

From Polar to Reed-Muller Codes: a Technique to Improve the Finite-Length Performance

Marco Mondelli, S. Hamed Hassani, and Rüdiger Urbanke

Abstract—We explore the relationship between polar and RM codes and we describe a coding scheme which improves upon the performance of the standard polar code at practical block lengths. Our starting point is the experimental observation that RM codes have a smaller error probability than polar codes under MAP decoding. This motivates us to introduce a family of codes that “interpolates” between RM and polar codes, call this family $\mathcal{C}_{\text{inter}} = \{C_\alpha : \alpha \in [0, 1]\}$, where $C_\alpha|_{\alpha=1}$ is the original polar code, and $C_\alpha|_{\alpha=0}$ is an RM code. Based on numerical observations, we remark that the error probability under MAP decoding is an increasing function of α . MAP decoding has in general exponential complexity, but empirically the performance of polar codes at finite block lengths is boosted by moving along the family $\mathcal{C}_{\text{inter}}$ even under low-complexity decoding schemes such as, for instance, belief propagation or successive cancellation list decoder. We demonstrate the performance gain via numerical simulations for transmission over the erasure channel as well as the Gaussian channel.

Keywords—Polar codes, RM codes, MAP decoding, SC decoding, list decoding.

I. INTRODUCTION

Polar Coding: Benefits and Drawbacks. Polar codes, which were introduced by Arıkan in [1], are a family of codes which provably achieve the capacity of a large class of channels, including binary-input memoryless output-symmetric channels (BMSCs), by means of encoding and decoding algorithms with complexity $\Theta(N \log N)$, N being the block length of the code.

In particular, for any BMSC W with capacity C and for any rate $R < C$, the block error probability under the proposed successive cancellation (SC) decoding, namely P_e^{SC} , scales roughly as $2^{-\sqrt{N}}$ as N grows large [2]. This result has been further refined and extended to the MAP decoder, showing that both $\log_2(-\log_2 P_e^{\text{SC}})$ and $\log_2(-\log_2 P_e^{\text{MAP}})$ behave as $\log_2(N)/2 + \sqrt{\log_2(N)/2} \cdot Q^{-1}(R/C) + o(\sqrt{\log_2(N)})$ for any fixed rate strictly less than capacity [3]. Consequently, even at moderate block lengths, error floors do not affect the performance of polar codes.

However, when we consider rates close to capacity, simulation results show that large block lengths are required in order to achieve a desired error probability. Therefore, it is interesting to explore the trade-off between the gap to capacity $C - R$ and the block length N when the error probability is a fixed value P_e . In particular, it has been observed that $C - R$ scales as $N^{-1/\mu}$, where μ denotes the scaling exponent

[4]. For transmission over the binary erasure channel (BEC), an estimation for the scaling exponent is known, namely $\mu \approx 3.627$. Therefore, compared to random codes which have a scaling exponent of 2, polar codes require larger block lengths to achieve the same rate and error probability. For a generic BMSC, taking as a proxy of the error probability the sum of the Bhattacharyya parameters, there exists a universal parameter μ' , such that reliable communication requires rates that satisfy $R < C - \alpha N^{-1/\mu'}$, where α is a positive constant [5], [6]. The exponent μ' is lower bounded by 3.553 and it has been conjectured that its value can be increased up to the scaling parameter of the BEC, i.e., $\mu' = \mu \approx 3.627$.

In order to improve the finite-length performance of polar codes, several decoding algorithms have been proposed. Maximum likelihood (ML) decoders are implemented via the Viterbi algorithm [7] and via sphere decoding [8], but are practical only for relatively short block lengths. A linear programming (LP) decoder is introduced in [9], and the performance under belief propagation (BP) decoding is considered in [10]. The stopping set analysis for the special case of the transmission over the BEC is also provided in [11]. A successive cancellation list (SCL) decoder is proposed in [12]. Empirically, the usage of L concurrent decoding paths yields a significant improvement in the achievable error probability and allows to obtain an error probability comparable to that under MAP decoding with practical values of the list size. However, it has been recently shown that, under MAP decoding, the introduction of any finite list does not change the scaling exponent [13]. In particular, for any BMSC and for any family of linear codes with unbounded minimum distance, list decoding cannot modify the scaling behavior for finite values of the list size. Analogously, under genie-aided SC decoding, the scaling exponent stays constant for any fixed number of helps from the genie, when transmission takes place over the BEC.

Reed-Muller Codes and Their Relation to Polar Coding. RM codes were introduced by Muller [14] and rediscovered shortly thereafter with an efficient decoding algorithm by Reed [15]. The relation between polar codes and RM codes is also pointed out in [1] and performance comparisons are carried out in [16], [17]. Furthermore, Dumer’s recursive decoding algorithm for RM codes [18] is similar to the SC decoder for polar codes [19]. Numerical simulations and analytical results suggest that RM codes have a bad performance under successive and iterative decoding, but they outperform polar codes under MAP decoding [1], [10]. However, no rigorous results are known and the fundamental problem concerning whether RM codes are capacity-achieving under MAP decoding, at

M. Mondelli, S. H. Hassani, and R. Urbanke are with the School of Computer and Communication Sciences, EPFL, CH-1015 Lausanne, Switzerland. Emails: {marco.mondelli, seyedhamed.hassani, ruediger.urbanke}@epfl.ch.

least for some channels with a sufficient amount of symmetry, remains open [20].

