Efficient Design and Decoding of Polar Codes

Peter Trifonov, *Member, IEEE*

***Abstract*—Polar codes are shown to be instances of both generalized concatenated codes and multilevel codes. It is shown that the performance of a polar code can be improved by representing it as a multilevel code and applying the multistage decoding algorithm with maximum likelihood decoding of outer codes. Additional performance improvement is obtained by replacing polar outer codes with other ones with better error correction performance. In some cases this also results in complexity reduction. It is shown that Gaussian approximation for density evolution enables one to accurately predict the performance of polar codes and concatenated codes based on them.**

I. INTRODUCTION

Polar codes were recently shown to achieve the capacity of discrete input memoryless output symmetric channels [1]. Classes of polar codes with high error exponents were proposed in [2], [3]. However, the practical performance of polar codes under the successive cancellation (SC) decoding reported up to now turns out to be worse than that of LDPC and Turbo codes. Furthermore, construction of polar codes requires employing density evolution. Careful implementation is needed to avoid quantization errors while computing the probability densities of log-likelihood ratios within the SC decoder. An implementation of density evolution with complexity $O(n\mu^2 \log \mu)$ was proposed in [4], where n is the length of the polar code to be constructed, and μ is the number of quantization levels, which has to be selected sufficiently high to achieve the required accuracy.

This paper demonstrates that polar codes can be efficiently constructed using Gaussian approximation for density evolution. Furthermore, it is shown that polar codes can be treated in the framework of multilevel coding. This enables one to improve the performance of polar codes by considering them as multilevel or, equivalently, generalized concatenated (GCC) ones, and using block-wise near-maximum-likelihood decoding of outer codes. In some cases this results also in reduced decoding complexity. The second contribution of the paper is a simple algorithm for construction of GCC with inner polar codes. If optimal outer codes are used, this algorithm constructs codes with substantially better performance compared to similar polar ones.

The relationship of polar and multilevel codes was first observed in the original paper [1], and the approximate instance of the SC decoding algorithm was reported already in [5] in the context of Reed-Muller codes considered as generalized concatenated ones. In this paper the theory of multilevel codes is systematically applied to improve the performance of polar codes and obtain new codes with better performance.

P. Trifonov is with the Distributed Computing and Networking Department, Saint-Petersburg State Polytechnic University, Polytechnicheskaya str., 21, office 104, 194021, Saint-Petersburg, Russia, Email: petert@dcn.ftk.spbstu.ru.

This work was partially presented at IEEE International Symposium on Wireless Communication Systems 2011.

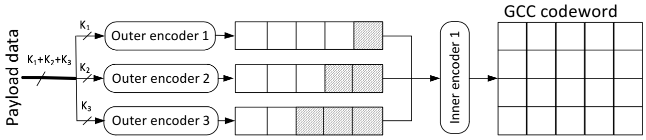


Fig. 1. Generalized concatenated code

The paper is ogranized as follows. Section II introduces the necessary background. Section III presents an algorithm for construction of polar codes based on Gaussian approximation. The relationship of polar, generalized concatenated and multilevel codes is studied in Section IV. Section V presents a construction of concatenated codes based on polar ones. Numeric results are given in Section VI. Finally, some conclusions are drawn.

II. BACKGROUND

A. Generalized concatenated codes

A generalized concatenated code ([6], [7], see [8] for detailed treatment) is constructed using a family[1] of (N, K_i, D_i) outer codes $\mathcal{C}_i$ over $GF(2^{b_i}), 1 \leq i \leq v$, and a family of nested inner (n, k_j, d_j) codes $\mathbb{C}_i$ over $GF(2)$, such that $k_j = \sum_{i=j}^{v} b_i, 1 \leq j \leq v$. Codes $\mathbb{C}_i, i > 1$, induce a recursive decomposition of code $\mathbb{C}_1$ into a number of cosets, so that

$$\mathbb{C}_i = \left\{ c + \sum_{s=1}^{b_i} u_s \mathbf{g}_{k_{i+1}+s} | c \in \mathbb{C}_{i+1}, u_s \in \{0, 1\} \right\},$$

where $\mathbf{g}_j$ denotes the rows of the generator matrix of $\mathbb{C}_1$.

The data are first encoded with outer codes to obtain codewords $(c_{1,1}, \ldots, c_{1,N}), \ldots, (c_{v,1}, \ldots, c_{v,N})$. Then for each $j = 1, \ldots, N$ the symbols $c_{ij}, 1 \leq i \leq v$, are expanded into b_i-tuples using some fixed basis of $GF(2^{b_i})$, and encoded with (n, k_1, d_1) inner code. This results in a $(Nn, \sum_{i=1}^{v} K_i b_i, \geq \min(D_1 d_1, \ldots, D_v d_v))$ linear binary code. It can be seen that the j-th symbols of outer codewords $\mathcal{C}_1, \ldots, \mathcal{C}_n$ successively select the subsets of the inner code $\mathbb{C}_1$. This eventually results in a single codeword being a subvector of a GCC codeword. Figure 1 illustrates the GCC encoding scheme.

GCC were shown to significantly outperform classical concatenated codes. In this paper only outer codes over $GF(2)$ will be considered.

[1]For the sake of simplicity we consider only the case of linear binary codes.

