c_0^{n-1} = u_0^{n-1}G^{(n)}$ of a code with frozen set $\mathcal{F} = [\varphi]$, $u_0^{\varphi-1} = \mathbf{0}$ and dimension $k = n - \varphi$ over a BEC.

The set of erased positions $\mathcal{E} \subseteq [n]$ is called an *erasure configuration*. When erasure configuration $\mathcal{E}$ occurs, the values $c_{\overline{\mathcal{E}}} = u_\varphi^{n-1}\hat{G}$ are available for the receiver, where $\hat{G} = G^{(n)}_{[\varphi],\overline{\mathcal{E}}}$ is the $k \times r$ submatrix of $G^{(n)}$ without the rows from $[\varphi]$ and without the columns from $\mathcal{E}$, $r = n - |\mathcal{E}|$. Denote by $\mathcal{U}$ the set of all $\hat{u}_\varphi^{n-1}$ such that $\hat{u}_\varphi^{n-1}\hat{G} = c_{\overline{\mathcal{E}}}$. One can see that

$$\mathcal{U} = \left\{ u_\varphi^{n-1} + a_0^{k-1} | a_0^{k-1} \in \mathrm{cs}^{\perp}(\hat{G}) \right\}, \quad (11)$$

where for a set of vectors $\mathcal{A} \subseteq \mathbb{F}^t$, by $\mathcal{A}^{\perp} \subseteq \mathbb{F}^t$ we denote the set of vectors $x_0^{t-1} : \forall y_0^{t-1} \in \mathcal{A} : x_0^{t-1} \cdot y_0^{t-1} = 0$, and $\mathrm{cs}(A)$ is the column space of matrix A. The value u_φ^{n-1} can be unambiguously recovered by the receiver after erasure configuration $\mathcal{E}$ iff $|\mathcal{U}| = 1$, i.e. $\mathcal{U} = \{u_\varphi^{n-1}\}$.

More generally, consider recoverability of the value of a linear combination $p_0^{k-1} \cdot u_\varphi^{n-1}$ after erasure configuration $\mathcal{E}$. The set of values of $p_0^{k-1} \cdot \hat{u}_\varphi^{n-1}$ for all $\hat{u}_\varphi^{n-1} \in \mathcal{U}$ is given by

$$\left\{ p_0^{k-1} \cdot (u_\varphi^{n-1} + a_0^{k-1}) \mid a_0^{k-1} \in \mathrm{cs}^{\perp}(\hat{G}) \right\}. \quad (12)$$

We say that vector p_0^{k-1} is $(\mathcal{E}, \varphi)$-*recoverable*, if the corresponding linear combination $p_0^{k-1} \cdot u_\varphi^{n-1}$ can be recovered unambiguously for given $c_{\overline{\mathcal{E}}}$, i.e., the set (12) contains only the correct value $p_0^{k-1} \cdot u_\varphi^{n-1}$. Expanding the brackets in (12), one can see that p_0^{k-1} is $(\mathcal{E}, \varphi)$-recoverable iff $\forall a_0^{k-1} \in \mathrm{cs}^{\perp}(\hat{G}) : p_0^{k-1} \cdot a_0^{k-1} = 0$, which leads to $p_0^{k-1} \in \mathrm{cs}^{\perp\perp}(\hat{G}) = \mathrm{cs}(\hat{G})$. Thus, the set of $(\mathcal{E}, \varphi)$-recoverable vectors is a linear space, which is equal to $\mathrm{cs}(\hat{G}) \in \mathbb{S}_k$.

Definition 2. Let $s \in \mathbb{S}_j$ be the space of all p_0^{j-1}, such that $(p_0^{j-1}, \mathbf{0}^{k-j})$ is $(\mathcal{E}, \varphi)$-recoverable. In this case, s is called the $(\mathcal{E}, \varphi, j)$-*space* and is denoted by $\chi_n^{(\varphi,j)}(\mathcal{E})$, and $\mathcal{E}$ is called an (s, φ, j)-*configuration*. The set of (s, φ, j)-configurations is denoted by $\xi_n^{(\varphi,j)}(s)$. Thus,

$$\chi_n^{(\varphi,j)}(\mathcal{E}) = \left\{ p_0^{j-1} \mid (p_0^{j-1}, \mathbf{0}^{k-j}) \in \mathrm{cs}\left(G^{(n)}_{[\varphi],\overline{\mathcal{E}}} \right) \right\}, \quad (13)$$

$$\xi_n^{(\varphi,j)}(s) = \left\{ \mathcal{E} \mid \chi_n^{(\varphi,j)}(\mathcal{E}) = s \right\}. \quad (14)$$

If $\mathcal{A}$ is a set, denote by $2^{\mathcal{A}}$ the set of all subsets of $\mathcal{A}$. Thus, function $\chi_n^{(\varphi,j)} : 2^{[n]} \to \mathbb{S}_j$, maps an erasure configuration, which is a subset of $[n]$, to a linear subspace of $\mathbb{F}^j$, and $\xi_n^{(\varphi,j)}$ returns the inverse image of $\chi_n^{(\varphi,j)}$. Note that $\chi_n^{(\varphi,j)}$ is not injective, so $\xi_n^{(\varphi,j)} : \mathbb{S}_j \to 2^{2^{[n]}}$.

In words, $\chi_n^{(\varphi,j)}(\mathcal{E})$ defines the set of vectors p_0^{j-1}, for which the value of linear combination $p_0^{j-1}\bullet u_\varphi^{\varphi+j-1}$ can be recovered after erasure configuration $\mathcal{E}$, provided that $u_0^{\varphi-1}=\mathbf{0}$. Conversely, $\xi_n^{(\varphi,j)}(s)$ defines the set of erasure configurations, after which the linear combination $p_0^{j-1}\bullet u_\varphi^{\varphi+j-1}$ can be deduced by the receiver if *and only if* $p\in s$.