Contribution of the Present Work. In this paper we propose an interpolation method between the polar code of block length N and rate R and an RM code of the same block length and rate. To do so, we describe a family of codes $\mathcal{C}_{\text{inter}} = \{C_\alpha : \alpha \in [0, 1]\}$ such that $C_\alpha|_{\alpha=1}$ is the original polar code, and $C_\alpha|_{\alpha=0}$ is an RM code. We remark that experimentally the error probability under MAP decoding increases with α . Even if MAP decoding is in general an NP-complete task, this result is relevant in practice because picking suitable codes from $\mathcal{C}_{\text{inter}}$ boosts the finite length performance of the original polar code also when low-complexity suboptimal algorithms are employed. In particular, a remarkable performance improvement is noticed adopting the SCL decoder proposed in [12] and the BP decoder. This performance gain could be substantial in the sense of the reduction of the scaling exponent: according to numerical simulations performed for $N = 2^{10}$ over the BEC, the error probability under MAP decoding for the transmission of C_α for α sufficiently small is very close to that of random codes. As a result, the usage of codes in $\mathcal{C}_{\text{inter}}$ potentially improves the speed at which capacity is reached.

Organization. Section II points out similarities and differences between the polar and the RM construction and describes explicitly the interpolating family $\mathcal{C}_{\text{inter}}$ for the special case of the transmission over the BEC. Starting from the analysis of the two extreme cases of MAP and SC decoding, Section III shows how to improve significantly the finite-length performance of polar codes by using codes of the form C_α decoded with low-complexity suboptimal schemes when transmission takes place over the BEC. The interpolation method between RM and polar codes is described for the transmission over a generic BMSC W in Section IV, where the simulation results for the binary additive white Gaussian noise channel (BAWGNC) are presented as a case study. Finally, Section V draws the conclusions of the paper.

II. FROM POLAR TO RM CODES: AN INTERPOLATION METHOD FOR THE BEC

Let $n \in \mathbb{N}$ and $N = 2^n$. Consider the $N \times N$ matrix G_N defined as follows,

$$G_N = F^{\otimes n}, \quad F = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \quad (1)$$

where $F^{\otimes n}$ denotes the n -th Kronecker power of F . As it has been formerly pointed out in [1], the generator matrices of both polar and RM codes are obtained by suitably selecting rows from $G_N = (g_1, \dots, g_N)^T$.

In particular, the *RM rule* for building a code of block length N and minimum distance 2^k for some fixed $k \in \{0, 1, \dots, n\}$ consists in choosing the rows of G_N with Hamming weight at least 2^k . Thus, the rate R of this code is given by

$$R = \frac{\sum_{i=k}^n \binom{n}{i}}{N}. \quad (2)$$

In general, if we require an RM code with fixed block length N and rate R , where R cannot be written in the form (2) for some $k \in \mathbb{N}$, we take as generator matrix any subset of NR rows of G_N with the highest Hamming weights. Notice that this criterion is channel-independent in the sense that it does not rely on the particular channel over which the transmission takes place.

On the other hand, the *polar rule* is channel-specific. Indeed, the N synthetic channels $W_N^{(i)}$ ($i \in \{0, \dots, N-1\}$) are obtained from N independent copies of the original channel W . The row g_i is associated to $W_N^{(i)}$ and the synthetic channels (and, therefore, the rows) with the lowest Bhattacharyya parameters are selected. In general, different channels W yield different choices of rows. Let us consider the simple case of the transmission over the BEC(ε) for fixed $\varepsilon \in (0, 1)$. In this particular scenario, the Bhattacharyya parameter Z_i associated to $W_N^{(i)}$ (and, therefore, to g_i) is given by

$$Z_i(\varepsilon) = f_{b_1^{(i)}} \circ f_{b_2^{(i)}} \circ \dots \circ f_{b_n^{(i)}}(\varepsilon), \quad (3)$$

where $f_0(x) = 1 - (1-x)^2$, $f_1(x) = x^2$, \circ denotes function composition, and $b^{(i)} = (b_1^{(i)}, b_2^{(i)}, \dots, b_n^{(i)})^T$ is the binary expansion of i over n bits, $b_1^{(i)}$ being the most significant bit and $b_n^{(i)}$ the least significant bit. In order to construct a code of block length N and rate R , we select the NR rows which minimize the expression (3).

The link between the RM rule and the polar rule is clarified by the following proposition.

Proposition 1: The polar code of block length N and rate R designed for transmission over a BEC(ε), when $\varepsilon \rightarrow 0$, is an RM code.

Proof: Suppose that the thesis is false, i.e., that we include g_{j^*} , but not g_{i^*} , with $w_H(g_{i^*}) > w_H(g_{j^*})$, where $w_H(\cdot)$ denotes the Hamming weight. Since $w_H(g_i) = 2^{\sum_{k=1}^n b_k^{(i)}} = 2^{w_H(b^{(i)})}$ for any $i \in \{0, \dots, N-1\}$ (Proposition 17 of [1]), then $w_H(b^{(i^*)}) > w_H(b^{(j^*)})$.

