Generalized Concatenated Codes Based on Polar Codes

Peter Trifonov, Piotr Semenov
Distributed Computing and Networking Department
Saint-Petersburg State Polytechnic University
Polytechnicheskaya str., 21, office 104
194021, Saint-Petersburg, Russia
Email: {petert,spk}@dcn.ftk.spbstu.ru

***Abstract*—Polar codes are demonstrated to be instances of both generalized concatenated codes and multilevel codes. It is shown that Gaussian approximation for density evolution enables one to accurately predict the performance of polar codes. A construction of generalized concatenated codes is proposed, which is based on the equal error probability design rule originally developed in the context of multilevel codes.**

I. Introduction

Polar codes were recently shown to achieve the capacity of discrete input memoryless symmetric channels [1]. Classes of polar codes with high error exponents were proposed in [2], [3]. However, their practical performance turns out to be not record breaking.

This paper introduces a construction of generalized concatenated codes based on polar codes, which provides much better performance compared to plain polar codes. The performance of the considered construction strongly depends on the choice of outer code rates. We propose to employ the techniques developed in the area of multilevel codes to solve the rate allocation problem. Furthermore, we show that the performance of polar codes can be efficiently analyzed using Gaussian approximation for density evolution. It appears also that polar codes can be themselves represented as generalized concatenated codes.

The paper is ogranized as follows. Section II introduces the necessary background. Section III presents the proposed code construction and rate allocation algorithm. Numeric results are given in Section IV. Finally, some conclusions are drawn.

II. Background

A. Generalized concatenated codes

Generalized concatenated codes ([4], [5], see [6] for detailed treatment) are based on a family[1] of (N, K_i, D_i) outer codes $\mathcal{C}_i$ over $GF(2^{m_i}), 1 \leq i \leq l$, and a family of nested inner (n, k_j, d_j) codes $\mathbb{C}_i$ over $GF(2)$, such that $k_j = \sum_{i=j}^{l} m_i, 1 \leq j \leq l$. The data are first encoded with outer codes to obtain codewords $(c_{11}, \ldots, c_{1N}), \ldots, (c_{l1}, \ldots, c_{lN})$. Then for each $j = 1, \ldots, N$ the symbols $c_{ij}, 1 \leq i \leq l$, are expanded into m_i-tuples using some fixed bases of $GF(2^{m_i})$, and encoded with (n, k_1, d_1) inner code. This results in a $(Nn, \sum_{i=1}^{l} K_i m_i, \min(D_1 d_1, \ldots, D_l d_l))$ linear binary code. Generalized concatenated codes were shown to significantly outperform classical concatenated codes.

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

B. Multilevel codes

Consider some signal constellation (single- or multi-dimensional) A consisting of 2^n symbols labeled with distinct binary vectors $(x_1, \ldots, x_n)$ [7], [8]. 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 $u_1^i = (u_1, \ldots, u_i)$, and $a(x_1, \ldots, x_n)$ is the symbol corresponding to label $(x_1, \ldots, x_n)$. Let $(c_{11}, \ldots, c_{1N}), \ldots, (c_{n1}, \ldots, c_{nN})$ be some codewords of binary codes $\mathcal{C}_1, \ldots, \mathcal{C}_n$. Then a codeword of the corresponding multilevel code is given by $(a(c_{11}, \ldots, c_{n1}), \ldots, a(c_{1N}, \ldots, c_{nN}))$.

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

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

and supplying it to the decoder of $\mathcal{C}_1$, which produces an estimate $(\widehat{c}_{11}, \ldots, \widehat{c}_{1N})$ 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 subsets $A(\widehat{c}_{1i})$ identified by the first decoder. If the estimates $\widehat{c}_{1i}$ are correct, this essentially improves the reliability of the input to the decoders 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 $(r_1, \ldots, r_N)$, but also $(c_{i1}, \ldots, c_{iN}), 1 \leq i < j$.

Multilevel codes can be considered as an instance of generalized concatenated codes [6] .