Remark 2. Let $j>k$, i.e. $j=k+h$ for some $h>0$. In this case, the conditional part of definition (13) is inconsistent. We extend the definition as follows. In Remark 1 we assume that symbols u_{n+h} for $h\geq 0$ are equal to zero. Hence, these symbols are always perfectly known for the receiver, so any $\mathcal{E}$ does not erase any symbol u_{n+h}. Observe that any vector from $\mathbb{F}^j\setminus\chi_n^{(\varphi,j)}(\mathcal{E})$ must be *not* $(\mathcal{E},\varphi)$-recoverable, so for any $\mathcal{E}$ and $q_0^{h-1}\in\mathbb{F}^h$, we must include vector $(\mathbf{0}^k,q_0^{h-1})$ in the set $\chi_n^{(\varphi,k+h)}(\mathcal{E})$. This leads to

$$\chi_n^{(\varphi,k+h)}(\mathcal{E})=\left\{(p,q)\;\middle|\;p\in\chi_n^{(\varphi,k)}(\mathcal{E}),q\in\mathbb{F}^h\right\}.$$

Similarly, we assume that $\xi_n^{(\varphi,k+h)}(s)=\emptyset$ for all s which do not contain $(\mathbf{0}^k,q)$ for some $q\in\mathbb{F}^h$.

Example 2. Consider $(s,0,2)$-configurations for the case of $n=2$, $c_0^1=u_0^1Q^{(2)}=(u_0+u_1,u_1)$. For erasure configuration $\mathcal{E}=\{0\}$, the only non-zero vector which is $(\mathcal{E},0)$-recoverable is $p=(0,1)$. That is, if symbol c_0 is erased, one can recover unambiguously only $u_1=c_1$. This means that $\{0\}\in\xi_2^{(0,2)}(\langle 01\rangle)$. All $(s,0,2)$-configurations are

$$\xi_2^{(0,2)}(\langle 01\rangle)=\{\{0\}\},\xi_2^{(0,2)}(\langle 10\rangle)=\emptyset,\xi_2^{(0,2)}(\langle 11\rangle)=\{\{1\}\},$$
$$\xi_2^{(0,2)}(\langle\rangle)=\{\{0,1\}\},\xi_2^{(0,2)}(\mathbb{F}^2)=\{\emptyset\}. \tag{15}$$

That is, there are no erasure configurations, such that only $\langle 10\rangle$ (i.e. symbol u_0) is unambiguously recoverable, and the whole vector u_0^1 can be unambiguously recovered only if there are no erasures. For the same case, the $(\mathcal{E},0,2)$-spaces are

$$\chi_2^{(0,2)}(\emptyset)=\mathbb{F}^2,\chi_2^{(0,2)}(\{0\})=\langle 01\rangle,$$
$$\chi_2^{(0,2)}(\{1\})=\langle 11\rangle,\chi_2^{(0,2)}(\{0,1\})=\langle\rangle.$$

Example 3. Consider the case of $\varphi=2$, $j=2$, $n=4$ and $c_0^3=u_0^3Q^{(4)}=(u_0+u_1+u_3,u_2+u_3,u_1+u_2+u_3,u_3)$. Since $\varphi=2$ implies $u_0^1=\mathbf{0}$, one has $c_0=c_3=u_3$, $c_1=c_2=u_2+u_3$ and one can restore u_3 by c_0 or c_3, and cannot restore other non-zero linear combinations of u_2^3 if both c_1 and c_2 are erased. Thus, $\xi_4^{(2,2)}(\langle 01\rangle)=\{\{1,2\},\{0,1,2\},\{1,2,3\}\}$.

E. Coset minimum weight and erasure configurations

For a subspace $s\in\mathbb{S}_j$, we denote the minimal cardinality of a (s,φ,j)-configuration as

$$\delta_n^{(\varphi,j)}(s)=\min_{\mathcal{E}\in\xi_n^{(\varphi,j)}(s)}|\mathcal{E}|, \tag{16}$$

assuming that the minimum over the empty set is $+\infty$.

Theorem 1. *Let $\varphi\in[n]$ and $j>0$. For any $p\in\mathbb{F}^j$,*

$$\min_{c\in\mathcal{C}_n^{(\varphi)}(p)}\mathbf{wt}(c)=\min_{s\in\mathbb{S}_j:p\notin s}\delta_n^{(\varphi,j)}(s).$$

Proof. Denote $\mathcal{A}=\left\{\mathrm{supp}(c)\middle|c\in\mathcal{C}_n^{(\varphi)}(p)\right\}$,

$$\mathcal{B}=\bigcup_{s\in\mathbb{S}_j:p\notin s}\xi_n^{(\varphi,j)}(s)=\bigcup_{s\in\mathbb{S}_j:p\notin s}\left\{\mathcal{E}\mid\chi_n^{(\varphi,j)}(\mathcal{E})=s\right\}$$
$$=\left\{\mathcal{E}\mid p\notin\chi_n^{(\varphi,j)}(\mathcal{E})\right\}.$$

Then the theorem can be reformulated as $\min_{\Omega\in\mathcal{A}}|\Omega|=\min_{\mathcal{E}\in\mathcal{B}}|\mathcal{E}|$.

If $\Omega\in\mathcal{A}$, then there exists u_φ^{n-1}, such that $p\bullet u_\varphi^{\varphi+j-1}=1$ and $\Omega=\mathrm{supp}(c_0^{n-1})$ for $c_0^{n-1}=(\mathbf{0}^\varphi,u_\varphi^{n-1})G^{(n)}$. In this case $c_{\overline{\Omega}}=\mathbf{0}$ and the all-zero value $\hat{u}_\varphi^{n-1}=\mathbf{0}$ also belongs to set (11) of possible values of u_φ^{n-1} for the given $c_{\overline{\Omega}}$, but $p\bullet\hat{u}_\varphi^{\varphi+j-1}=0$. Thus, the value of $p\bullet u_\varphi^{\varphi+j-1}$ is not recoverable after erasure configuration Ω, which implies $p\notin\chi_n^{(\varphi,j)}(\Omega)\implies\Omega\in\mathcal{B}$. So, $\Omega\in\mathcal{A}\implies\Omega\in\mathcal{B}$ and $\min\limits_{\Omega\in\mathcal{A}}|\Omega|\geq\min\limits_{\mathcal{E}\in\mathcal{B}}|\mathcal{E}|$.