From formula (3), one deduces that $Z_i(\varepsilon)$ is a polynomial in ε with minimum degree equal to $2^{w_H(b^{(i)})}$. Hence,

$$\lim_{\varepsilon \rightarrow 0} \frac{Z_{i^*}(\varepsilon)}{Z_{j^*}(\varepsilon)} = 0,$$

which means that there exists $\delta > 0$ s.t. for all $\varepsilon < \delta$, $Z_{i^*}(\varepsilon) < Z_{j^*}(\varepsilon)$. Consequently, a polar code designed for transmission over a BEC(ε), with $\varepsilon < \delta$, which includes g_{j^*} must also include g_{i^*} . This is a contradiction. ■

Recall that the transmission takes place over $W = \text{BEC}(\varepsilon)$. Let C_α be the polar code of block length N and rate R designed for a BEC($\alpha\varepsilon$). When $\alpha = 1$, C_α reduces to the polar code for the channel W , while, when $\alpha \rightarrow 0$, C_α becomes an RM code by Proposition 1. Consider the family of codes $\mathcal{C}_{\text{inter}}$ defined as,

$$\mathcal{C}_{\text{inter}} = \{C_\alpha : \alpha \in [0, 1]\}. \quad (4)$$

The codes in $\mathcal{C}_{\text{inter}}$ provide an *interpolation method* to pass smoothly from a polar code to an RM code of the same rate and block length. Indeed, consider the generator matrices of the codes in $\mathcal{C}_{\text{inter}}$ which are obtained reducing α from 1

to 0. We start from the generator matrix of the polar code and the successive matrices are obtained by changing one row at a time. In particular, numerical simulations show that the row which is included in the next code (associated to a smaller α) has a higher Hamming weight than the row which was removed from the previous code (associated to a higher α). Heuristically, this happens for the following reason. The row indices chosen by C_α are the ones which minimize the associated Bhattacharyya parameters $Z_i(\alpha\varepsilon)$ given by (3). As $f_1(x) \leq f_0(x)$ for any $x \in [0, 1]$, applying f_1 instead of f_0 makes the Bhattacharyya parameter decrease. However, also the order in which the functions are applied is important, since $f_0 \circ f_1(x) \leq f_1 \circ f_0(x)$ for any $x \in [0, 1]$: if we fix $w_H(b^{(i)})$, Z_i is minimized by applying first all the functions f_1 and then the functions f_0 . Therefore, the goodness of the index i depends both on the number of 1's in its binary expansion $b^{(i)}$ and on the positions of these 1's. On the other hand, when designing an RM code only $w_H(b^{(i)})$ matters and, for α small enough, C_α tends to an RM code. As a result, as α goes from 1 to 0, the value of $Z_i(\alpha\varepsilon)$ depends more and more on $w_H(b^{(i)})$ than on the position of the 1's in $b^{(i)}$.

III. IMPROVING THE FINITE-LENGTH PERFORMANCE OF POLAR CODES FOR THE BEC

The focus of this section is on the performance of the codes in C_{inter} when transmission takes place over the BEC(ε). We start considering the MAP decoder and then move to the SC decoder introduced by Arkan. By taking into account low-complexity suboptimal decoding schemes which outperform the original SC algorithm (e.g., SCL and BP), we highlight the advantage of employing codes of the form C_α . The simulation results of this section refer to codes of fixed block length $N = 2^{10}$ and rate $R = 0.5$. The number of Monte Carlo trials is $M = 10^5$.

A. Motivation: MAP Decoding

Since it has been observed that under MAP decoding picking the rows of G_N according to the RM rule significantly improves the performance with respect to the polar choice [10], it is interesting to analyze the error probability $P_e^{\text{MAP}}(\alpha, \varepsilon)$ under MAP decoding for the transmission of the code C_α over the BEC(ε). Although MAP decoding is in general an NP-complete task, for the particular case of the BEC it is equivalent to the inversion of a suitable matrix and, therefore, can be performed in $\mathcal{O}(N^3)$.

First of all, fix the value of ε and consider how P_e^{MAP} varies as a function of α . As it is shown in Figure 1 for four distinct values of ε , $P_e^{\text{MAP}}(\alpha, \varepsilon)$ is increasing in α . In short, the proposed interpolation method to pass from the polar code $C_\alpha|_{\alpha=1}$ to an RM code $C_\alpha|_{\alpha=0}$ yields a family of codes with decreasing MAP error probability. This conjecture, if proved, would imply that RM codes are capacity-achieving for the BEC, which is a long-standing open problem in coding theory. Another evidence in support of this statement is as follows. As it has been pointed out in Section II, the polar rule differs from the RM rule in the fact that not only the number, but also the position of the 1's in $b^{(i)}$ matters in the choice of