C. Polar codes

Consider a binary input symmetric memoryless channel with output probability density function $W(y|x)$. 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 $W^n(y_1^n | x_1^n) = \prod_{i=1}^{n} W(y_i | x_i)$, $G_n = B_n F^{\otimes s}$, $n = 2^s$, $F = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, $\otimes s$ denotes s-times

Kronecker product of a matrix with itself, and B_n is a 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}) 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 successive cancellation 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], [9]:

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

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 $L_1^{(i)}(y_i) = \log \frac{W(y_i|0)}{W(y_i|1)}$.

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 $L_1^{(i)}(y_i)$. 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. However, implementing density evolution both accurately and efficiently turns out to be a challenging task.

III. CONCATENATED CODING SCHEME

A. Multilevel codes based on the polarization transformation

It can be seen that the concepts of equivalent subchannels and the successive cancellation 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 $a(u) = uG_n, u \in GF(2)^n$. The crucial assumption underlying the successive cancellation decoding algorithm is that the decisions $\widehat{u}_1^{i-1}$ at the previous steps are correct. It is known that the capacities of the equivalent subchannels induced by the polarizing transformation converge to 0 and 1 quite slowly, i.e. the probability of error for $i < n$ may be quite high, resulting thus in error propagation. It is natural to encode the data to be transmitted over different subchannels with appropriate outer codes $\mathcal{C}_i$ of length N to reduce the probability of this event, similarly to the construction of multilevel codes.

That is, one can construct N vectors $u^{(j)} = (u_{1,j}, \dots, u_{n,j}), 1 \le j \le N$, where $(u_{i,1}, \dots, u_{i,N}) \in \mathcal{C}_i$, and construct a multilevel codeword $(u^{(1)}G_n, \dots, u^{(N)}G_n)$. Then decoding can be performed by computing the log-likelihood ratios according to (3)–(4), and passing them to a decoder of $\mathcal{C}_i$, which produces a codeword estimate $(\widehat{u}_{i,1}, \dots, \widehat{u}_{i,N}) \in \mathcal{C}_i$. It can be utilized in the subsequent steps of the successive cancellation decoding algorithm.

The similarity of the above described algorithm with multistage decoding suggests employing multilevel code design rules for selection of parameters of the proposed coding scheme. The following rules are of most interest [8]:

1) *Capacity rule.* The rate R_i of $\mathcal{C}_i$ should be chosen equal to the capacity C_i of the i-th subchannel. According to [8] and (3)–(4), one obtains

$$\begin{aligned} C_i = I(y_1^n; u_i|u_1^{i-1}) &= \mathbf{E}_{u_1^{i-1}}[C(A(u_1^{i-1}))] \\ &\quad - \mathbf{E}_{u_1^i}[C(A(u_1^i))], \quad (5) \end{aligned}$$

where

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

is the capacity when using the subset B of $\{0,1\}^n$ for transmission over the vector channel $W^n(y_1^n|x_1^n)$. Obviously, employing this rule results in a capacity-achieving concatenated code, provided that the outer codes can achieve the capacity too. However, evaluating (5) seems to be a difficult task. Furthermore, besides the polar codes themselves there are still no capacity achieving code constructions (except for the case of binary erasure channel).

2) *Coding exponent rule.* The rate R_i of $\mathcal{C}_i$ should be selected as a solution of the equation

$$\mathcal{E}_i(R_i) = -\frac{1}{N} \log_2 p_e,$$

where p_e is the target error probability, and $\mathcal{E}_i(R_i) = \max_{0 \le \rho \le 1} (E_i(\rho) - \rho R_i)$. This rule ensures that the decoding error probabilities of $\mathcal{C}_i$ are approximately the same, provided that random codes are used. However, employing it requires evaluation of $E_i(\rho) = \mathbf{E}_{u_1^{i-1}} \left[-\log_2 E_i(\rho, u_1^{i-1}) \right]$, where

$$E_i(\rho, u_1^{i-1}) = \int_{y_1^n} \left(\sum_{u_i=0}^{1} \frac{1}{2} f^{\frac{1}{1+\rho}}(y_1^n|u_1^i) \right)^{1+\rho} dy_1^n, \quad (6)$$