If $\mathcal{E}\in\mathcal{B}$, then $p\notin\chi_n^{(\varphi,j)}(\mathcal{E})$, which by Definition 2 implies $(p,\mathbf{0}^{k-j})\notin\mathrm{cs}(\hat{G})$ and $\exists a_0^{k-1}\in\mathrm{cs}^\perp(\hat{G}):(p,\mathbf{0}^{k-j})\bullet a_0^{k-1}=1$, which implies $p\bullet a_0^{j-1}=1$. Denote $\hat{c}_0^{n-1}=(\mathbf{0}^\varphi,a_0^{k-1})G^{(n)}$. Since $p\bullet a_0^{j-1}=1$, by Definition 1 one has $\hat{c}_0^{n-1}\in\mathcal{C}_n^{(\varphi)}(p)$, and therefore $\mathrm{supp}(\hat{c})\in\mathcal{A}$. On the other hand, $\hat{c}_{\overline{\mathcal{E}}}=a_0^{k-1}\hat{G}=\mathbf{0}$, which means $\mathrm{supp}(\hat{c})\subseteq\mathcal{E}$. So, $\forall\mathcal{E}\in\mathcal{B}\;\;\exists\Omega\in\mathcal{A}:\Omega\subseteq\mathcal{E}$, hence, $\min\limits_{\Omega\in\mathcal{A}}|\Omega|\leq\min\limits_{\mathcal{E}\in\mathcal{B}}|\mathcal{E}|$. □

Corollary 1. *For any $j>0$:*

$$d_n^{(\varphi)}=\min\left\{\delta_n^{(\varphi,j)}(s)\middle|s\in\mathbb{S}_j:(1,\mathbf{0}^{j-1})\notin s\right\}.$$

IV. Bound on Minimum Distance of Convolutional Polar Codes

The structure of the convolutional polarizing transformation $Q^{(n)}$, $n=2^m$, enables one to compute easily $\delta_n^{(\varphi,j)}(s)$, defined in (16), for $j=3$. By computing the values of $\delta_n^{(\varphi,3)}(s)$, one can obtain the values of $d_n^{(\varphi)}$ by Corollary 1 and the lower bound on the minimum distance by Lemma 1.

Consider transmission of $c_0^{n-1}=u_0^{n-1}Q^{(n)}$, such that $u_0^{\varphi-1}=\mathbf{0}$, through a BEC and let the erasure configuration be $\mathcal{E}$. The intuition behind recursive computing of $\delta_n^{(\varphi,3)}(s)$ is as follows.

Consider the case of $\varphi=2\psi+1<n-1$. Denote $x_0^{n/2-1}=u_0^{n-1}X^{(n)}$, $z_0^{n/2-1}=u_0^{n-1}Z^{(n)}$, $\mathcal{E}'=\mathcal{E}\cap[\frac{n}{2}]$, $\mathcal{E}''=\left\{i\geq 0|i+\frac{n}{2}\in\mathcal{E}\right\}$. Recall that $\chi_n^{(2\psi+1,3)}(\mathcal{E})$ is the set of all p_0^2, such that the value of $p_0^2\bullet u_{2\psi+1}^{2\psi+3}$ can be deduced from c_0^{n-1} after erasure configuration $\mathcal{E}$. Similarly, $\chi_{n/2}^{(\psi,3)}(\mathcal{E}')$ and $\chi_{n/2}^{(\psi,3)}(\mathcal{E}'')$ are the sets of q_0^2 and r_0^2, s.t. $q_0^2\bullet x_\psi^{\psi+2}$ and $r_0^2\bullet z_\psi^{\psi+2}$ are recoverable from $c_0^{n/2-1}$ and $c_{n/2}^{n-1}$ after erasure configurations $\mathcal{E}'$ and $\mathcal{E}''$, under the assumption $x_0^{\psi-1}=\mathbf{0}$ and $z_0^{\psi-1}=\mathbf{0}$, respectively. By (4)–(5) one obtains $x_i=u_{2i}+u_{2i+1}+u_{2i+2}$ and $z_i=u_{2i+1}+u_{2i+2}$ for $i<\frac{n}{2}-1$, which, together with $u_0^{2\psi}=\mathbf{0}$, implies $x_0^{\psi-1}=z_0^{\psi-1}=\mathbf{0}$, so the above assumption holds. Furthermore, since u_0^{n-1} was processed by the m-th layer of the CvPT before transmission, the value of elements of $u_{2\psi+1}^{2\psi+3}$, as well as the value of any linear combination $p_0^2\bullet u_{2\psi+1}^{2\psi+3}$, can be deduced only from known linear combinations of elements of x_ψ^{n-1} and z_ψ^{n-1}. However, for

any $x_{\psi+3}^{n/2-1}$, $z_{\psi+3}^{n/2-1}$ and $u_{2\psi+1}^{2\psi+3}$, one can find $u_{2\psi+4}^{n-1}$, such that $(\mathbf{0}^{2\psi+1}, u_{2\psi+1}^{n-1}) = \left(\mathbf{0}^{\psi}, x_{\psi}^{n/2-1}, \mathbf{0}^{\psi}, z_{\psi}^{n/2-1}\right) Q^{(n)}$ as follows: set u_{2i+2} to $x_{i+1} + z_{i+1}$ for $i = \frac{n}{2}-2, \dots, \psi+1$, set u_{n-1} to $z_{n/2-1}$, and set u_{2i+1} to $z_i + u_{2i+2}$ for $i = \frac{n}{2}-2, \dots, \psi+2$. So, for any $p \in \mathbb{F}^3$, even complete knowledge of $x_{\psi+3}^{n/2-1}$ and $z_{\psi+3}^{n/2-1}$ does not provide the value $p \cdot u_{2\psi+1}^{2\psi+3}$. Thus, recoverable linear combinations $q_0^2 \cdot x_{\psi}^{\psi+2}$ and $r_0^2 \cdot z_{\psi}^{\psi+2}$ contain all information about recoverable linear combinations $p_0^2 \cdot u_{2\psi+1}^{2\psi+3}$, and therefore $\chi_n^{(2\psi+1,3)}(\mathcal{E})$ can be uniquely deduced from given $\chi_{n/2}^{(\psi,3)}(\mathcal{E}')$ and $\chi_{n/2}^{(\psi,3)}(\mathcal{E}'')$. The similar consideration for $\varphi = 2\psi+2$ leads to the fact that $\chi_n^{(2\psi+2,3)}(\mathcal{E})$ can also be deduced from $\chi_{n/2}^{(\psi,3)}(\mathcal{E}')$ and $\chi_{n/2}^{(\psi,3)}(\mathcal{E}'')$.

