An Approximate Method for Construction of Polar codes with Kernels over $\mathbb{F}_{2^t}$

Liudmila Karakchieva, Peter Trifonov *Member, IEEE*

***Abstract*—An approximate method for evaluation of the reliability of the symbol subchannels induced by a polarizing transformation with non-binary kernels is proposed. We show that if polar codes are combined with the channel adapter, the capacities of the subchannels can be estimated recursively.**

***Index Terms*—Polar codes, non-binary kernels, channel adapter.**

I. Introduction

Polar codes are linear block codes, which achieve symmetric capacity of memoryless channels and have low encoding and decoding complexity [1]. Classical polar codes are based on the 2×2 matrix $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ and can be constructed by several methods, in particular, by binary erasure channel recursion [1], degrading and upgrading transformations [2], Gaussian approximation [3], as well as the Monte-Carlo method. However, the rate of polarization of the Arikan kernel is quite low, which results in poor finite-length performance of the corresponding polar codes.

Much higher rate of polarization can be obtained by employing large kernels [4], [5]. Rate of polarization provided by non-binary kernels (e.g. Reed-Solomon) far exceeds that of binary kernels with similar dimension. Reed-Solomon kernel was also shown to provide optimal scaling exponent [5].

To the best of our knowledge, only the erasure channel recursion [6] and Monte-Carlo code construction methods [1] are currently available for the case of large non-binary kernels. However, the erasure channel recursion is quite inaccurate for generic channels, while the Monte-Carlo method is very time consuming.

A generalization of the Gaussian approximation method to the case of binary large kernels was suggested in [7]. It relies on construction of histograms of the LLRs produced by a kernel processor to obtain a family of curves, which relate the capacity of the underlying channel with the capacities of the subchannels induced by a kernel. Construction of such histograms is almost infeasible for the case of non-binary kernels. In this paper we present an alternative approach to construction of such curves, extending therefore the Gaussian approximation method to the case of large non-binary kernels.

II. Background

Let us consider an $l \times l$ non-singular matrix K over $\mathbb{F}_q$, $q = p^t > 2$, such $\mathbb{F}_p(\overline{K}) = \mathbb{F}_q$ for any standard form $\overline{K}$ of K, where $\mathbb{F}_p(K)$ is the field extension of $\mathbb{F}_p$ generated by the adjunction of all elements of K. It is possible to show that such matrix polarizes any q-ary memoryless source and q-ary input memoryless output-symmetric channel [5].

Reed-Solomon kernel is given by matrix K with elements $K_{ij} = \alpha_j^{l-i-1}$, where $0 \le i, j < l$, α_j are some distinct elements of $\mathbb{F}_q$, and $l \le q$ is kernel dimension. Hence, the last k rows of K represent a generator matrix of an (l, k) RS code.

Rate of polarization of the RS kernel is given by $E(l) = \log(l!)/(l\log(l))$, which is the highest possible value for a given $l \le q$. For example, for $l = 4$ and $l = 8$ the Reed-Solomon kernels provide rate of polarization $E(4) \approx 0.57312$ and $E(8) \approx 0.63747$ respectively, which is much higher compared to 0.5, rate of polarization of the Arikan kernel. Recall that binary kernels achieve $E(l) > 0.5$ only for $l \ge 16$.

An $(n = l^m, k)$ non-binary polar code is a code generated by k rows of matrix $G_m = B_{l,m}K^{\otimes m}$, where $u_i = 0$, $i \in \mathcal{F}$, $\mathcal{F} \subset \{0, \ldots, n-1\}$ is a set of indices of frozen symbols, $|\mathcal{F}| = n-k$, and $B_{l,m}$ is the digit-reversal permutation matrix, which provides the mapping

$$\pi\Big(\sum_{i=0}^{m-1} j_i l^i\Big) = \sum_{i=0}^{m-1} j_i l^{m-1-i}, 0 \le j_i < l.$$

Let us consider a q-ary input memoryless output-symmetric channel with transition probability function $W_0^{(0)}(y|c) = W(y|c)$, $c \in \mathbb{F}_q$. We consider here the case of $q = 2^t$.