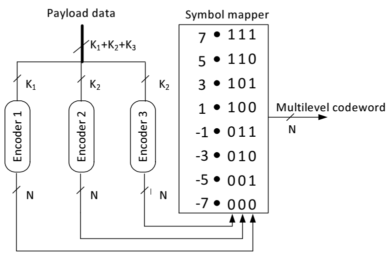


Fig. 2. Multilevel code based on 8-PAM

B. Multilevel codes

Consider some signal constellation (single- or multi-dimensional) A consisting of 2^n symbols labeled with distinct binary vectors $(x_1, \dots, x_n)$ [9], [10]. Let

$$A(u_1^{i-1}) = \left\{ a(x_1^n) \in A | x_1^{i-1} = u_1^{i-1}, x_i^n \in \{0,1\}^{n-i+1} \right\},$$

where $z_a^b = (z_a, \dots, z_b)$, and $a(x_1, \dots, x_n)$ is the symbol of A corresponding to label $(x_1, \dots, x_n)$. Let $(c_{11}, \dots, c_{1N}), \dots, (c_{n1}, \dots, c_{nN})$ be some codewords of binary codes $\mathcal{C}_1, \dots, \mathcal{C}_n$. Then a codeword of the corresponding multilevel code is given by $(a(c_{11}, \dots, c_{n1}), \dots, a(c_{1N}, \dots, c_{nN}))$. In other words, the j-th symbols of codes $\mathcal{C}_1, \dots, \mathcal{C}_n$ identify a single element of constellation A, which is used as the j-th symbol of a multilevel code codeword. This approach is exactly the same as the one used by the GCC encoder. Figure 2 illustrates this construction for the case of 8-PAM signal constellation.

Having received a vector of noisy symbols $(r_1, \dots, r_N)$, the multistage decoding algorithm proceeds by computing the log-likelihood ratios

$$L_i = \ln \frac{\sum_{a \in A(1)} P\{a|r_i\}}{\sum_{a \in A(0)} P\{a|r_i\}}, 1 \le i \le N, \quad (1)$$

and supplying it to the decoder of $\mathcal{C}_1$, which produces an estimate $(\widehat{c}_{11}, \dots, \widehat{c}_{1N})$ for the corresponding codeword. The codeword of $\mathcal{C}_2$ can be recovered in the same way, but the original signal constellation A should be replaced in (1) with its subset $A(\widehat{c}_{1i})$ identified by the first decoder. If the estimates $\widehat{c}_{1i}$ are correct, this essentially improves the reliability of the input to the decoder of $\mathcal{C}_2$. This algorithm proceeds recursively for all levels of the code. That is, at the j-th stage the decoder observes the output of a virtual channel given by not only $(r_1, \dots, r_N)$, but also $(c_{i1}, \dots, c_{iN}), 1 \le i < j$.

Multilevel codes can be treated as an instance of GCC [8] .

C. Polar codes

Consider a binary input output symmetric memoryless channel with output probability density function $W(y|x), y \in \mathcal{Y}, x \in \mathbb{F}_2$. It can be transformed into a vector channel given by $W_n(y_1^n|u_1^n) = W^n(y_1^n|u_1^n G_n)$, where $W^n(y_1^n|x_1^n) = \prod_{i=1}^n W(y_i|x_i)$, $G_n = B_s F^{\otimes s}$, $n = 2^s$, $F = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, $\otimes s$ denotes s-times Kronecker product of a matrix with itself, and B_s is a $2^s \times 2^s$ bit reversal permutation matrix. This channel is obtained by transmitting the elements of $x_1^n = u_1^n G_n$ over n copies of the original channel $W(y_i|x_i)$. The vector channel can be further decomposed into equivalent subchannels

$$W_n^{(i)}(y_1^n, u_1^{i-1}|u_i) = \frac{1}{2^{n-1}} \sum_{u_{i+1}^n} W_n(y_1^n|u_1^n). \quad (2)$$

Here $(y_1^n, u_1^{i-1}) \in \mathcal{Y}^n \times \mathbb{F}_2^{i-1}$ corresponds to the output of the i-th subchannel, and u_i to its input. The values of u_1^{i-1} are assumed to be available at the receiver side. For example, they can be obtained as (presumably correct) decisions made by the decoder for other channels. It was shown in [1] that the sum capacity of the transformed channel is equal to the capacity of the original vector channel W^n, and for $n \to \infty$ the capacities of $W_n^{(i)}$ converge either to 0 or to 1. Symbols u_i to be transmitted over low-capacity subchannels can be frozen (i.e. set to 0 at the transmitter side). This results in a linear block code.

Given y_1^n and estimates $\widehat{u}_1^{i-1}$ of u_1^{i-1}, the SC decoding algorithm attempts to estimate u_i. This can be implemented by computing the following log-likelihood ratios $L_n^{(i)}(y_1^n, \widehat{u}_1^{i-1}) = \log \frac{W_n^{(i)}(y_1^n, \widehat{u}_1^{i-1}|u_i=0)}{W_n^{(i)}(y_1^n, \widehat{u}_1^{i-1}|u_i=1)}$ [1], [11]:

$$\begin{aligned} L_n^{(2i-1)}&(y_1^n, \widehat{u}_1^{2i-2}) \\ &= 2\tanh^{-1}(\tanh(L_{n/2}^{(i)}(y_1^{n/2}, \widehat{u}_{1,e}^{2i-2} \oplus \widehat{u}_{1,o}^{2i-2})/2) \\ &\qquad \times \tanh(L_{n/2}^{(i)}(y_{n/2+1}^n, \widehat{u}_{1,e}^{2i-2})/2)), \quad (3) \\ L_n^{(2i)}&(y_1^n, \widehat{u}_1^{2i-1}) = L_{n/2}^{(i)}(y_{n/2+1}^n, \widehat{u}_{1,e}^{2i-2}) \\ &\qquad + (-1)^{\widehat{u}_{2i-1}} L_{n/2}^{(i)}(y_1^{n/2}, \widehat{u}_{1,e}^{2i-2} \oplus \widehat{u}_{1,o}^{2i-2}) (4) \end{aligned}$$

where $\widehat{u}_{1,e}^i$ and $\widehat{u}_{1,o}^i$ are subvectors of $\widehat{u}_1^i$ with even and odd indices, respectively, and $L_1^{(i)}(y_i) = \log \frac{W(y_i|0)}{W(y_i|1)}$. By employing the min-sum approximation, one obtains the decoding algorithm for Reed-Muller codes presented in [5].