Let $\psi = \left\lfloor \frac{\varphi-1}{2} \right\rfloor$, $\mathbb{S}_3 = \{\mathcal{T}_i\}_{i=0}^{15}$. For any $l \in [16]$, consider $(\mathcal{T}_l, \varphi, 3)$-erasure configuration $\mathcal{E}$ for which the minimum in (16) is achieved, i.e. $\chi_n^{(\varphi,3)}(\mathcal{E}) = \mathcal{T}_l$ and $|\mathcal{E}| = \delta_n^{(\varphi,3)}(\mathcal{T}_l)$. Obviously, $|\mathcal{E}| = |\mathcal{E}'| + |\mathcal{E}''|$. Let $\chi_{n/2}^{(\psi,3)}(\mathcal{E}') = \mathcal{T}_i$, $\chi_{n/2}^{(\psi,3)}(\mathcal{E}'') = \mathcal{T}_j$. Then, $\mathcal{E}'$ and $\mathcal{E}''$ are also the minimum-weight $(\mathcal{T}_i, \psi, 3)$- and $(\mathcal{T}_j, \psi, 3)$- erasure configurations, respectively, i.e. $|\mathcal{E}'| = \delta_{n/2}^{(\psi,3)}(\mathcal{T}_i)$, and $|\mathcal{E}''| = \delta_{n/2}^{(\psi,3)}(\mathcal{T}_j)$. We know that $\mathcal{T}_l$ can be deduced from $\mathcal{T}_i$ and $\mathcal{T}_j$, i.e., for each φ and n there is a function $\mathbf{T}_n^{(\varphi)}(i,j)$, which returns $\mathcal{T}_l$ for given i and j, and for considered minimum-weight $\mathcal{E}$, $\mathcal{E}'$, $\mathcal{E}''$ one can obtain $\delta_n^{(\varphi,3)}(\mathbf{T}_n^{(\varphi)}(i,j)) = \delta_{n/2}^{(\psi,3)}(\mathcal{T}_i) + \delta_{n/2}^{(\psi,3)}(\mathcal{T}_j)$.

It turns out that $\mathbf{T}_n^{(\varphi)} = \mathbf{T}_{n'}^{(\varphi')}$ if $\varphi \equiv \varphi' \mod 2$, i.e., there are only two different functions $\mathbf{T}_n^{(\varphi)}$: one for odd φ and another one for even φ. They are defined as $\mathbf{T}_o, \mathbf{T}_e : [16] \times [16] \to \mathbb{S}_3$, such that

$$\begin{aligned}\mathbf{T}_o(i,j) = \{p_0^2 | \exists p' \in \mathcal{T}_i, p'' \in \mathcal{T}_j : \\ (p_0^2, 0, 0)^T = X_{\overline{[1]},*}^{(6)} p'^T + Z_{\overline{[1]},*}^{(6)} p''^T\} \qquad (17)\\ \mathbf{T}_e(i,j) = \{p_0^2 | \exists p' \in \mathcal{T}_i, p'' \in \mathcal{T}_j : \\ (p_0^2, 0)^T = X_{\overline{[2]},*}^{(6)} p'^T + Z_{\overline{[2]},*}^{(6)} p''^T\}. \qquad (18)\end{aligned}$$

The above consideration is summarized in the following theorem.

Theorem 2. *Denote $\Delta_{n,l}^{(\varphi)} = \delta_n^{(\varphi,3)}(\mathcal{T}_l)$ for $l \in [16]$, $n = 2^m$. Then, for the CvPT , for $0 \le \psi < \frac{n}{2}$:*

$$\Delta_{n,l}^{(2\psi+1)} = \min_{i,j} \left\{ \Delta_{n/2,i}^{(\psi)} + \Delta_{n/2,j}^{(\psi)} \middle| \mathbf{T}_o(i,j) = \mathcal{T}_l \right\}, \qquad (19)$$

$$\Delta_{n,l}^{(2\psi)} = \min_{i,j} \left\{ \Delta_{n/2,i}^{(\psi-1)} + \Delta_{n/2,j}^{(\psi-1)} \middle| \mathbf{T}_e(i,j) = \mathcal{T}_l \right\}. \qquad (20)$$

The base of the recursion is

$$\delta_1^{(0,1)}(\langle\rangle) = 1,\ \delta_1^{(0,1)}(\langle 1\rangle) = 0. \qquad (21)$$

Proof. The proof is in the Appendix. □

Remark 3. Note that formulae (19)–(20) include the cases of $\Delta_{n,l}^{(n-2)} = \delta_n^{(n-2,3)}(\mathcal{T}_l)$ and $\Delta_{n,l}^{(n-1)} = \delta_n^{(n-1,3)}(\mathcal{T}_l)$. They can be obtained according to the assumption in Remark 2 as follows. For $s \in \mathbb{S}_{i+h}$, denote the set of tails of length i by $s|_i = \left\{ p_0^{i-1} \mid p_0^{i+h-1} \in s \right\}$. We assume that any erasure configuration does not erase u_{n-1+h} for any $h > 0$, i.e.

$$\delta_n^{(n-i,i+h)}(s) = \begin{cases} \delta_n^{(n-i,i)}(s|_i), & \text{if } \forall p \in \mathbb{F}^h : (\mathbf{0}^i, p) \in s \\ +\infty, & \text{otherwise} \end{cases}$$

The same assumption is applied for computing the values of $\Delta_{1,l}^{(0)} = \delta_1^{(0,3)}(\mathcal{T}_l)$ from the values $\delta_1^{(0,1)}(s)$ for $s \in \mathbb{S}_1$ that are given by the base (21) of the recursion. This assumption, though not natural since symbols u_{n+h}, $h \ge 0$ do not exist, allows one to employ the unified formulae (19)–(20) for the cases of $\varphi > n-3$.