The subchannels induced by the polarizing transformation G_m with the kernel K are given by

$$W_m^{(i)}(y_0^{n-1}, u_0^{i-1}|u_i) = \frac{1}{q^{n-1}} \sum_{u_{i+1}^{n-1} \in \mathbb{F}_q^{n-i-1}} \prod_{j=0}^{n-1} W_0^{(0)}(y_j|(u_0^{n-1}G_m)_j), 0 \le i < n.$$

It is convenient to define probabilities

$$W_\lambda^{(lj+i)}(u_0^{lj+i}|y_0^{n'-1}) = \sum_{u_{lj+i+1}^{lj+l-1}} \prod_{s=0}^{l-1} W_{\lambda-1}^{(j)}\big((u_{lt}^{lt+l-1}K)_s, 0 \le t \le j|y_{\frac{n'}{l}s}^{\frac{n'}{l}s+\frac{n'}{l}-1}\big), \quad (1)$$

where $n' = l^\lambda$. These probabilities can be computed as described in [8].

The successive cancellation algorithm can be used to decode polar codes. This method makes decisions

$$\hat{u}_i = \begin{cases} \arg\max_{u_i \in \mathbb{F}_q} W_m^{(i)}(\hat{u}_0^{i-1}, u_i|y_0^{n-1}), & i \notin \mathcal{F} \\ \text{the frozen value of } u_i & i \in \mathcal{F} \end{cases}.$$

The authors are with ITMO University, Russia. Email: {lvkarakchieva, pvtrifonov}@itmo.ru

This work was supported by the Ministry of Science and Higher Education of Russian Federation, project (Goszadanie) no. 2019-0898.

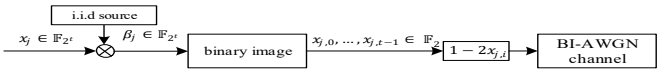


Fig. 1: BI-AWGN channel with binary image mapper and channel adapter

III. The capacity of subchannels

The reliability of subchannels $W_m^{(i)}(Y_0^{n-1}, U_0^{i-1}|U_i)$ can be characterized by their mutual information $I_m^{(i)}(Y_0^{n-1}, U_0^{i-1}; U_i)$, where Y_0^{n-1} is a random vector corresponding to channel output, and U_0^i is a random vector corresponding to information symbols. From the definition of mutual information, we obtain

$$\begin{aligned} I_m^{(i)}(Y_0^{n-1}, U_0^{i-1}; U_i) &= H(U_i) - H(U_i|Y_0^{n-1}, U_0^{i-1}) \\ &= -\sum_{a\in\mathbb{F}_q} p(U_i = a)\log_2\big(p(U_i = a)\big) \\ &+ \sum_{U_0^i\in\mathbb{F}_q^{i+1}} \int_{Y_0^{n-1}\in\mathbb{R}^n} W_m^{(i)}(Y_0^{n-1}, U_0^i) \\ &\times \log_2\big(W_m^{(i)}(U_i|Y_0^{n-1}, U_0^{i-1})\big)dY. \end{aligned} \tag{2}$$

The complexity of computing this expression grows exponentially with i. For the case of the Arikan kernel, these computations can be approximated with polynomial complexity by employing channel degrading and upgrading transformations [2]. This approach was generalized in [9] to the case of non-binary codes with 2×2 Arikan-like kernel. However, to the best of our knowledge, no techniques for computing $I_m^{(i)}(Y_0^{n-1}, U_0^{i-1}; U_i)$ are available for larger non-binary kernels.

We consider transmission of binary images of codeword symbols over BI-AWGN channel. That is, given $x_j = \sum_{i=0}^{t-1} x_{j,i}\alpha^i$, where α is a primitive element of $\mathbb{F}_{2^t}$, symbols $x_{j,0}, ..., x_{j,t-1} \in \mathbb{F}_2$ are transmitted. However, this results in a non-uniform distribution of errors in the symbols of the non-binary codeword, while the actual errors in the subchannels in $W_m^{(i)}$ are approximately equiprobable. We propose to overcome this problem by employing a channel adapter, similarly to [10]. Fig. 1 illustrates the proposed transmission scheme. The channel adapter multiplies codeword symbols by independent random values β_j uniformly distributed over $\mathbb{F}_{2^t} \setminus \{0\}$, which are known to the receiver.