It is sufficient to perform the error probability analysis only for the case of all-zero codeword. Density evolution can be used to compute the probability density functions $p_i(x)$ of $L_n^{(i)}(y_1^n, \widehat{u}_1^{i-1})$ from the PDF of $L_1^{(i)}(y_i)$ [12]. Then the error probability for the i-th subchannel can be obtained as $\pi_i = \int_{-\infty}^0 p_i(x)dx$. To obtain (n,k) polar code, one should set at the transmitter $u_i = 0$ for $n-k$ subchannels with the highest π_i. That is, the polar code generator matrix is given by $G = AF^{\otimes s}$, where A is a $k \times n$ submatrix of B_s obtained by taking the rows corresponding to the active subchannels. It was shown in [4] that density evolution for polar codes can be implemented with complexity $O(n\mu^2 \log \mu)$, where μ is the number of quantization levels, which has to be set sufficiently high to avoid catastrophic loss of precision.

III. Design of polar codes based on Gaussian approximation

The main drawback of the polar code construction method based on density evolution is its high computational complexity. The most practically important case corresponds to the AWGN channel. In this scenario $L_1^{(i)}(y_i) \sim \mathcal{N}(\frac{2}{\sigma^2}, \frac{4}{\sigma^2})$, provided that the all-zero codeword is transmitted. It was suggested in [13] to approximate the distributions of intermediate values arising in the belief propagation decoding algorithm for LDPC codes with Gaussian ones. This substantially simplifies the analysis. Since the transformations performed by the SC decoding algorithm are essentially the same as in the case of belief propagation decoding, this approach can be extended to the case of polar codes. Namely, the values given by (3)–(4) can be considered as Gaussian random variables with $\mathbf{D}[L_n^{(i)}] = 2\,\mathbf{E}[L_n^{(i)}]$, where $\mathbf{E}$ and $\mathbf{D}$ are the mean and variance, respectively. This enables one to compute only the expected value of $L_n^{(i)}$, drastically reducing thus the complexity. In the case of polar codes this approach reduces to

$$\begin{aligned}\mathbf{E}[L_n^{(2i-1)}] &= \phi^{-1}\left(1-\left(1-\phi\left(\mathbf{E}[L_{n/2}^{(i)}]\right)\right)^2\right) && (5)\\ \mathbf{E}[L_n^{(2i)}] &= 2\,\mathbf{E}[L_{n/2}^{(i)}], && (6)\end{aligned}$$

where

$$\phi(x)=\begin{cases}1-\frac{1}{\sqrt{4\pi x}}\int_{-\infty}^{\infty}\tanh\frac{u}{2}e^{-\frac{(u-x)^2}{4x}}dx, & x>0\\ 1, & x=0.\end{cases}$$

The error probability for each subchannel is given by [14]

$$\pi_i \approx Q\left(\sqrt{\mathbf{E}[L_n^{(i)}]/2}\right), 1\le i\le n. \quad (7)$$

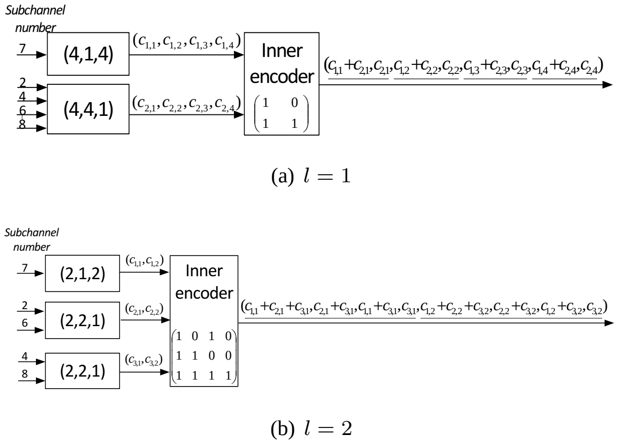
It can be seen that the cost of computing π_i is given by $O(n\log n)$. Similar approach was considered in [15].

IV. Decomposition of polar codes

Direct calculation of (7) shows that the rate of channel polarization is quite low, i.e. for practical values of codelength n there are many subchannels with quite high error probability π_i. These subchannels have to be used for data transmission in order to obtain a code with reasonable rate. However, the errors occuring in these subchannels at some steps of the standard SC decoding algorithm cannot be corrected at the subsequent steps, and the overall performance of a polar code is dominated by the performance of the worst subchannel. The proposed approach avoids this problem by performing joint decoding over a number of subchannels.