Remark 4. Formula (20) in the case of $\Delta_{n,l}^{(0)}$ leads to computing $\Delta_{n/2,i}^{(-1)} = \delta_{n/2}^{(-1,3)}(\mathcal{T}_i)$, which is formally equal, for a given $\mathcal{T}_i$, to the minimum weight of an erasure configuration which erases values $p \cdot u_{-1}^2$ for and only for $p \in \mathcal{T}_i$. For the symbols u_{-i}, $i > 0$, we do not employ the same assumption as in Remark 3. If one assumes that symbols with negative indices are always known and employs functions $\mathbf{T}_o$ and $\mathbf{T}_e$, one would obtain that input symbols on the current layer of the convolutional polarizing transformation u_{-2}, u_{-1}, and input symbol $x_{-1} = u_{-2} + u_{-1} + u_0$ on the next layer are always known, which implies that u_0 is always known. This would result in incorrect value of $\Delta_{n,l}^{(0)}$. Thus, we assume that u_{-i} for $i > 0$ are always erased, which leads to

$$\chi_n^{(-i,j)}(\mathcal{E}) = \left\{ (\mathbf{0}^i, p) \mid p \in \chi_n^{(0,j-i)}(\mathcal{E}) \right\}, 0 < i \le j.$$

The values $d_n^{(i)}$ for the case of a CvPC can be computed with Algorithm 1. The three-dimensional array τ of the subspaces of $\mathbb{F}^3$ is initialized in lines 1.2–1.7, such that $\tau[0][i][j] = \mathbf{T}_e(i,j)$ and $\tau[1][i][j] = \mathbf{T}_o(i,j)$. The values $\Delta_{1,*}^{(0)}$ are computed in lines 1.8–1.11. Function `M1Cluster`, presented in Algorithm 2, is called to obtain $\Delta_{n,*}^{(-1)}$ for $n = 1$ and $n = 2^\lambda$, respectively, in lines 1.12 and 1.20.

The values of $\Delta_{2^\lambda,*}^{(\varphi)}$ for $-1 \le \varphi < 2^\lambda$ are computed by Theorem 2 in lines 1.13–1.21 and stored in array C', using values of $\Delta_{2^{\lambda-1},*}^{(\psi)}$ for $-1 \le \psi < 2^{\lambda-1}$, which are stored in array C. The values $d_n^{(i)}$ are obtained as $d[i], i \in [n]$.

The asymptotic complexity of the Algorithm 1 is defined by the complexity of the main loop 1.13–1.21. The complexity of the λ-th iteration of the loop is defined by the complexity of the loop in lines 1.15–1.19, which consists of 2^λ iterations, each of them has complexity $O(1)$. Thus, the overall asymptotic complexity is $\sum_{\lambda=1}^{\log n} O(2^\lambda) = O(n)$.

In Fig. 2 the lower bound on minimum distance, computed by (10), for CvPCs of lengths 64, 1024, 16384 is presented. The codes are obtained via the Monte-Carlo method by minimization of the E_b/N_0 needed to achieve the SC decoding error probability 10^{-3}. For comparison, we also report the results for Arikan polar codes, which are optimized in the same way. One can see that the CvPCs can have lower, equal or higher minimum distance, compared to Arikan polar codes.

Unlike the case of Arikan polarizing transformation $A^{(n)}$, the weight of the i-th row of CvPT $Q^{(n)}$ is not necessarily equal to $d_n^{(i)}$. Thus, the bound (10) is not exact at least for codes with $\mathcal{I} = \{i\}$. In general, it is not known, for which cases the bound is exact. However, by employing the low-weight codeword search algorithm presented in [8], we verified that the bound is exact for CvPCs with $m = 5, \dots, 13$, rates $\frac{1}{20}, \dots \frac{19}{20}$ and target FER of SC decoding 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, and 10^{-6}.

Algorithm 1: Computing $d_n^{(i)}, n = 2^m$ for all $i \in [n]$

Input: m

$X \leftarrow \begin{pmatrix} 11000 \\ 01110 \\ 00011 \end{pmatrix}, Z \leftarrow \begin{pmatrix} 11000 \\ 00110 \\ 00001 \end{pmatrix}$

for $i, j = 0 \dots 15$ **do**
  $\tau[0 \dots 1][i][j] \leftarrow \emptyset$
  for $(p, q) \in \mathcal{T}_i \times \mathcal{T}_j$ **do**
    $r \leftarrow pX + qZ$
    if $r_3^4 = \mathbf{0}^2$ **then** $\tau[1][i][j] \leftarrow \tau[1][i][j] \cup r_0^2$
    if $r_4 = 0$ **then** $\tau[0][i][j] \leftarrow \tau[0][i][j] \cup r_1^3$

for $i = 0 \dots 15$ **do**
  if $\mathcal{T}_i = \mathbb{F}^3$ **then** $C[0][0][i] \leftarrow 0$
  else if $\mathcal{T}_i = \langle 010, 001 \rangle$ **then** $C[0][i] \leftarrow 1$
  else $C[0][i] \leftarrow +\infty$

$C[-1] \leftarrow \texttt{M1Cluster}(C[0])$

for $\lambda = 1 \dots m$ **do**
  $C'[0 \dots 2^\lambda - 1][0 \dots 15] \leftarrow +\infty$
  for $\varphi = 0 \dots 2^\lambda - 1$ **do**
    $\psi = \lceil \frac{\varphi}{2} \rceil - 1$
    for $i, j = 0 \dots 15$ **do**
      let $l : \mathcal{T}_l = \tau[\varphi \bmod 2][i][j]$
      $C'[\varphi][l] \leftarrow \min\{C[\psi][i] + C[\psi][j], C'[\varphi][l]\}$
  $C'[-1] \leftarrow \texttt{M1Cluster}(C'[0])$
  $\texttt{swap}(C, C')$

for $i = 0 \dots n-1$ **do** $d[i] \leftarrow \min_{s \in \mathbb{S}^3 : (1,0,0) \notin s} C[i][s]$

return $d[0 \dots n-1]$

Algorithm 2: M1Cluster

Input: C, array of $\Delta_{n,*}^{(0)}$
Output: D, array of $\Delta_{n,*}^{(-1)}$

$D[0 \dots 15] \leftarrow +\infty$
for $i = 0 \dots 15$ **do**
  if $\mathcal{T}_i \subseteq \langle 100, 010 \rangle$ **then**
    let $j : \mathcal{T}_j = \{(a, p_0^1) \mid (p_0^1, 0) \in \mathcal{T}_i, a \in \mathbb{F}\}$
    $D[j] \leftarrow \min\{C[i], D[j]\}$
return $D[0 \dots 15]$

V. Convolutional Polar Subcodes

In general, the SC decoding algorithm for codes, represented in the form (1), does not provide ML decoding. The Tal-Vardy list decoding algorithm [11] for polar codes can be immediately extended to the case of CvPC, using the techniques presented in [4]. With sufficiently large list size L the SCL algorithm delivers near-ML decoding. The SCL decoding error probability of convolutional polar codes is lower than that of classical polar codes, but still can be improved by extending the construction of randomized polar subcodes [10] to the case of the convolutional polarizing transformation.