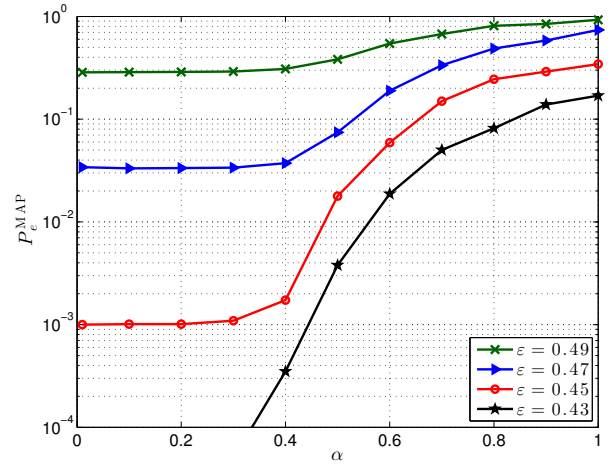


Figure 1. Error probability P_e^{MAP} under MAP decoding for the transmission of C_α over the BEC(ε), when α varies in $[0, 1]$ and ε is given four distinct values. The block length is $N = 2^{10}$ and the rate is $R = 0.5$. Observe that P_e^{MAP} is increasing in α for all values of ε , which means that the minimum error probability is achieved by the RM code $C_\alpha|_{\alpha=0}$.

the row indices. In particular, polar codes prefer to set the 1's in the least significant bits of the binary expansion of i . However, if one is concerned with achieving the capacity of the BEC under MAP decoding, the specific order of the 1's in the binary expansions of the row indices does not play any role. Indeed, denote by \mathcal{F} the set of row indices of G_N which are not chosen for the generator matrix of the polar code (these indices are *frozen*, since they are not used for the transmission of information bits) and let \mathcal{F}^c be its complement. Then, it is possible to arbitrarily permute the binary expansions $b^{(i)}$ ($i \in \mathcal{F}^c$) and still get a set of row indices which yields a capacity-achieving family of codes under MAP decoding. This fact is formalized in the following proposition.

Proposition 2: Denote by \mathcal{F}^c the set of row indices chosen by polar coding. Let $\pi : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ be a permutation and let P_π be the associated permutation matrix. Construct the code C_π by taking the rows of G_N whose indices have binary expansions $P_\pi b^{(i)}$. Let $\varepsilon \in (0, 1)$ and denote by $P_e^{\mathcal{D}}(C_\pi)$ the error probability under the decoder \mathcal{D} for the transmission of C_π over the BEC(ε). Then, $P_e^{\text{MAP}}(C_\pi) \leq P_e^{\text{SC}}(C_l)$, C_l being the original polar code.

Proof: As observed in [10], there exist $n!$ different representations of the polar code C_l of block length $N = 2^n$ obtained by permuting the n layers of connections. Let us apply the permutation τ to these layers and then run the SC algorithm, denoting by $P_e^{\text{SC}, \tau}(C_l)$ the error probability for transmission over the BEC(ε). The application of the permutation τ affects the Bhattacharyya parameter Z_i associated to the synthetic channel $W_N^{(i)}$, which is now given by

$$Z_i(\varepsilon) = f_{\tau(b_1^{(i)})} \circ f_{\tau(b_2^{(i)})} \circ \dots \circ f_{\tau(b_n^{(i)})}(\varepsilon).$$

On the other hand, the generator matrix (and, consequently, the set \mathcal{F}^c) does not change, because the code stays the same.

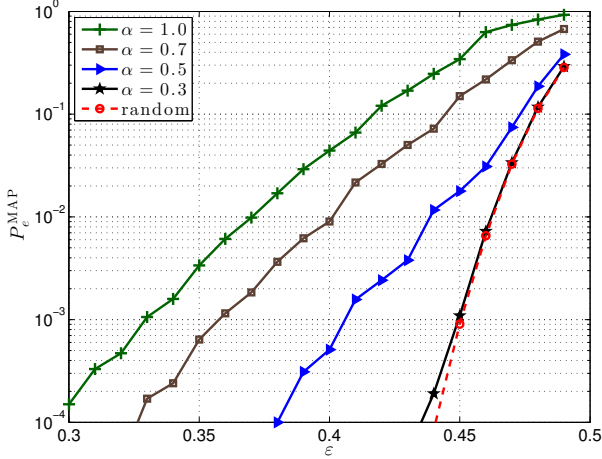


Figure 2. Error probability P_e^{MAP} under MAP decoding for the transmission of C_α over the BEC(ϵ), when ϵ varies in $\{0.30, 0.31, \dots, 0.49\}$ and α is given four distinct values. The block length is $N = 2^{10}$ and the rate is $R = 0.5$. Remark that already for $\alpha = 0.3$ the error performance of C_α is comparable to that of random codes.