To obtain a simple method for construction of non-binary polar codes, we propose to assume that the subchannels $W_m^{(i)}$ behave in a way similar to a binary-input AWGN channel combined with a binary image transmitter and appropriate channel adapter. That is, we assume that the symbols x_j to be transmitted over these subchannels are multiplied by random non-zero values $\beta_j \in \mathbb{F}_{2^t}$, known to the receiver[1], the value $\beta_j x_j$ is transformed into its binary image, and the obtained t bits are transmitted over BI-AWGN channel.

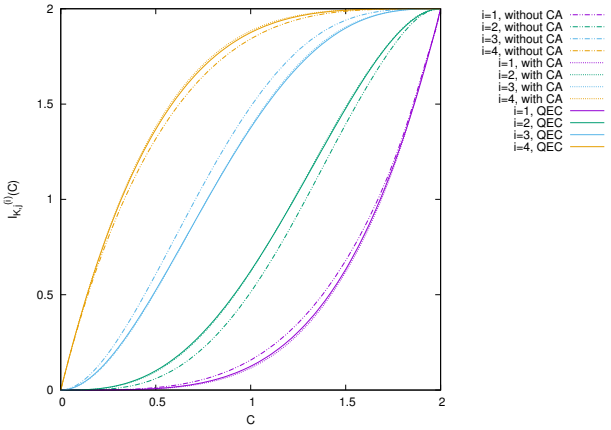
This assumption implies that the capacity of subchannel $W_m^{(i)}$ induced by m-layered polarizing transformation and some q-ary input memoryless output symmetric channel W depends only on m, i, the kernel being used, and the capacity $C = I(W)$ of W. This enables one to approximately evaluate the capacities of the subchannels as [7]

$$I_m^{(li+j)}(C) \approx I_{K,j}(I_{m-1}^{(i)}(C)), 0 \le j < l, \tag{3}$$

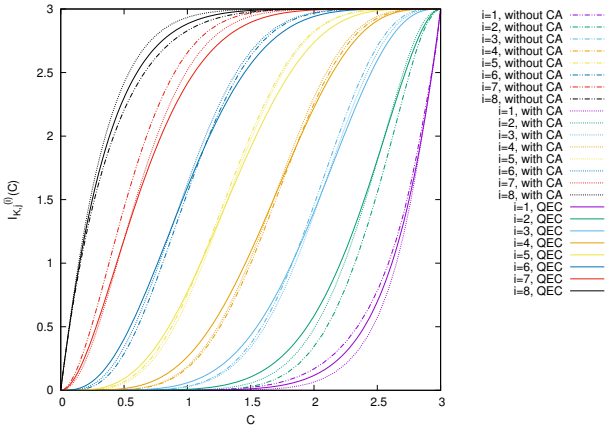
where $I_0^{(0)} = C$, and $I_{K,j}(C)$ are a family of kernel-specific functions, which give the capacity of $W_1^{(j)}$, provided that the underlying channel has capacity C.

Hence, the proposed code construction method involves the following steps:

1) (one-time preprocessing): Obtain functions $I_{K,j}(C), 0 \le j < l$ for a given kernel K and various values of C.
2) Compute the capacity $C = I(W)$ of the underlying channel W.
3) Compute recursively the capacities of the subchannels via (3).
4) Let $\mathcal{F}$ be given by $n - k$ indices i of subchannels with the smallest values of $I_m^{(i)}$.



(a) 4×4 kernel over $\mathbb{F}_4$



(b) 8×8 kernel over $\mathbb{F}_8$

Fig. 2: Subchannel capacity functions

IV. Computing capacity functions

It was suggested in [7] for the case of $q = 2$ to obtain functions $I_{K,j}(C)$ from the histograms of output values of a kernel processor, i.e. the device, which computes

[1]In general, one should use a random bijective mapping $\mathbb{F}_q \to \mathbb{F}_q$ as a channel adapter. However, we have found that multiplication by random non-zero values already provides quite good results.