By Lemma 1, any codeword $c_0^{n-1} = u_0^{n-1} G^{(n)}$ of weight d corresponds to vector u_0^{n-1} with at least one symbol $u_i = 1, i \in \mathcal{I}$, such that $d_n^{(i)} \le d$. In the case of polar codes, $d_n^{(i)}$ is equal to the weight of the i-th row of $A^{(n)}$. In the case of CvPCs one can obtain $d_n^{(i)}$ by Algorithm 1.

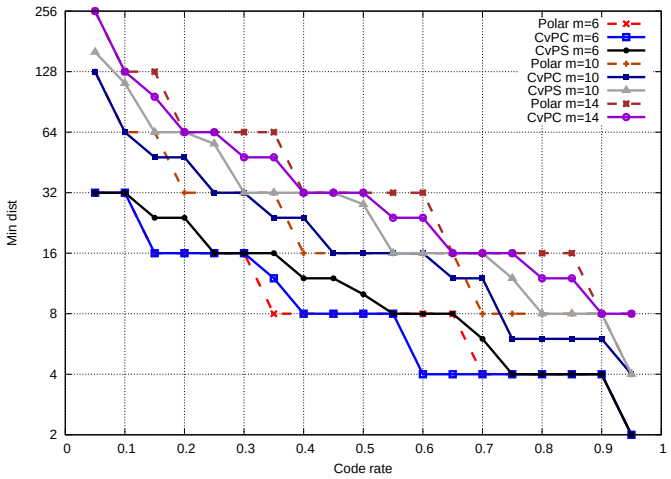


Fig. 2: Minimum distance of polar codes, CvPCs and CvPSs, constructed for AWGN channel for target FER 0.001.

Polar subcodes [10] are a generalization of polar codes, where some symbols $u_\varphi, \varphi \in \mathcal{D}$, called dynamic frozen symbols, are not set to zero, but to linear combinations of previous symbols $u_i, i < \varphi$. Such generalization allows one to obtain lower SCL decoding error probability. This approach can be immediately extended to the case of convolutional polarizing transformation. Namely, the dynamic freezing constraints should be constructed, so that they involve all non-frozen symbols u_i with the smallest $d_n^{(i)}$, but the indices of dynamic frozen symbols $i \in \mathcal{D}$ should be as small as possible, so that the SCL decoding algorithm can process these constraints at the earliest possible phases, minimizing thus the probability of the correct path being killed.

This results in the following code construction algorithm:

1) Construct an $(n, k+f)$ convolutional polar code, i.e. assign $u_\mathcal{S} = \mathbf{0}$ for the static frozen set $\mathcal{S} \subset [n]$ of the worst $n-k-f$ bit subchannels.
2) Choose the dynamic frozen set $\mathcal{D} \subseteq [n] \setminus \mathcal{S}$ as the set of f indices of the minimum-weight bit subchannels with the largest indices, that are not static frozen. Set

$$u_i = \sum_{j \in \mathcal{I}} V_{i,j} u_j, \ i \in \mathcal{D}, \ \mathcal{I} = [n] \setminus \mathcal{F},$$

where the frozen set $\mathcal{F} = \mathcal{S} \cup \mathcal{D}$ consists of indices of static frozen or dynamic frozen symbols, and $V_{i,j}$ are distributed uniformly over $\mathbb{F}$.

The set $\mathcal{I}$ for a CvPC optimized for SC decoding can be chosen either by evolution of erasure probabilities proposed in [2], or by Monte-Carlo simulations of the genie-aided SC decoder. Due to lack of analysis techniques for the list SC decoding algorithm, the optimal value of f should be determined by simulations.

Another component of the construction introduced in [10] is type-B dynamic freezing constraints, which are imposed on the symbols transmitted over the least reliable yet unfrozen subchannels. These constraints speed up error propagation for

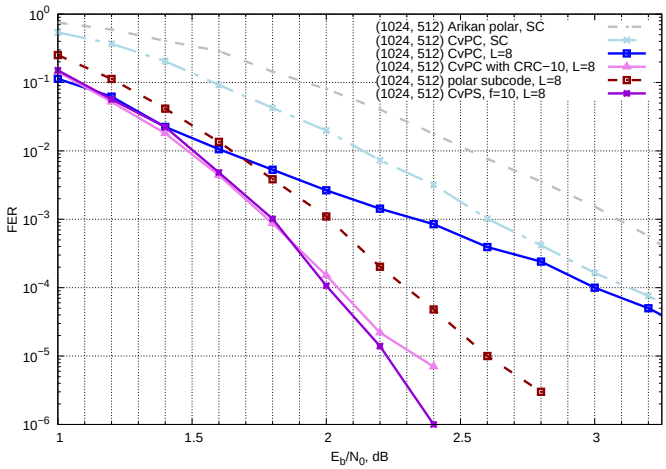


Fig. 3: Performance of $(1024, 512)$ CvPS with $f = 10$

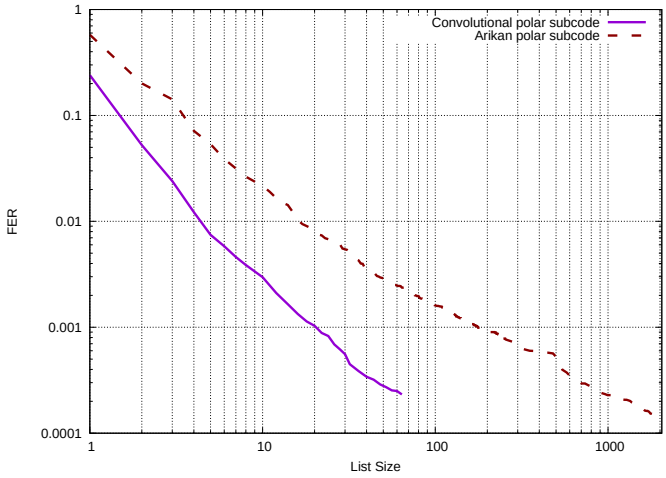


Fig. 4: Performance of a $(4096, 2048)$ CvPS with $f = 12$ in the AWGN channel with $E_b/N_0 = 1.25$ dB

incorrect paths in the list SC algorithm, so that the probabilities (2) of these paths decrease quickly, reducing thus the probability of a correct path being killed. However, simulations of moderate-length CvPS show that type-B dynamic frozen symbols do not provide any noticeable gain in the case of CvPSs.