A. Generalized concatenated polar codes

The recursive structure of polar codes enables one to consider them as GCC. Namely, the generator matrix of a polar code can be represented as $G=AF^{\otimes s}=A(F^{\otimes(s-l)}\otimes F^{\otimes l})$, where A is a full-rank matrix with at most one non-zero element in each column. Then the encoding operation can be considered as partitioning of the data vector u into 2^l subvectors, multiplication of these subvectors by some submatrices given by rows of $F^{\otimes(s-l)}$, row-wise arrangement of the obtained vectors into a table, and column-wise multiplication of this table by matrix $F^{\otimes l}$. This is equivalent to encoding the data with a GCC based on 2^l outer codes $\mathcal{C}_i$ of length $N=2^{s-l}$, and inner codes $\mathbb{C}_i$ of length $n=2^l$ generated by rows $i,\dots,2^l$ of matrix $B_lF^{\otimes l}$. The generator matrices of the $(1+R(i,l))$-th outer code $\mathcal{C}_i$ is obtained by taking rows $1+R(j,s-l)$ of $F^{\otimes(s-l)}$, such that row $1+R(i2^{s-l}+j,s)$



(a) $l=1$

(b) $l=2$

Fig. 3. Representation of $(8,5,2)$ polar code as GCC

of $F^{\otimes s}$ is included into the generator matrix of the original polar code, where $0\le i<2^l, 0\le j<2^{s-l}$, and

$$R\left(\sum_{j=0}^{m-1}2^j i_j, m\right)=\sum_{j=0}^{m-1}2^j i_{m-1-j}, i_j\in\{0,1\}.$$

Observe that both $\mathbb{C}_i$ and $\mathcal{C}_i$ are also instances of polar codes. This will be denoted by degree-l decomposition.

Example 1. *Consider a $(8,5,2)$ polar code with generator matrix [16]*

$$G=\begin{pmatrix}1&1&0&0&0&0&0&0\\1&1&1&1&0&0&0&0\\1&1&0&0&1&1&0&0\\1&0&1&0&1&0&1&0\\1&1&1&1&1&1&1&1\end{pmatrix}. \quad (8)$$

This matrix corresponds to active subchannels 2,4,6,7,8 of the polarizing transformation given by $F^{\otimes 3}$. For $l=1$, this code can be decomposed into $(4,1,4)$ and $(4,4,1)$ outer codes (the generator matrix of the latter one is given by $B_2F^{\otimes 2}=\begin{pmatrix}1&0&0&0\\1&0&1&0\\1&1&0&0\\1&1&1&1\end{pmatrix}$). Inner code $\mathbb{C}_1$ is given by the row space of F (see Figure 3(a)). Alternatively, for $l=2$ the parameters of outer codes are $(2,0,\infty),(2,1,2),(2,2,1),(2,2,1)$, and the inner code $\mathbb{C}_1$ is generated by $B_2F^{\otimes 2}$. However, one can eliminate the first empty outer code and the first row from the generator matrix of the inner code, and obtain the constuction shown in Figure 3(b).

The above described decomposition of polar codes enables one to perform block-wise decoding of outer codes. This reduces the probability of propagation of incorrect information bit estimates, which sacrifices the performance of the SC decoding algorithm. Since the length of outer codes is relatively small, one can efficiently implement near-maximum-likelihood decoding algorithms for them. On the other hand, low-complexity SC decoding based on expressions (3)–(4) can be used for processing of inner codes.

The performance of the proposed algorithm is not worse than that of the original SC decoder. To see this observe that the SC algorithm essentially follows the same scheme, but recursively employs itself for decoding of the outer codes. Obviously, if the SC decoder is able to make correct decisions for the payload symbols corresponding to a single outer code, the ML decoder for this code can do it too.

It was shown in [2] that channel polarization can be also performed by high-dimensional kernels F. The proposed decomposition method applies to this case too.

B. Multilevel polar codes

GCC introduced in Section IV-A can be also treated in the framework of multilevel coding. It can be seen that the concepts of equivalent subchannels and the SC decoding algorithm are very similar to the construction of multilevel codes and the multistage decoding algorithm.

In the context of polar codes, signal constellation A is given by 2^n binary n-vectors $a(u)$, which can be obtained as $a(u) = uB_lF^{\otimes l}, u \in GF(2)^n$, where $n = 2^l$. This constellation is recursively partitioned into subsets $A(u_1^i)$ by fixing the values of $u_1, \ldots, u_i$. The elements of u are obtained as codeword symbols of outer codes $\mathcal{C}_i$ of length $N = 2^{s-l}$. That is, one can construct N vectors $u^{(j)} = (c_{1,j}, \ldots, c_{n,j}), 1 \leq j \leq N$, where $(c_{i,1}, \ldots, c_{i,N}) \in \mathcal{C}_i, 1 \leq i \leq n$, and obtain a multilevel codeword $(u^{(1)}B_lF^{\otimes l}, \ldots, u^{(N)}B_lF^{\otimes l})$.

Example 2. *Let us proceed with the code given by* (8)*. For* $l = 1$ *the signal constellation is given by* $GF(2)^2$*. It is partitioned into subsets* $A(0) = \{00, 11\}$ *and* $A(1) = \{10, 01\}$*. Codeword symbols of* $(4, 1, 4)$ *code* $\mathbb{C}_1$ *are used to select a subset, while the symbols of the* $(4, 4, 1)$ *code* $\mathbb{C}_2$ *identify the particular constellation elements to be transmitted. For* $l = 2$ *the signal constellation is* $GF(2)^4$*, but since* $\mathbb{C}_1$ *is an empty code, it is effectively reduced to the set of all even-weight vectors of length* 4*. On Figure 3 the subvectors of the polar codeword corresponding to a single constellation element (i.e. codeword of the inner code) are underlined.*

Observe that the decoding algorithm outlined in the previous section represents an instance of multistage decoding. Indeed, it involves computing the log-likelihood ratios for $u_{i,j}$ according to (3)–(4), and passing them to a decoder of $\mathcal{C}_i$, which produces a codeword estimate $(\widehat{c}_{i,1}, \ldots, \widehat{c}_{i,N}) \in \mathcal{C}_i$. This codeword is utilized in the subsequent step of the multistage decoding algorithm to select an appropriate coset of $\mathbb{C}_1$ (i.e. a subset of the signal constellation). These operations are performed for all n levels of the constellation partitioning chain. Block-wise decoding of outer codes enables one to reduce the error probability for the case of unreliable subchannels $W_{Nn}^{(i)}, 1 \leq i \leq nN$. The complexity of this algorithm will be analyzed in Section V-D.