Therefore, the probability that the SC decoder fails when applying the permutation τ to the layers of the code C_l equals the probability that the SC decoder fails when the code C_τ is employed. In formulas, for any permutation τ ,

$$P_e^{\text{SC}, \tau}(C_l) = P_e^{\text{SC}}(C_\tau).$$

Denote by OSC the algorithm which runs SC decoding over all the $n!$ possible overcomplete representation of a polar code. When transmission takes place over the BEC, the OSC decoder fails if and only if there exists an information bit which cannot be decoded by any of these $n!$ SC decoders. Let $P_e^{\text{OSC}}(C_\pi)$ be the error probability under OSC decoding for transmission of the code C_π over the BEC(ϵ). Then, $P_e^{\text{OSC}}(C_\pi) \leq P_e^{\text{SC}, \tau}(C_\pi)$ for any τ . Taking $\tau = \pi^{-1}$ and recalling that MAP decoding minimizes the error probability, we obtain that

$$P_e^{\text{MAP}}(C_\pi) \leq P_e^{\text{OSC}}(C_\pi) \leq P_e^{\text{SC}, \pi^{-1}}(C_\pi) = P_e^{\text{SC}}(C_l),$$

which gives us the desired result. ■

In Figure 2 we fix the value of α and we analyze P_e^{MAP} as a function of ϵ . It is interesting to remark that already for $\alpha = 0.3$, the error probability for the transmission of C_α is very close to that of random coding, which not only achieves capacity, but does so with a more favorable tradeoff between N and $C - R$. Indeed, random codes have a scaling exponent $\mu = 2$, while the scaling exponent of polar codes is $\mu = 3.627$.

B. SC Decoding

After dealing with optimal MAP decoding, let us analyze the performance of the codes in $\mathcal{C}_{\text{inter}}$ under SC decoding. As can be seen in Figure 3 for four distinct values of ϵ , the error probability $P_e^{\text{SC}}(\alpha, \epsilon)$ under SC decoding for transmission of the code C_α over the BEC(ϵ) is a decreasing function of α . Hence, the best performance are obtained using the polar code

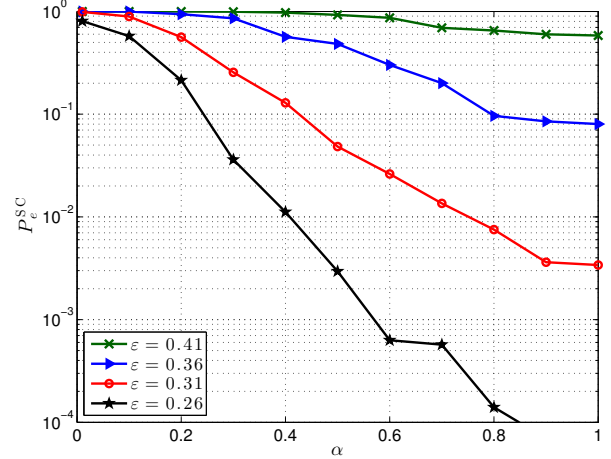


Figure 3. Error probability P_e^{SC} under SC decoding for the transmission of C_α over the BEC(ϵ), when α varies in $[0, 1]$ and ϵ is given four distinct values. The block length is $N = 2^{10}$ and the rate is $R = 0.5$. Observe that P_e^{SC} is decreasing in α , which means that the minimum P_e^{SC} is achieved by the original polar code $C_\alpha|_{\alpha=1}$.

$C_\alpha|_{\alpha=1}$. The theoretical reason of this behavior lies in the fact that P_e^{SC} can be well approximated by the sum of the Bhattacharyya parameters of the synthetic channels which are selected by the polar code for transmission of the information bits [21]. Formally, let $\mathcal{F}^c(\alpha)$ be the set of indices which are selected by the polar code C_α . Then,

$$P_e^{\text{SC}}(\alpha) \lesssim \sum_{i \in \mathcal{F}^c(\alpha)} Z_i(\epsilon). \quad (5)$$

The bound (5) is tight and $\sum_{i \in \mathcal{F}^c(\alpha)} Z_i(\epsilon)$ is minimized for $\alpha = 1$.

C. Something Between the Two Extremes: List Decoding and Belief Propagation

Consider the SCL scheme introduced in [12] and denote by $P_e^{\text{SCL}}(\alpha, \epsilon, L)$ the error probability under SCL decoding with list size L for transmission of the polar code C_α over the BEC(ϵ). Clearly, if $L = 1$, this scheme reduces to the SC algorithm originally proposed by Arkan, while for $L \geq 2^{NR}$, the SCL decoder is equivalent to the MAP decoder, since the list is big enough to contain all the possible 2^{NR} codewords. Therefore, as L increases, we gradually pass from SC decoding to MAP decoding.

If we fix α and we let L grow, $P_e^{\text{SCL}}(\alpha, \epsilon, L)$ monotonically decreases from $P_e^{\text{SC}}(\alpha, \epsilon)$ to $P_e^{\text{MAP}}(\alpha, \epsilon)$. Values of α close to 1 imply that $P_e^{\text{SCL}}(\alpha, \epsilon, L)$ gets close to the MAP error probability for small values of the list size. If α is reduced, a bigger list size is required to obtain performance comparable to MAP decoding since the underlying SC algorithm gets worse, but $P_e^{\text{MAP}}(\alpha, \epsilon)$ becomes significantly smaller. In other words, a smaller α implies a slower converge (in terms of L) toward a smaller error probability. This trade-off between MAP error