VI. Performance of Convolutional Polar Subcodes

In Fig. 3 the performance of $(1024, 512)$ CvPS, polar code and polar subcode is presented for $f = 10$ for the case of AWGN channel. The polar code and the polar subcode are constructed for the AWGN channel with $E_b/N_0 = 2$ dB using Gaussian approximation of density evolution [12], and the CvPS is constructed for the same channel using Monte-Carlo simulations for subchannels qualities. One can see that the CvPS outperforms the randomized polar subcode [10], the CvPC [2] and the CvPC concatenated with CRC-10.

In Fig. 4 performance of a $(4096, 2048)$ CvPS with $f = 12$ type-A dynamic frozen symbols is presented. Transmission

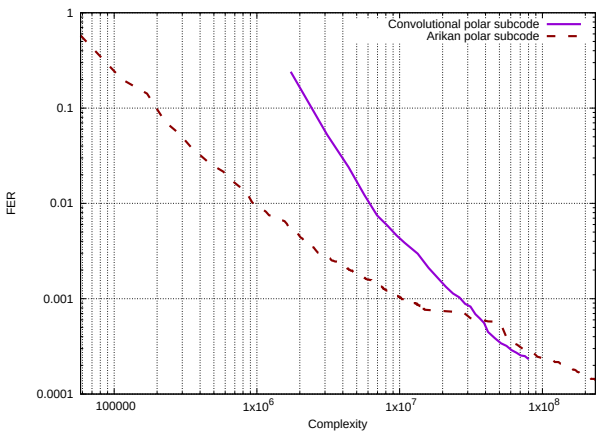


Fig. 5: SCL decoding complexity of $(4096, 2048)$ CvPS

of BPSK-modulated symbols over the AWGN channel with $E_b/N_0 = 1.25$ dB is considered. The decoding algorithm is the SCL decoding with different values of list size that are shown in the x-axis. The performance of the CvPS is compared to that of a polar subcode with $f = 12$ type-A dynamic frozen symbols and 52 type-B dynamic frozen symbols. One can see that the CvPS under SCL decoding with the same list size outperforms the classical polar subcode. The smaller list size can be used to achieve the same FER, which allows less sophisticated hardware implementation.

In Fig. 5 the complexity (the number of operations) of SCL decoding, based on the expressions derived in [4], of the described above codes is compared for list size $L = 1 \dots 64$ for the CvPS and $L = 1 \dots 1024$ for the polar subcode. The complexity is obtained as the number of additions and comparisons of LLRs. The complexity of SC decoding for CvPS is approximately $46.5n \log n$, as shown in [4]. The complexity of SC decoding of polar codes is $n \log n$. However, as shown in [2], the CvPT induces stronger polarization than Arikan polarizing transformation, so the smaller list size is needed to achieve the same FER. This leads to the smaller complexity needed to achieve FER less than $6 \cdot 10^{-4}$ in the case of CvPS, because achieving this FER requires list size $L = 352$ for the polar subcode and only $L = 28$ for the CvPS. Furthermore, for a large list size the SCL decoding is near-ML, and, for sufficiently good channel, FER of ML-decoding is primarily determined by the minimum distance and the error coefficient. Dynamic frozen symbols decrease the error coefficient and may even increase the minimum distance of a CvPS. In Fig. 2 one can see that the minimum distance of CvPSs is higher than that of CvPCs.

VII. Conclusions

In this paper a tight lower bound on the minimum distance of convolutional polar codes is provided. Furthermore, a generalization of the randomized construction of polar subcodes to the case of convolutional polarizing transformation is proposed. Simulations show that the proposed code construction has lower frame error rate under SCL decoding [4] compared to polar subcodes with the same list size. The complexity for achieving the same FER with convolutional polar subcodes

can be lower than in the case of polar subcodes [10] based on the Arikan polarizing transformation.

APPENDIX

Proof of Theorem 2. For erasure configuration $\mathcal{E} \subseteq [n]$, denote $\mathcal{E}' = \mathcal{E} \cap [n/2]$ and $\mathcal{E}'' = \{j - n/2 \mid j \in \mathcal{E} \setminus [n/2]\}$. We now consider the case of $\varphi = 2\psi + 1$ and prove (19).

Note that $u_0^{2\psi} = \mathbf{0}^{2\psi+1}$ implies $x_0^{\psi-1} = z_0^{\psi-1} = \mathbf{0}^{\psi}$. By (3) one obtains $\hat{Q} = \left(\hat{X}\hat{Q}' \ , \ \hat{Z}\hat{Q}''\right)$, where $\hat{Q} = Q^{(n)}_{\overline{[2\psi+1]},\overline{\mathcal{E}'}}$, $\hat{Q}' = Q^{(n/2)}_{\overline{[\psi]},\overline{\mathcal{E}'}}$, $\hat{Q}'' = Q^{(n/2)}_{\overline{[\psi]},\overline{\mathcal{E}''}}$, $\hat{X} = X^{(n)}_{\overline{[2\psi+1]},\overline{[\psi]}}$, $\hat{Z} = Z^{(n)}_{\overline{[2\psi+1]},\overline{[\psi]}}$. By (13), $p_0^2 \in \chi_n^{(\varphi,3)}(\mathcal{E})$ iff there exists q:

$$(p_0^2, \mathbf{0}^{k-3})^T = \hat{Q}q^T = (\hat{X}\hat{Q}', \hat{Z}\hat{Q}'')q^T = \hat{X}\hat{Q}'q'^T + \hat{Z}\hat{Q}''q''^T,$$

where $q = (q', q'')$, $k = n - \varphi$, which implies, in particular,

$$(\hat{X}\hat{Q}'q'^T)_{\overline{[3]}} = (\hat{Z}\hat{Q}''q''^T)_{\overline{[3]}}. \tag{22}$$