It appears that the subchannels in the sense of polar codes (see (2)) are equivalent to subchannels in the sense of multilevel codes. Indeed, the likelihood ratio for $c_{i,j}$ (for brevity, the second index will be omitted in this derivation) in the case of polar codes of length n depends both on real channel output y_1^n and "genie hint" $c_1^{i-1} = u_1^{i-1}$. That is,

$$\begin{aligned} \lambda_i(y_1^n, u_1^{i-1}) &= \frac{W_n^{(i)}(y_1^n, u_1^{i-1}|u_i = 0)}{W_n^{(i)}(y_1^n, u_1^{i-1}|v_i = 1)} \\ &= \frac{W_n^{(i)}(y_1^n|u_1^{i-1}, u_i = 0)P\left\{u_1^{i-1}|u_i = 0\right\}}{W_G^{(i)}(y_1^n|u_1^{i-1}, u_i = 1)P\left\{u_1^{i-1}|u_i = 1\right\}} \\ &= \frac{W_n^{(i)}(y_1^n|u_1^{i-1}, v_i = 0)}{W_n^{(i)}(u_1^n|u_1^{i-1}, u_i = 1)}. \end{aligned}$$

This is essentially the likelihood ratio for the case of the subchannel at level i of the multilevel code, provided that the decisions at the previous levels are correct. Since the distributions of likelihood ratios for subchannels of polar and multilevel codes are identical, their capacities are the same. The representation of polar codes as multilevel ones seems to be more natural, since it avoids the expansion of channel output alphabet by treating u_1^{i-1} as channel parameters.

V. Concatenated codes based on polar codes

It must be recognized that the GCC obtained by decomposing a polar code may not be optimal from the point of view of multilevel coding. The similarity of the polar code constuction and the above described decoding algroithm with multilevel codes and multistage decoding, respectively, suggests employing multilevel code design rules for selection of parameters of the coding scheme described above. That is, the performance of a polar code under the multistage decoding with block-wise maximum-likelihood decoding of outer codes can be improved by changing the set of frozen bits. Furthermore, if the algorithm used to perform block-wise decoding of outer codes does not take into account their structure, one can use any linear block code with suitable parameters, not necessary polar, as $\mathcal{C}_i$. This enables one to employ outer codes with better error correction performance.

The following subsections present a reformulation of the multilevel code design rules (see [10]) to the case of the signal constellation given by the row space of matrix $B_lF^{\otimes l} = \mathbb{F}_2^n$.

A. Capacity rule

The rate R_i of $\mathcal{C}_i$ should be chosen equal to the capacity C_i of the i-th subchannel of the multilevel code, which is induced by matrix $B_lF^{\otimes l}$. According to [10], one obtains

$$C_i = I(y_1^n; u_i|u_1^{i-1}) = \mathbf{E}_{u_1^{i-1}}[C(A(u_1^{i-1}))] - \mathbf{E}_{u_1^i}[C(A(u_1^i))], \quad (9)$$

where

$$C(B) = \int_{\mathbb{R}^n} \sum_{a \in B} \frac{W^n(y_1^n|a)}{|B|} \log_2 \left(\frac{|B|W^n(y_1^n|a)}{\sum\limits_{b \in B} W^n(y_1^n|b)} \right) dy_1^n \quad (10)$$

is the capacity when using the subset B of $\mathbb{F}_2^n$ for transmission over the vector channel $W^n(y_1^n|x_1^n)$. In the case of binary-input memoryless output symmetric channels, one can drop the expectation operator in (9) to obtain $C_i = C(A^{(i-1)}) - C(A^{(i)})$, where $A^{(i)} = A(\underbrace{0, \ldots, 0}_{i \text{ times}})$. It can be seen that the latter set is a linear block code $\mathbb{C}_i$ generated by $l-i$ last rows of $B_l F^{\otimes l}$. The expression (10) can be further simplified to

$$C(A^{(i)}) = \int_{\mathbb{R}^N} \prod_{j=1}^{N} W(y_j|0) \log_2 \left(\frac{|\mathbb{C}_i| \prod_{j=1}^{N} W(y_j|0)}{\sum_{b\in\mathbb{C}_i} \prod_{j=1}^{N} W(y_j|b_j)} \right) dy_1^N.$$

Hence, the capacity of the i-th subchannel of the multilevel polar code can be computed as

$$C_i = \int_{\mathbb{R}^N} \prod_{j=1}^{N} W(y_j|0) \log_2 \left(\frac{2 \sum_{b\in\mathbb{C}_{i+1}} \prod_{j=1}^{N} W(y_j|b_j)}{\sum_{b\in\mathbb{C}_i} \prod_{j=1}^{N} W(y_j|b_j)} \right) dy_1^N. \quad (11)$$

Obviously, employing this rule results in a capacity-achieving concatenated code, provided that the outer codes can achieve the capacity too. However, evaluating (11) seems to be a difficult task.

B. Balanced minimum distances rule

The classical approach to the design of GCC is to select $D_i d_i \approx const$. However, as it was shown in [10], this forces one to select for some channels codes with rate exceeding their capacities, while the error correction capability of other codes may be excessive for their channels. This results in too high error coefficient of the obtained code. It can be seen that the Reed-Muller codes are designed according to this rule.