$W_\lambda^{(lj+i)}(u_0^{lj+i}|y_0^{n'-1})$, their approximations or the corresponding LLRs. While it may be possible to extend this approach to the non-binary case, as described in [11], [12], [13], the number of entries in such histograms, i.e. the complexity of their construction, grows exponentially with q. Hence, this approach is not feasible in practice for $q > 3$.

It was shown in [14] that one can avoid construction of multi-dimensional histograms by performing a Monte-Carlo simulation of an algorithm for evaluation of the associated symbol probabilities $(W_m^{(i)}(Y_0^{n-1}, U_0^{i-1}|U_i)$ in the considered case), computing their entropies, and averaging them over sufficiently many channel output instances, i.e.

$$I_{K,j}(C) \approx \frac{1}{T}\sum_{s=1}^{T} \hat{I}_{K,j}(Y_{0,s}^{l-1}, U_{0,s}^{i-1}; U_i), \tag{4}$$

where the subscript s denotes the s-th realization of the corresponding random variable, T is the number of samples, and $\hat{I}_{K,j}(Y_{0,s}^{l-1}, U_{0,s}^{i-1}; U_i)$ denotes an estimate of the capacity of the corresponding subchannel, obtained from the values $W_1^{(j)}(U_i = u_i|Y_{0,s}^{l-1}, U_{0,s}^{i-1}), u_i \in \mathbb{F}_{2^t}$, which is given by

$$\hat{I}_{K,j}(Y_{0,s}^{l-1}, U_{0,s}^{i-1}; U_i) = t + \sum_{a\in\mathbb{F}_{2^t}} w(a)\log_2\big(w(a)\big),$$

where $w(a) = W_1^{(i)}(U_i = a|Y_{0,s}^{l-1}, U_{0,s}^{i-1})$, and U_i is a random variable corresponding to the i-th kernel input symbol. In other words, $\hat{I}_{K,j}(Y_{0,s}^{l-1}, U_{0,s}^{i-1}; U_i)$ can be computed by (2) for $m = 1$ given $Y_{0,s}^{l-1}$, the s-th realization of Y_0^{l-1}. It was shown in [14] that for sufficiently large T such estimates converge to the true values of the corresponding mutual information functions.

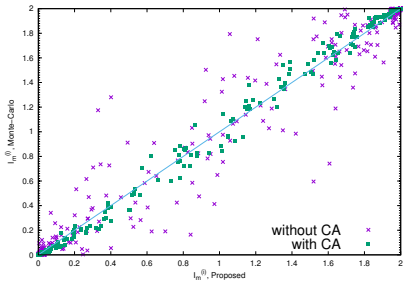
The complexity of construction of the tables, which approximate functions $I_{K,j}(C)$, is $O(TlC_p)$, where C_p is the complexity of single evaluation of (1). The complexity of code construction with the proposed method is $O(n)$ evaluations of $I_{K,j}(C)$, i.e. interpolation over the pre-computed tables. Observe that the complexity of the Monte-Carlo code construction method is $O(Tn\log_q(n)C_p)$.

The proposed code construction method reduces to one-time construction of tables, which is implemented by applying the SC decoder to a code consisting of a single instance of the polarization kernel, while the classical Monte-Carlo method requires one to run the SC decoder for a multi-layered polarizing transformation for every value of channel SNR and code length.

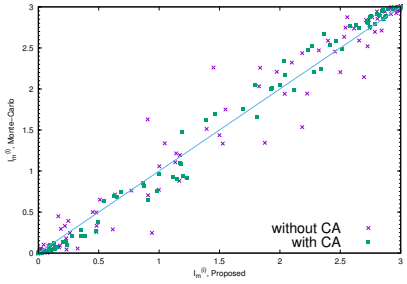
V. Numeric results

In our simulations we consider transmission of the binary image of the codewords of polar codes with Reed-Solomon kernels over BI-AWGN channel. Both the cases of channel adapter with random β_j (i.e. with channel adapter (CA), see Fig. 1), and $\beta_j = 1$ (i.e. without channel adapter), were considered.