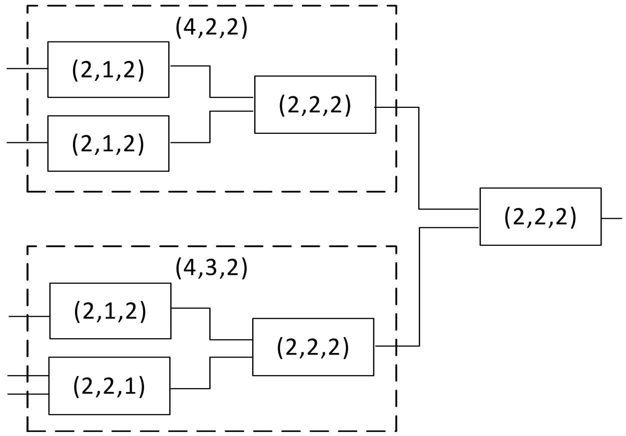

Fig. 1. $(8,5,2)$ polar code as a generalized concatenated code

and

$$f(y_1^n|u_1^i) = \sum_{u_{i+1}^n} \frac{W^n(y_1^n|u_1^n G)}{2^{n-i}}.$$

Evaluating (6) requires one essentially to enumerate all possible vectors u_1^n, and is therefore also impractical.

3) *Equal error probability rule*. More reasonable approach can be based on considering a specific family of outer codes, and selecting their parameters so that the decoding error probability is approximately the same for all subchannels. Some subchannels may be assigned zero-rate codes under this scheme. These subchannels correspond to frozen bits.
4) *Balanced minimum distances rule*. The classical approach to the design of generalized concatenated codes is to select $D_i d_i \approx const$. However, as it was shown in [8], 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.

The described construction can be considered as an instance of generalized concatenated codes. Indeed, the i-th inner code is given by $\mathbb{C}_i = \mathbb{C}_{i+1} \cup (e_i G_n + \mathbb{C}_{i+1})$, where e_i is the i-th unit vector, and $\mathbb{C}_{n+1} = \{(0,\dots,0)\}$. The outer codes are given by $\mathcal{C}_i \subset GF(2)^N$. Since the construction of polar codes is based on recursive application of the polarizing transformation, they can be also considered as an instance of generalized concatenated codes. The ability of polar codes to achieve the capacity implies the existence of capacity achieving generealized concatenated codes. Furthermore, it turns out that the decoding algorithm derived in [10] for the case of Reed-Muller codes (which can be considered as an instance of GCC designed according to the balanced distances rule) can be obtained from (3)–(4) by employing the min-sum approximation.

Example 1. *Consider a $(8,5,2)$ polar code with generator matrix [11]*

$$G = \begin{pmatrix} 1&1&0&0&0&0&0&0\\ 1&1&1&1&0&0&0&0\\ 1&1&0&0&1&1&0&0\\ 1&0&1&0&1&0&1&0\\ 1&1&1&1&1&1&1&1 \end{pmatrix}.$$

It can be obtained as a 3-level generalized concatenation of $(2,1,2)$ and $(2,2,1)$ codes, as shown in Figure 1. The last level of concatenation is based on partitioning of the set of 2-vectors into subsets $A_0 = \{00,11\}$ and $A_1 = \{10,01\}$. The codeword symbols of $(4,2,2)$ code are used to select a particular subset, while the ones of $(4,3,2)$ code select particular tuples to be transmitted.

It was shown in [2] that channel polarization can be also performed by high-dimensional kernels F based on nested BCH codes. This construction can be also treated in the framework of generalized concatenated codes.

Concatenated coding schemes similar to the one described above were proposed in [12], [13]. However, these papers do not address the problem of outer code rate optimization in a systematic way.