C. Equal error probability rule

More practical approach can be based on selection of outer codes $\mathcal{C}_i$ so that the decoding error probability is approximately the same for all subchannels. This requires one to be able to compute the decoding error probability for all possible component codes. For instance, one can derive distance profiles for each level of the multilevel code (see [17]), and employ union bound to estimate the decoding error probability in the case of multistage decoding. Alternatively, assuming the validity of Gaussian approximation introduced in section III, one can study (e.g. via simulations) the performance of possible component codes in the case of AWGN channel with noise variance $2/L_{2^l}^{(i)}$, and use these results to estimate their performance in the equivalent subchannels of the multilevel code. In what follows, the latter approach will be used, since it is simpler to implement and allows one to take into account the performance of non-maximum likelihood decoding algorithms for outer codes.

The probability of incorrect decoding of a binary linear block code $\mathcal{C}$ can be obtained as $p_e \leq \sum_{c\in\mathcal{C}\setminus\{\mathbf{0}\}} P\left\{w(c)<0\right\}$, where $w(c)=\sum_{i:c_i\neq 0} L_i$, and $L_i=\ln\frac{P\{c_i=0|y_i\}}{P\{c_i=1|y_i\}}$ [14]. In the case of multilevel polar codes, L_i are computed by the SC decoding algorithm for the inner code. Assuming the validity of Gaussian approximation

CodeOptimization(σ, R, N, l)

$\mathbf{E}[L_1^{(1)}] \leftarrow 2/\sigma^2$
Compute $m_i=\mathbf{E}[L_{2^l}^{(i)}], 1\leq i\leq 2^l$ via (5)–(6)
$P' \leftarrow 1; P'' \leftarrow 0$
while $P'-P'' > \epsilon P'$
do $\tilde{P} \leftarrow (P'+P'')/2$
  $t_i \leftarrow \arg\max_{t:P_t(m_i)\leq\tilde{P}} K_t, 1\leq i\leq 2^l$
  $K \leftarrow \sum_{i=1}^{2^l} K_{t_i}$
  if $K < RN2^l$
    then $P'' \leftarrow \tilde{P}$
    else $P' = \tilde{P}$
return $(K_{t_1},\dots,K_{t_{2^l}}), \tilde{P}$

Fig. 4. Design of a GCC according to the equal error probability rule

for (3)–(4), $w(c)$ can be also approximated as a Gaussian random variable. Hence, one obtains

$$p_e \leq \sum_{j=1}^{N} A_j Q\left(\sqrt{\frac{\mathbf{E}[L_i]}{2} j}\right),$$

where A_i are weight spectrum coefficients of code $\mathcal{C}$, and d is its minimum distance. Since it is in general difficult to obtain code weight spectrum, and union bound is known to be not tight in the low-SNR region, one can use simulations to obtain a performance curve for the case of AWGN channel and some fixed (probably, non-ML) decoding algorithm, and use least squares fitting to find suitable α and δ, so that the decoding error probability is given by

$$p_e(m) \approx \alpha Q\left(\sqrt{\frac{m}{2}\delta}\right), \quad (12)$$

where $m=\mathbf{E}[L_i]$.

Assume now that the outer codes $\mathcal{C}_i$ are selected from some family of error-correcting codes (not necessary polar) of length N. Let K_t, D_t and $P_t(m)$ be the dimension, minimum distance and decoding error probability function for the t-th code, respectively, where m is the expected value of LLR. Let us further assume that $K_0=P_0(m)=0$ and $P_i(m) < P_j(m) \Leftrightarrow K_i < K_j$ (this is true if $K_i<K_j \Leftrightarrow D_i > D_j$, and m is sufficiently large). Figure 4 presents a simple algorithm for construction of a generalized concatenated (multilevel) code of rate R according to the equal error probability rule. The algorithm employs the bisection method to approximately solve the equation $\sum_{i=1}^{2^l} K(i,P)=RN2^l$, where $K(i,P)$ is the maximum dimension of a code capable of achieving error probability P at the i-th subchannel. The parameter ϵ is a sufficiently small constant, which affects the precision of the obtained estimate for P. The code is optimized for the case of AWGN channel with noise variance σ^2. The algorithm returns the dimensions of optimal codes for each level, as well as an estimate for the decoding error probability for each code.

The SC/multistage decoder produces an error if decoding of any of the component codes is incorrect. Therefore, the overall error probability of the GCC can be computed as

$$\begin{aligned} P &= 1-P\left\{C_1,\ldots,C_n\right\} \\ &= 1-P\left\{C_1\right\}P\left\{C_2|C_1\right\}\cdots P\left\{C_n|C_1,\ldots,C_{n-1}\right\} \\ &\approx 1-\prod_{i=1}^{n}(1-P_{t_i}(m_i)) \approx 1-(1-\tilde{P})^n, \end{aligned} \quad (13)$$

where C_i denotes the event of correct decoding of the outer code at the i-th level, $\tilde{P}$ is the quantity computed by the above algorithm, and t_i is the index of the code selected for the i-th subchannel. This expression enables semi-analytic prediction of the performance of the concatenated code, based on the available performance results for component outer codes.

Concatenated coding schemes similar to the one described above were proposed in [18], [19]. However, these papers do not address the problem of outer code rate optimization in a systematic way.