Fig. 2 illustrates the subchannel capacity functions $I_{K,i}(C)$ for Reed-Solomon kernels over $\mathbb{F}_4$ and over $\mathbb{F}_8$ for AWGN channel (dashed lines) and the q-ary erasure channel (QEC) (solid lines). It can be seen that in the case of $\mathbb{F}_4$ the functions obtained for AWGN channel with channel adapter are very close to those obtained for QEC. This means that the capacity function for QEC can be reliably used to obtain codes for AWGN channel. However, this is clearly not the case for codes over $\mathbb{F}_8$. Observe also, that the curves obtained for AWGN channel with and without channel adapter are quite different. This is due to highly non-uniform distribution of errors at the output of the BI-AWGN channel without the channel adapter.



(a) $m = 5$, 4×4 RS kernel over $\mathbb{F}_4$



(b) $m = 3$, 8×8 RS kernel over $\mathbb{F}_8$

Fig. 3: Accuracy of the proposed method of subchannel reliability estimation

The kernel capacity functions shown in Fig. 2 were used to estimate the capacities of bit subchannels of the polarizing transformations via the recursive function (3). Fig. 3 presents the Monte-Carlo simulated $I_m^{(i)}$ for each subchannel vs the estimated values of $I_m^{(i)}$ for the case of transmission of the binary image of the output of the polarizing transformation over the AWGN channel at $E_s/N_0 = -1$ dB with genie-aided SC decoder. It can be seen that the kernel capacity functions obtained with CA enable one to order the subchannels $W_m^{(i)}$ more accurately according to their reliability, compared to the case of the scheme without CA That is, the Monte-Carlo simulated $I_m^{(i)}$ approximately monotonically increase with $I_m^{(i)}$ in the case of the scheme with CA. This means that the proposed approach can be used to order bit subchannels according to their reliability. On the other hand, the results obtained for the scheme without CA are quite chaotic.

Fig. 4–5 present the performance of polar codes with 8×8 and 4×4 RS kernels constructed for $E_s/N_0 = -1$ dB by Monte-Carlo simulations and by the proposed recursive approximation (3). It can be seen that employing the channel adapter enables one to obtain better performance. The proposed approximation method together with the channel adapter enables one to obtain codes with almost the same performance as in the case of subchannel reliabilities obtained via Monte-Carlo simulations.