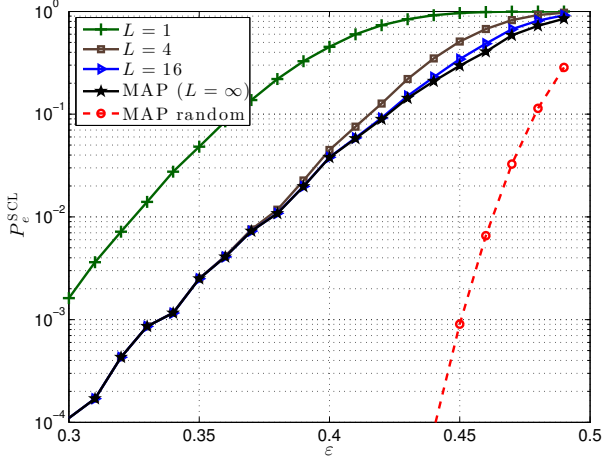
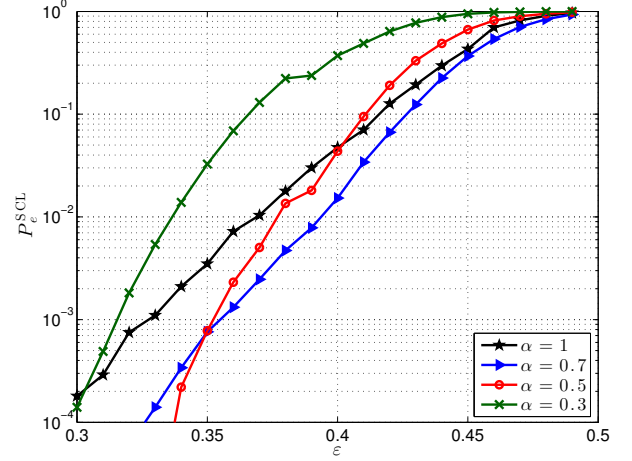
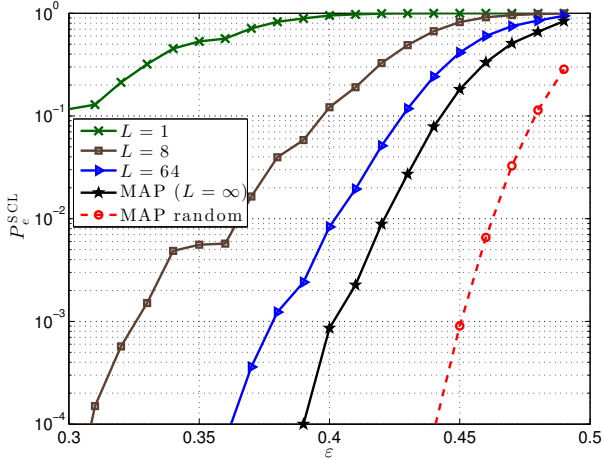
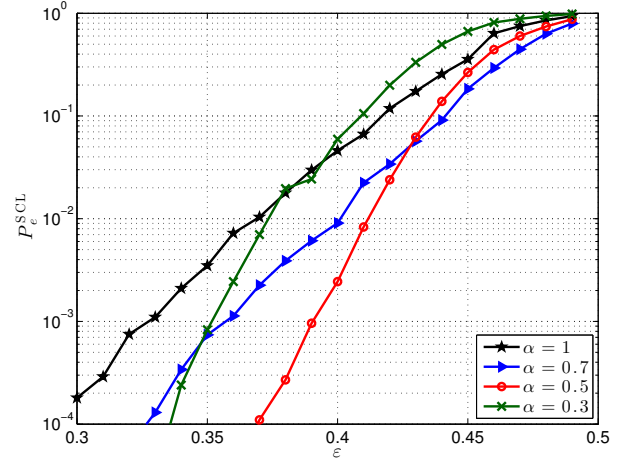
(a) $\alpha = 0.9$ (a) $L = 8$ (b) $\alpha = 0.4$ (b) $L = 64$

Figure 4. Error probability P_e^{SCL} under SCL decoding for the transmission of C_α over the BEC(ε), when ε varies in $\{0.30, 0.31, \dots, 0.49\}$ and for different values of the list size L . The block length is $N = 2^{10}$ and the rate is $R = 0.5$. As a benchmark, we represent also the error probability under MAP decoding for the transmission of C_α (in black) and for the transmission of a random code (in red). Observe that if α is big (upper plot), P_e^{SCL} converges to P_e^{MAP} already small values of the list size. On the other hand, if α is small (lower plot), bigger list sizes are required to get to the error probability of MAP decoding, which in return becomes much smaller in value and, therefore, much closer, to the error probability of a random code.

probability and list size required to reach it is illustrated in Figure 4 for $\alpha = 0.9$ and $\alpha = 0.4$, where, as a benchmark, we represent also the average error probability under MAP decoding for the transmission of random codes.