D. Decoding complexity

One can use any suitable algorithm to implement soft-decision decoding of outer codes in the GCC obtained either by decomposing a polar code, or constructed explicitly using the algorithm in Figure 4. Box-and-match algorithm is one of the most efficient methods to perform near maximum-likelihood decoding of short linear block codes [20]. Its worst-case complexity for the case of (N,K) code with order t reprocessing is given by $O((N-K)K^t) = O(N^{t+1})$, although in practice it turns out to be much more efficient. Decoding of a concatenated code of length $\nu = Nn$ involves decoding of N inner codes using the SC algorithm, and decoding of n outer codes. Therefore the overall complexity is given by $O(N^{t+1}nC_b + Nn\log nC_s)$, where C_b and C_s are some factors which reflect the cost of elementary operations performed by these algorithms. While the overall complexity is asymptotically dominated by the cost of box-and-match decoding and is higher than that of the SC algorithm, which has complexity $O(\nu\log\nu C_s)$, the proposed approach may result in practice in lower number of arithmetic operations, since the length of the component codes is much smaller than the length ν of the original code, and the cost C_b of elementary operations of the former algorithm (add and compare) is much smaller than the cost C_s of evaluating $\tanh(x)$.

VI. Numeric results

Figure 5 presents simulation results illustrating the accuracy of bit error rate analysis based on the Gaussian approximation. Simulations were performed for the case of $2^{10}\times 2^{10}$ polarizing transformation and AWGN channel with noise variance $N_0/2 = 1$. Error-free values $\widehat{u}_1^{i-1} = u_1^{i-1}$ were used in the SC decoding algorithm while estimating u_i to eliminate error propagation. Transmission of 10^6 data blocks was simulated. Each point on the figure corresponds to a particular subchannel and presents actual vs. estimated bit error rate. It can be seen that except for a few very bad channels Gaussian approximation provides very accurate results, although it slightly overestimates the error probability. The discrepancy in the low-BER range is caused mostly by the simulation inaccuracy.

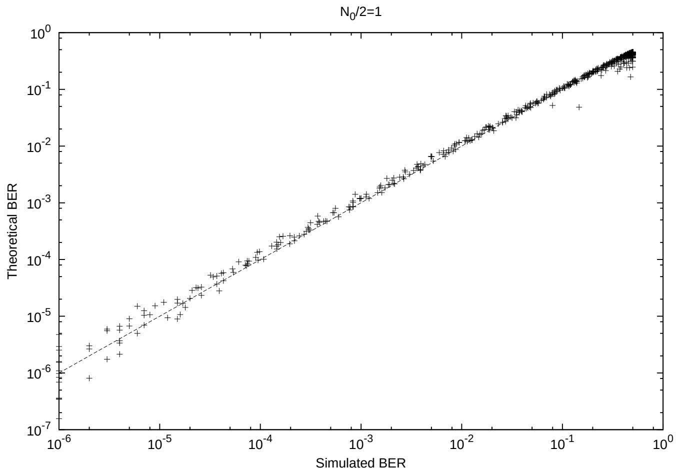


Fig. 5. Accuracy of Gaussian approximation

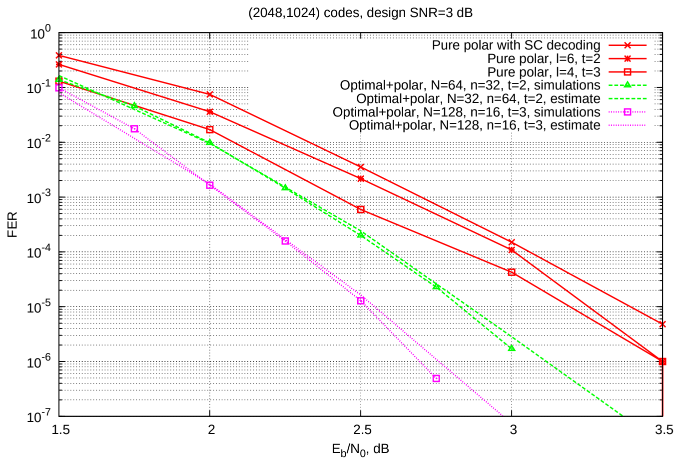


Fig. 6. Performance of polar and concatenated codes

Observe that there are many subchannels with medium bit error rate, which require additional layer of coding to achieve reliable data transmission.

Figure 6 presents the performance of polar codes of length 2048 designed using the Gaussian approximation method for the case of AWGN channel with $E_b/N_0 = 3$ dB. Both pure SC and multistage decoding algorithms were considered. For multistage decoding, degree l decomposition of the original polar code was performed, and box-and-match algorithm with order t reprocessing was used for decoding of outer polar codes [20]. Table I presents the normalized decoding time T_i/T_0 for the considered cases, where T_0 is the time needed

TABLE I
RELATIVE DECODING COMPLEXITY FOR $(2048,1024)$ CODES

Design SNR	2 dB	3 dB
Pure polar with SC decoding	1	1
Pure polar, $l=6, t=2$	0.24	0.24
Pure polar, $l=4, t=3$	4.92	2.9
Optimal+polar, $N=32, n=64, t=2$	0.26	0.24
Optimal+polar, $N=64, n=32, t=3$	0.31	0.36
Optimal+polar, $N=128, n=16, t=3$	4.34	3.11

to decode plain polar code using the SC algorithm, and T_i is the time needed to decode the corresponding code using the multistage decoding algorithm.

It can be seen that block-wise decoding of outer codes provides up to 0.25 dB performance gain compared to SC decoding. Higher values of N do not provide any noticeable performance improvement. The figure presents also the performance of GCC based on inner polar codes and outer optimal linear block codes [21], [22] with multistage decoding[2]. It can be seen that increasing the length of outer codes provides additional 0.5 dB performance gain. This is due to much higher minimum distance of optimal codes compared to polar codes of the same length, obtained by decomposing the polar code of length Nn. It can be also seen that expression (13) provides very good estimate for the decoding error probability of the concatenated code. For long outer codes the actual performance turns out to be slightly better. This is due to slightly pessimistic estimates of subchannel quality produced by the Gaussian approximation for density evolution, as it was shown in Figure 5. Furthermore, in some cases the proposed decomposition results in more efficient decoding. This is due to high efficiency of the box-and-match algorithm for short codes, which does not need to evaluate the $\tanh(\cdot)$ function.