B. Evaluating the quality of equivalent subchannels

Implementing the equal error probability rule requires one to be able to calculate the quality of equivalent subchannels. This can be performed by constructing the PDF of $L_1^{(i)}(y_i)$ and applying to it density evolution. The most practically important case corresponds to the AWGN channel. In this scenario $L_1^{(i)}(y_i) \sim \mathcal{N}(\frac{2}{\sigma^2}, \frac{4}{\sigma^2})$. It was suggested in [14] to approximate the PDFs of the values given by (3)–(4) with suitable Gaussian distributions. 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

$$\mathbf{E}[L_n^{(2i-1)}] = \phi^{-1}\left(1 - \left(1 - \phi\left(\mathbf{E}[L_{n/2}^{(i)}]\right)\right)^2\right) \quad (7)$$

$$\mathbf{E}[L_n^{(2i)}] = 2\,\mathbf{E}[L_{n/2}^{(i)}], \quad (8)$$

where $\phi(x) \approx e^{-0.4527x^{0.86}+0.0218}$. Channel symmetry implies that $\mathbf{D}[L_m^{(i)}] = 2\,\mathbf{E}[L_m^{(i)}]$. The error probability for each subchannel can be obtained as

$$\pi_i \approx Q\left(\sqrt{\mathbf{E}[L_n^{(i)}]/2}\right). \quad (9)$$

Evaluating these expressions in many cases leads to more accurate results compared to the standard density evolution, since the latter approach may suffer from quantization errors.

C. Rate allocation algorithm

It is well known that 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

$$w(c)=\sum_{i:c_i\neq 0} L_i, \tag{10}$$

and $L_i = \ln\frac{P\{c_i=0|y_i\}}{P\{c_i=1|y_i\}}$. While the log-likelihood ratios obtained from (3) are not Gaussian, for sufficiently good codes the number of summands in (10) is large, and this quantity can be also approximated as a Gaussian random variable by the central limit theorem. Hence, one obtains

$$p_e \leq \sum_{j=1}^{N} A_j Q\left(\sqrt{\frac{\mathbf{E}[L_i]}{2}j}\right) \approx A_d Q\left(\sqrt{\frac{\mathbf{E}[L_i]}{2}d}\right), \tag{11}$$

where A_i are weight spectrum coefficients of code $\mathcal{C}$, and d is its minimum distance. While it is in general difficult to obtain code weight spectrum, 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 A_d and d.

Let K_l and $P_l(m)$ be the dimension and decoding error probability function for the l-th code in the considered family of possible outer codes, respectively, where m is the expected value of LLR, and $K_0 = P_0(m) = 0$. Let us further assume that $P_i(m) < P_j(m) \Leftrightarrow K_i < K_j$. The following simple algorithm can be used to construct a generalized concatenated code with rate R according to the equal error probability rule. The algorithm employs the bisection method to approximately solve the equation $\sum_{i=1}^{n} K(i,P) = NnR$, 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. It is assumed that the code will be used for transmitting data over 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:

1) Let $\mathbf{E}[L_1^{(1)}] = 2/\sigma^2$. Compute $m_i = \mathbf{E}[L_n^{(i)}], 1 \leq i \leq n$ using (7)–(8).
2) Let $P' = 1, P'' = 0$.
3) If $P' - P'' < \epsilon P'$ go to step 7:
4) Let $\tilde{P} = (P' + P'')/2$.
5) For each i find $l_i = \arg\max_{l:P_l(m_i)\leq \tilde{P}} K_l$. Let $K = \sum_{i=1}^{n} K_{l_i}$.
6) If $K < RNn$, then $P'' = \tilde{P}$, else $P' = \tilde{P}$. Go to step 3.
7) Return $(K_{l_1}, \ldots, K_{l_n}), \tilde{P}$.

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

$$\begin{aligned} P &= 1 - P\{C_1, \ldots, C_n\} \\ &= 1 - P\{C_1\} P\{C_2|C_1\} \cdots P\{C_n|C_1, \ldots, C_{n-1}\} \\ &\approx 1 - \prod_{i=1}^{n}(1 - P_{l_i}(m_i)) \approx 1 - (1-\tilde{P})^n, \end{aligned} \tag{12}$$

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 l_i is the index of the code selected for the i-th subchannel.