In order to show that the usage of codes in $\mathcal{C}_{\text{inter}}$ significantly improves the finite-length performance of polar codes for practical values of the list size, fix L and consider the transmission of C_α for different values of α . The results for $L = 8$ and $L = 64$ are represented in Figure 5. The code

Figure 5. Error probability P_e^{SCL} under SCL decoding for the transmission of C_α over the BEC(ε), when ε varies in $\{0.30, 0.31, \dots, 0.49\}$ and for different values of α . The block length is $N = 2^{10}$ and the rate is $R = 0.5$. Already when $L = 8$ (upper plot), a performance improvement is obtained reducing α with respect to the original polar code $C_\alpha|_{\alpha=1}$. If the list size is increased to $L = 64$ (lower plot), the advantage in considering codes C_α with a smaller value of the tuning parameter α is even more evident.

$C_\alpha|_{\alpha=0.7}$ outperforms the original polar scheme already when $L = 8$. If the decoder is allowed to take $L = 64$, the improvement in performance is even more significant and, for example, the target error probability $P_e = 10^{-3}$ can be obtained for $\varepsilon = 0.39$ if we employ $C_\alpha|_{\alpha=0.5}$, while $\varepsilon = 0.35$ is required if we employ the original polar code $C_\alpha|_{\alpha=1}$. Remark that if the target error probability to be met is very low, it is convenient to consider codes C_α with small α , since they will be able to achieve it for higher erasure probabilities of the BEC. Indeed, observe that in the case $L = 64$, $C_\alpha|_{\alpha=0.3}$ outperforms $C_\alpha|_{\alpha=0.7}$ for $P_e^{\text{SCL}} < 10^{-3}$. This effect is due to

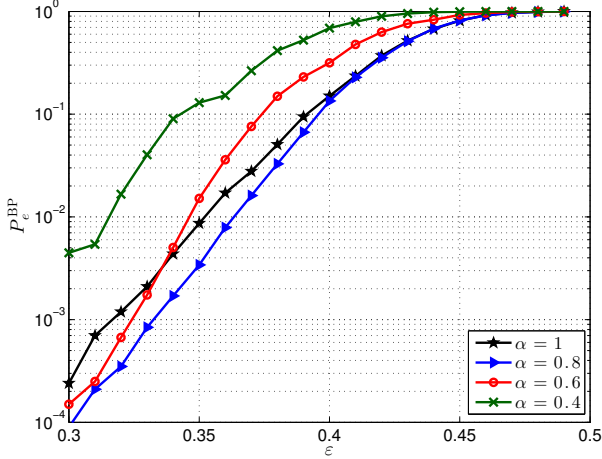


Figure 6. Error probability P_e^{BP} under BP decoding for the transmission of C_α over the $\text{BEC}(\varepsilon)$, when ε varies in $\{0.30, 0.31, \dots, 0.49\}$ and α is given four distinct values. The block length is $N = 2^{10}$ and the rate is $R = 0.5$. Remark that the optimal performance is obtained with the code $C_\alpha|_{\alpha=0.8}$.

the fact that, for any fixed rate less than capacity, P_e^{SC} scales with N as $2^{-\sqrt{N}}$ and, hence, polar codes are not affected by error floors.

In general, it is convenient to consider codes of the form C_α whenever the decoding algorithm yields better results than the SC decoder. As another example, consider the case of the BP decoder. It has been already pointed out that the polar choice of the row indices to be selected from G_N is not optimal for the BP algorithm [10], [11], but no systematic rule capable of outperforming polar codes is known. As can be seen in Figure 6, the interpolating family $\mathcal{C}_{\text{inter}}$ contains codes which achieve a smaller error probability than that of the original polar code $C_\alpha|_{\alpha=1}$ for an appropriate choice of the parameter α .

IV. GENERALIZATION TO ANY BMSC

This section is devoted to the generalization of the ideas expressed for the BEC in Sections II and III to the transmission over a BMSC W . In particular, first we propose a method for constructing the family of codes $\mathcal{C}_{\text{inter}}$ and, then, we analyze the performance for the transmission over a BAWGNC.

A. General Construction of an Interpolating Family

Suppose that the transmission takes place over the BMSC W and let $Z(W)$ be its Bhattacharyya parameter. In order to construct the interpolating family $\mathcal{C}_{\text{inter}}$, we consider the family of channels $\mathcal{W}_{\text{inter}}$ ordered by degradation [22] such that the element of the family with the biggest Bhattacharyya parameter is W itself and the element of the family with the smallest Bhattacharyya parameter is the perfect channel W^{opt} , in which the output is equal to the input with probability 1. There are many ways of performing such a task. In particular, we can set

$$\mathcal{W}_{\text{inter}} = \{W_\alpha : \alpha \in [0, 1]\}, \quad (6)$$

where $W_\alpha = W$ with probability α , $W_\alpha = W^{\text{opt}}$ with probability $1 - \alpha$, and the receiver knows which channel has been used. In formulas, $W_\alpha = \alpha W + (1 - \alpha)W^{\text{opt}}$.

Since the convex combination of BMS channels is a BMS channel, W_α is also a BMSC with Bhattacharyya parameter $Z_\alpha = \alpha Z$. Denote by C_α the polar code for transmission over W_α . Then, the interpolating family $\mathcal{C}_{\text{inter}}$ is defined as in (4). This is a reasonable choice for $\mathcal{C}_{\text{inter}}$ because of the following result, which extends Proposition 1.