Denote $a = q'\hat{Q}'^T$, $b = q''\hat{Q}''^T$. Thus, $a \in \mathrm{cs}(\hat{Q}')$, $b \in \mathrm{cs}(\hat{Q}'')$. Then (22) implies $a\hat{X}^T_{\overline{[3]},*} = b\hat{Z}^T_{\overline{[3]},*}$, so from (4)–(5) one obtains

$$a\begin{pmatrix} 000000\ldots \\ 100000\ldots \\ 111000\ldots \\ 001110\ldots \\ \ldots \end{pmatrix} = b\begin{pmatrix} 000000\ldots \\ 100000\ldots \\ 011000\ldots \\ 000110\ldots \\ \ldots \end{pmatrix},$$

which leads to the system of equations

$$\begin{cases} a_i + a_{i+1} = b_i, & i = 1 \ldots n/2 - \psi - 2 \\ a_i = b_i, & i = 2 \ldots n/2 - \psi - 1 \end{cases} \tag{23}$$

It is easy to see that (23) implies $a_i = b_i = 0$ for $i \geq 3$. Let $k' = n/2 - \psi$. By the above consideration, for any $p \in \mathbb{F}^3$ one has $(p, \mathbf{0}^{k-3}) \in \mathrm{cs}(\hat{Q})$ iff there exists $p', p'' \in \mathbb{F}^3$, s.t. $(p', \mathbf{0}^{k'-3}) \in \mathrm{cs}(\hat{Q}')$, $(p'', \mathbf{0}^{k'-3}) \in \mathrm{cs}(\hat{Q}'')$, and

$$(p, \mathbf{0}^3)^T = \begin{pmatrix} 100 \\ 110 \\ 010 \\ 011 \\ 001 \\ 001 \end{pmatrix}(p')^T + \begin{pmatrix} 100 \\ 100 \\ 010 \\ 010 \\ 001 \\ 001 \end{pmatrix}(p'')^T. \tag{24}$$

Note that two last elements of vector in the left-hand side equals 0, and two last rows in the right hand size of (24) are identical, so last rows of these matrices can be removed. The resulting matrices are equal to $X^{(6)}_{\overline{[1]},*}$ and $Z^{(6)}_{\overline{[1]},*}$, respectively. Recalling (13), one obtains that $\chi_n^{(2\psi+1,3)}(\mathcal{E})$ consists of all p_0^2, for which there exist $p' \in \chi_{n/2}^{(\psi,3)}(\mathcal{E}')$, $p'' \in \chi_{n/2}^{(\psi,3)}(\mathcal{E}'')$:

$$(p_0^2, \mathbf{0}^2)^T = X^{(6)}_{\overline{[1]},*}p'^T + Z^{(6)}_{\overline{[1]},*}p''^T. \tag{25}$$

Observe that (25) is equivalent to the equation in the right-hand side of (17). Obviously, $|\mathcal{E}| = |\mathcal{E}'| + |\mathcal{E}''|$ and the minimal cardinality of $(\mathcal{T}_l, 2\psi+1, 3)$-configuration $|\mathcal{E}|$ for each $\mathcal{T}_l \in \mathbb{S}_3$ can be found exactly as it is stated in (19).

Equality (20) can be proved similarly.

REFERENCES

[1] A. J. Ferris and D. Poulin, "Branching MERA codes: a natural extension of polar codes," *CoRR*, vol. abs/1312.4575, 2013. [Online]. Available: http://arxiv.org/abs/1312.4575

[2] A. J. Ferris, C. Hirche, and D. Poulin, "Convolutional polar codes," *CoRR*, vol. abs/1704.00715, 2017. [Online]. Available: http://arxiv.org/abs/1704.00715

[3] 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, Jul. 2009.

[4] R. Morozov and P. Trifonov, "Efficient SC decoding of convolutional polar codes," in *International Symposium on Information Theory and its Applications (ISITA)*, Oct. 2018, pp. 442–446.

[5] H. Saber, Y. Ge, R. Zhang, W. Shi, and W. Tong, "Convolutional polar codes: LLR-based successive cancellation decoder and list decoding performance," in *International Symposium on Information Theory (ISIT)*, Jun. 2018, pp. 1480–1484.

[6] T. Prinz and P. Yuan, "Successive cancellation list decoding of BMERA codes with application to higher-order modulation," in *International Symposium on Turbo Codes and Iterative Information Processing (ISTC)*, Dec. 2018, pp. 1–5.

[7] N. Hussami, S. B. Korada, and R. Urbanke, "Performance of polar codes for channel and source coding," in *International Symposium on Information Theory* (ISIT), Jul. 2009, pp. 1488–1492.

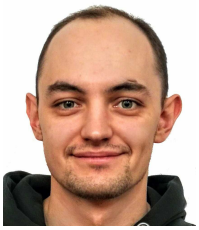
[8] A. Canteaut and F. Chabaud, "A new algorithm for finding minimum-weight words in a linear code: Application to McEliece's cryptosystem and to narrow-sense BCH codes of length 511," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 367–378, Jan. 1998.

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

[10] P. Trifonov and G. Trofimiuk, "A randomized construction of polar subcodes," in *International Symposium on Information Theory* (ISIT), Jun. 2017, pp. 1863–1867.

[11] I. Tal and A. Vardy, "List decoding of polar codes," *IEEE Transactions On Information Theory*, vol. 61, no. 5, pp. 2213–2226, May 2015.

[12] P. Trifonov, "Efficient design and decoding of polar codes," *IEEE Transactions on Communications*, vol. 60, no. 11, pp. 3221 – 3227, Nov. 2012.



Ruslan Morozov (S'16-M'19) was born in Saint Petersburg, Russia, in 1991. He received the MSc. degree in computer science in 2014. He is currently a researcher with the Institute of Computer Science and Technology, Saint-Petersburg Polytechnic University. His research interests include coding theory and its applications.

Peter Trifonov (S'02-M'05) was born in Saint Petersburg, Russia, in 1980. He received the PhD (Candidate of Science) degree from Saint-Petersburg Polytechnic University in 2005, and Doctor of Science degree from the Institute for Information Transmission Problems in 2018. He is currently an Associate Professor with the Institute of Computer Science and Technology, Saint-Petersburg Polytechnic University. His research interests include coding theory and its applications in telecommunications and other areas. He is an associate editor of IEEE Transactions on Communications, and the chairman of the IEEE-Russia IT Joint Sections Information Theory Society Chapter.