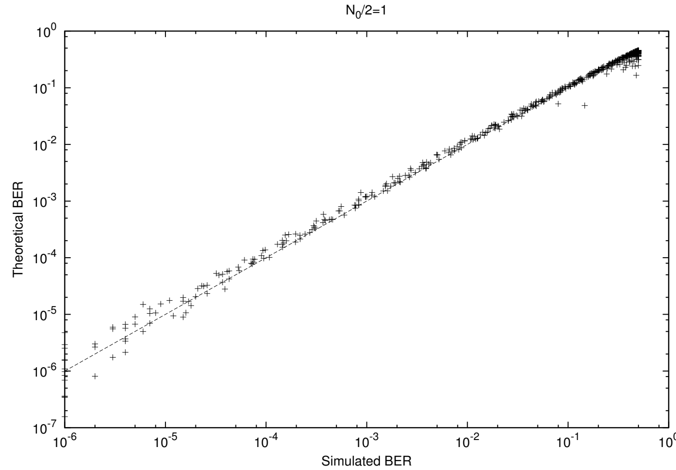

Fig. 2. Accuracy of Gaussian approximation

TABLE I
DESIGN OF $(1016, 508)$ CODE, $E_b/N_0 = 3$ dB

i	1	5	3	7	2	6	4	8
$\widehat{\pi}_i$	0.48	0.25	0.3	0.044	0.33	0.065	0.097	0.0024
π_i	0.44	0.24	0.29	0.044	0.32	0.065	0.1	0.0023
K_i	0	42	29	105	22	99	91	120

IV. NUMERIC RESULTS

Figure 2 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 successive cancallation 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. Observe that there are many subchannels with medium bit error rate, which require additional layer of coding to achieve reliable data transmission.

Next we present performance results for the case of GCC based on outer BCH codes decoded with box-and-match algorithm [15] with reprocessing order 4 and inner polar codes.

Table I illustrates the design of a rate 0.5 concatenated code based on BCH codes of length 127 and polar codes of length 8. Here $\widehat{\pi}_i$ denotes the empiric bit error probability for the i-th equivalent subchannel, provided that all previous decisions of the successive cancellation decoder are correct. One can see that these values are very close to the theoretical error probability π_i given by (9). Observe also that most of the sub-

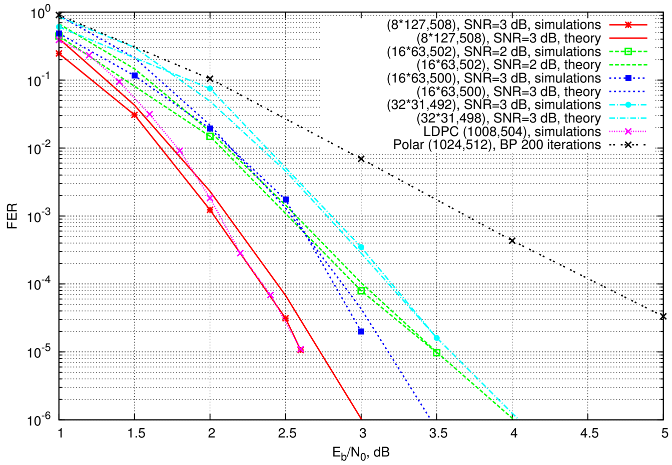

Fig. 3. Performance of concatenated codes

channels have quite high error probability. In a real successive cancellation decoder this results in severe error propagation, which causes the existing short polar codes to perform much worse compared to other ones. The error propagation problem can be mitigated by employing appropriate outer codes. The dimensions K_i of outer codes shown in the table were obtained with the optimization algorithm presented in Section III-C.

Figure 3 presents simulation results for codes of length nN obtained from various polar and BCH codes of length n and N, respectively, as well as their theoretical decoding error probability. It can be seen that the code described in Table I provides the same performance as a similar irregular LDPC code in the high-SNR region, and outperforms it for low SNR. This behaviour is due to near-ML decoding of outer component codes. Figure 3 shows also that increasing the length of outer codes results in better performance of concatenated codes. This is due to higher flexibility in selection of code parameters provided by longer codes. It appears also that running the rate assignment algorithm for different values of σ^2 results in different codes. Those obtained for smaller values of σ^2 (compare the codes for $SNR = 3$ dB and $SNR = 2$ dB) have usually higher minimum distance and perform better at high SNR. It can be also seen that the actual code performance is quite close to the one predicted by (12). The performance of the obtained concatenated codes is much better than the one of plain polar code decoded with belief propagation algorithm [16].