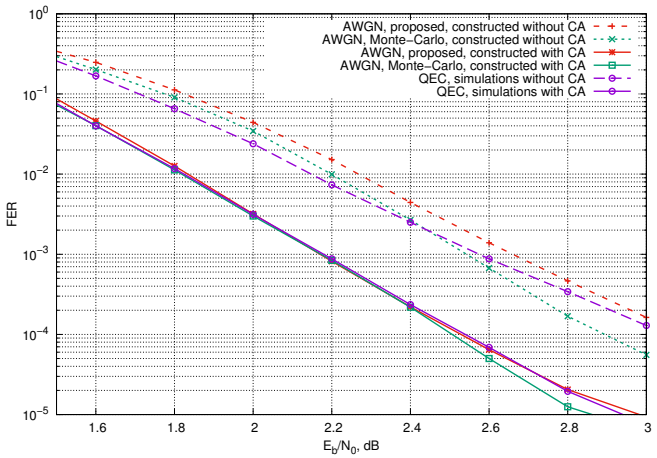


Fig. 4: Performance of $(4^5, 512)$ polar codes

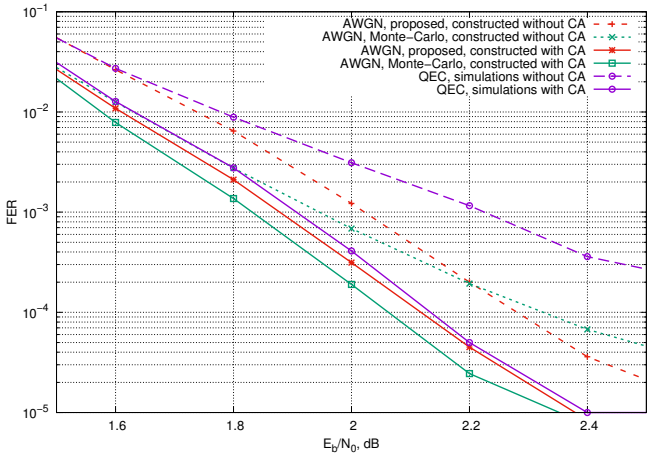


Fig. 5: Performance of $(8^3, 256)$ polar codes

Furthermore, Fig. 4–5 illustrate the performance of codes constructed for QEC with the same capacity as in the above case. Simulations were done with channel adapter, unless stated otherwise. It can be seen that channel adapter is essential for obtaining good performance. In the case of $\mathbb{F}_4$, the code obtained for QEC provides approximately the same performance as the one obtained via the proposed approximation method. This is due to capacity functions shown in Fig. 2a for the case of QEC and BI-AWGN channel with channel adapter being approximately equal. However, in the case of $\mathbb{F}_8$ there is a non-negligible performance gap between the codes obtained for the QEC and BI-AWGN channel with CA. The code obtained with the proposed method provides better performance compared to the one constructed for QEC.

Table I shows the number of frozen positions in which the codes constructed for QEC recursion and by the proposed method differ from codes constructed by Monte-Carlo method. It can be seen that introducing of channel adapter makes the constructed code to be closer to the code constructed by Monte-Carlo method. Furthermore, in the case of $\mathbb{F}_4$ the number of inconsistencies in frozen sets for both methods

TABLE I: Comparison of codes constructed by Monte-Carlo method, QEC recursion and proposed approximate method

Constructed code	RS kernel	Channel adapter	Proposed	QEC
$(4^5, 512)$	$K_{4\times4}$	yes	2	2
$(4^5, 512)$	$K_{4\times4}$	no	20	38
$(8^3, 256)$	$K_{8\times8}$	yes	4	6
$(8^3, 256)$	$K_{8\times8}$	no	8	10

corresponding to the last two columns is the same. Increasing the kernel size results in more inconsistencies between the codes. This explains the performance gap between the codes obtained for the QEC and BI-AWGN channel (see Fig. 5).

VI. Conclusions

A novel method for construction of polar codes with non-binary kernels was presented.

It relies on a family of functions for computing capacities of the subchannels induced by a single instance of the kernel, which are recursively used for computing the capacities of subchannels induced by the polarizing transformation based on such kernel. Furthermore, it was shown that employing the channel adapter enables one to obtain better performance compared to the case of straightforward transmission of a binary image of polar code.

References

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

[2] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6562–6582, October 2013.

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

[4] 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, pp. 6253–6264, December 2010.

[5] R. Mori and T. Tanaka, "Source and channel polarization over finite fields and Reed-Solomon matrices," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2720–2736, May 2014.

[6] H. Pfister and R. Urbanke, "Near-optimal finite-length scaling for polar codes over large alphabets," in *Proceedings of IEEE International Symposium on Information Theory*, 2016.

[7] P. Trifonov, "On construction of polar subcodes with large kernels," in *Proceedings of IEEE International Symposium on Information Theory*, Paris, France, 2019.

[8] H. Griesser and V. R. Sidorenko, "A posteriory probability decoding of nonsystematically encoded block codes," *Problems of Information Transmission*, vol. 38, no. 3, 2002.

[9] T. C. Gulcu, M. Ye, and A. Barg, "Construction of polar codes for arbitrary discrete memoryless channels," *IEEE Transactions On Information Theory*, vol. 64, no. 1, January 2018.

[10] G. Li, I. Fair, and W. A. Krzymien, "Density evolution for nonbinary LDPC codes under gaussian approximation," *IEEE Transactions on Information Theory*, vol. 55, no. 3, March 2009.

[11] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Transactions On Communications*, vol. 49, no. 10, October 2001.

[12] A. J. Grant, "Convergence of non-binary iterative decoding," in *Proceedings of IEEE Global Telecommunications Conference*, 2001.

[13] H. Chen and A. Haimovich, "Exit charts for turbo trellis-coded modulation," *IEEE Communications Letters*, vol. 8, no. 11, November 2004.

[14] J. Kliewer, B. C. Ng, and L. Hanzo, "Efficient computation of exit functions for nonbinary iterative decoding," *IEEE Transactions On Communications*, vol. 54, no. 12, December 2006.