VII. Conclusions

It was shown in this paper that polar codes can be considered as multilevel (generalized concatenated) ones, and the techniques developed in the area of multilevel coding and multistage decoding can be applied to their analysis. In particular, this enables one to perform joint decoding for a number of information symbols using any maximum likelihood decoding algorithm for short linear block codes. This results in performance improvement, since the standard SC decoding algorithm cannot correct the erroneous decisions made at early steps. Furthermore, this enables one to use arbitrary codes as outer ones in this construction. It was shown in this paper that this results in significant performance improvement, and, in some cases, in complexity reduction. It was also demonstrated that the performance of polar codes and concatenated codes based on them can be efficiently studied using the Gaussian approximation for density evolution. This enables one to predict their performance in the high-SNR region without simulations.

Acknowledgment

The author thanks the anonymous reviewers for many helpful comments, which have greatly improved the quality of the paper.

This work was supported by Russian Ministry of Education and Science under the contract 7.514.11.4101.

References

[1] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. On Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.
[2] S. B. Korada, E. Sasoglu, and R. Urbanke, "Polar codes: Characterization of exponent, bounds, and constructions," *IEEE Trans. On Inf. Theory*, vol. 56, no. 12, December 2010.
[3] R. Mori and T. Tanaka, "Non-binary polar codes using Reed-Solomon codes and algebraic geometry codes," in *Proc. of IEEE Inf. Theory Workshop*, 2010.
[4] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Trans. On Inf. Theory*, 2011, submitted for publication.
[5] G. Schnabl and M. Bossert, "Soft-decision decoding of Reed-Muller codes as generalized multiple concatenated codes." *IEEE Trans. on Inf. Theory*, vol. 41, no. 1, pp. 304–308, 1995.
[6] E. Blokh and V. Zyablov, "Coding of generalized concatenated codes," *Problems of Inf. Transmission*, vol. 10, no. 3, pp. 45–50, 1974.
[7] V. Zinov'ev, "Generalized cascade codes," *Problems of Inf. Transmission*, vol. 12, no. 1, pp. 5–15, 1976.
[8] M. Bossert, *Channel coding for telecommunications*. Wiley, 1999.
[9] H. Imai and S. Hirakawa, "A new multilevel coding method using error correcting codes," *IEEE Trans. on Inf. Theory*, vol. 23, no. 3, pp. 371–377, May 1977.
[10] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. On Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, July 1999.
[11] R. Mori and T. Tanaka, "Performance of polar codes with the construction using density evolution," *IEEE Comm. Letters*, vol. 13, no. 7, July 2009.
[12] ——, "Performance and construction of polar codes on symmetric binary-input memoryless channels," in *Proc. of IEEE Int. Symp. on Inf. Theory*, 2009.
[13] 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. on Inf. Theory*, vol. 47, no. 2, February 2001.
[14] J. G. Proakis, *Digital communications*. McGraw Hill, 1995.
[15] S. B. Korada, A. Montanari, E. Telatar, and R. Urbanke, "An empirical scaling law for polar codes," in *Proc. of IEEE Int. Symp. on Inf. Theory*, 2010.
[16] E. Arikan, "A performance comparison of polar codes and Reed-Muller codes," *IEEE Comm. Letters*, vol. 12, no. 6, June 2008.
[17] J. Huber, "Multilevel codes: Distance profiles and channel capacity," in *ITG-Fachbericht 130, Conf. Rec.*, October 1994, pp. 305–319.
[18] E. Arikan and G. Markarian, "Two-dimensional polar coding," in *Proc. of 10'th Int. Symp. on Comm. Theory and Applications*, Ambleside, UK, 2009.
[19] M. Seidl and J. B. Huber, "Improving successive cancellation decoding of polar codes by usage of inner block codes," in *Proc. of 6th Int. Symp. on Turbo Codes and Iterative Information Processing*, 2010, pp. 103 – 106.
[20] A. Valembois and M. Fossorier, "Box and match techniques applied to soft-decision decoding," *IEEE Trans. on Inf. Theory*, vol. 50, no. 5, May 2004.
[21] M. Grassl, "Bounds on the minimum distance of linear codes and quantum codes," Online available at http://www.codetables.de, 2007, accessed on 2011-12-17.
[22] ——, "Searching for linear codes with large minimum distance," in *Discovering Mathematics with Magma — Reducing the Abstract to the Concrete*, ser. Algorithms and Computation in Mathematics, W. Bosma and J. Cannon, Eds. Heidelberg: Springer, 2006, vol. 19, pp. 287–313.

[2]The dimensions of outer codes for the case $N = 128$ are $0, 12, 4, 92, 2, 86, 72, 119, 1, 77, 55, 116, 37, 114, 112, 126$.

Peter Trifonov was born in St.Petersburg, USSR in 1980. He received the MSc degree in computer science in 2003, and PhD (Candidate of Science) degree from St.Petersburg State Polytechnic University in 2005. Currently he is an Associate Professor at the Distributed Computing and Networking department of the same university. His research interests include coding theory and its applications in telecommunications and other areas. Since January, 2012 he is serving as a vice-chair of the IEEE Russia Joint Sections Information Theory Society Chapter.