V. Conclusions

In this paper it was shown that the channel polarization transformation can be treated in the framework of multilevel coding and multistage decoding. This enable one to utilize the corresponding code design rules to obtain good generalized concatenated codes. By properly selecting the rates of the component codes one can avoid error propagation, which can substantially degrade the performance of short plain polar codes. It was also shown that the quality of the equivalent subchannels induced by the polarizing transformation can be efficiently studied using Gaussian approximation for density evolution. The codes obtained with this approach substantially outperform plain polar codes, and approach the performance of similar irregular LDPC codes. However, this comes in general at the expence of higher decoding complexity of component codes. The proposed approach enables one also to accurately predict the performance of the constructed codes, and avoid costly simulations, which may be infeasible at high SNR. However, this requires one to be able to precisely estimate the performance of component codes.

The performance can be further improved by more careful selection of outer codes, as well as by employing an iterative decoding algorithm for the concatenated code.

Acknowledgements

This work was partially supported by the grant MK-1976.2011.9 of the President of Russia.

References

[1] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions On Information Theory*, vol. 55, no. 7, July 2009.
[2] S. B. Korada, E. Sasoglu, and R. Urbanke, "Polar codes: Characterization of exponent, bounds, and constructions," *IEEE Transactions On Information 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 *Proceedings of IEEE Information Theory Workshop*, 2010.
[4] E. Blokh and V. Zyablov, "Coding of generalized concatenated codes," *Problems of Information Transmission*, vol. 10, no. 3, pp. 45–50, 1974.
[5] V. Zinov'ev, "Generalized cascade codes," *Problems of Information Transmission*, vol. 12, no. 1, pp. 5–15, 1976.
[6] M. Bossert, *Channel coding for telecommunications*. Wiley, 1999.
[7] H. Imai and S. Hirakawa, "A new multilevel coding method using error correcting codes," *IEEE Transactions on Information Theory*, vol. 23, no. 3, pp. 371–377, May 1977.
[8] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Transactions On Information Theory*, vol. 45, no. 5, pp. 1361–1391, July 1999.
[9] R. Mori and T. Tanaka, "Performance of polar codes with the construction using density evolution," *IEEE Communications Letters*, vol. 13, no. 7, July 2009.
[10] G. Schnabl and M. Bossert, "Soft-decision decoding of Reed-Muller codes as generalized multiple concatenated codes," *IEEE Transactions on Information Theory*, vol. 41, no. 1, January 1995.
[11] E. Arikan, "A performance comparison of polar codes and Reed-Muller codes," *IEEE Communications Letters*, vol. 12, no. 6, June 2008.
[12] E. Arikan and G. Markarian, "Two-dimensional polar coding," in *Proceedings of 10'th International Symposium on Communication Theory and Applications*, Ambleside, UK, 2009.
[13] M. Seidl and J. B. Huber, "Improving successive cancellation decoding of polar codes by usage of inner block codes," in *Proceedings of 6th International Symposium on Turbo Codes and Iterative Information Processing*, 2010, pp. 103 – 106.
[14] 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 Transactions on Information Theory*, vol. 47, no. 2, Feb. 2001.
[15] A. Valembois and M. Fossorier, "Box and match techniques applied to soft-decision decoding," *IEEE Transactions on Information Theory*, vol. 50, no. 5, May 2004.
[16] E. Arikan, "Polar code encoder and belief propagation decoder in Matlab," 2010, software Item. [Online]. Available: http://www.vicewicom.eu/General-Area/Contents/Software-Simulation-and-data-Library-SSL/Polar-code-encoder-and-belief-propagation-decoder-in-Matlab