Proposition 3: Let W be a BMSC, W^{opt} be the perfect channel and $\alpha \in [0, 1]$. Denote by C_α the polar code of block length N and rate R designed for transmission over the BMSC $W_\alpha = \alpha W + (1 - \alpha)W^{\text{opt}}$. Then, when $\alpha \rightarrow 0$, C_α is an RM code.

Proof: When transmission takes place over the BMSC W_α , the Bhattacharyya parameter $Z_i(W_\alpha)$ of the i -th synthetic channel $W_{\alpha, N}^{(i)}$ ($i \in \{0, \dots, N - 1\}$) has the form (3), where ε is replaced by $Z_\alpha = \alpha Z$, $f_1(x) = x^2$, and $f_0(x)$ can be bounded as [1]

$$x \leq f_0(x) \leq 2x - x^2. \quad (7)$$

Suppose that g_{j^*} is included in the generator matrix of the code, but not g_{i^*} , with $w_H(g_{i^*}) > w_H(g_{j^*})$. Then, using (7), Z_{i^*} can be upper bounded by a polynomial in α with minimum degree $w_H(g_{i^*})$ and Z_{j^*} can be lower bounded by a polynomial in α with minimum degree $w_H(g_{j^*})$. Thus, for α small enough $Z_{i^*} < Z_{j^*}$ and we reach a contradiction. ■

Remark that if $W = \text{BEC}(\varepsilon)$, then $W_\alpha = \text{BEC}(\alpha\varepsilon)$. In general, there might be more natural ways to obtain the family of codes $\mathcal{C}_{\text{inter}}$, according to the particular choice of the channel W . Indeed, in Section IV-B which deals with the case of the BAWGNC, the interpolating family is constructed in a different way.

Once obtained a family of codes of the form C_α , where $C_\alpha|_{\alpha=1}$ is the polar code designed for transmission over the channel W and $C_\alpha|_{\alpha=0}$ is an RM code, numerical simulations show that the error probability under MAP decoding is an increasing function of α . On the other hand, under SC decoding, the optimal performance is still achieved using $C_\alpha|_{\alpha=1}$. If one considers low-complexity decoding algorithms which get close to the error probability under MAP decoding, the finite-length performance of polar codes is significantly improved by using the code C_α for a suitable choice of the parameter α .

B. Case Study: $W = \text{BAWGNC}(\sigma^2)$

Let $W = \text{BAWGNC}(\sigma^2)$ and define C_α as the polar code designed for transmission over $W_\alpha = \text{BAWGNC}(\alpha\sigma^2)$. As $\alpha \rightarrow 0$, W_α tends to the perfect channel W^{opt} and C_α becomes an RM code. In order to show the performance improvement guaranteed by the usage of codes in the interpolating family $\mathcal{C}_{\text{inter}}$ defined as in (4), consider the SCL decoder. To be coherent with the simulation setup of [12], the numerical simulations refer to codes of fixed block length $N = 2^{11}$ and rate $R = 0.5$. The number of Monte Carlo trials is $M = 10^5$. The codes are optimized for an SNR = 2 dB, namely, $\sigma^2 = 0.6309$. The results of Figure 7 are qualitatively

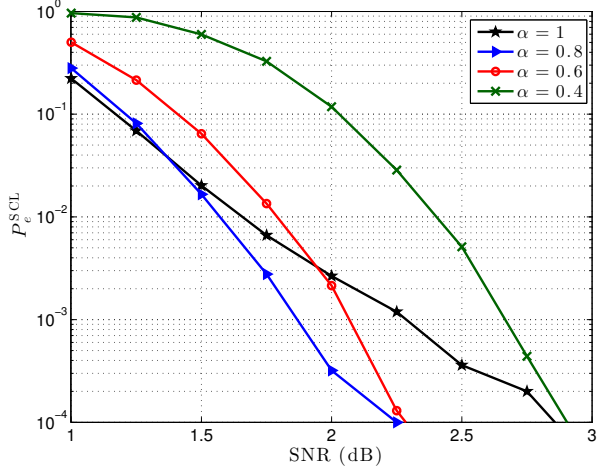
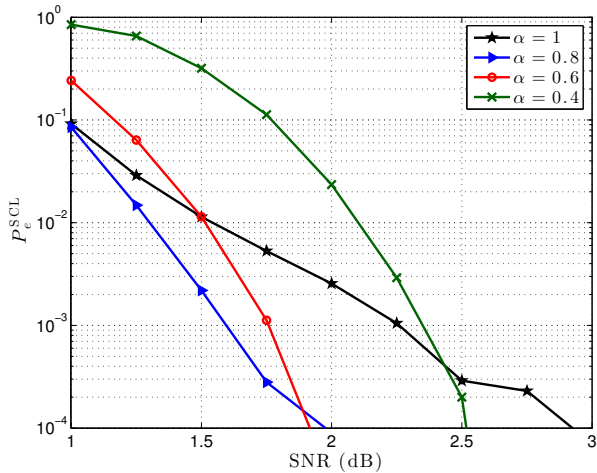
(a) $L = 8$ (b) $L = 64$

Figure 7. Error probability P_S^{SCL} under SCL decoding for the transmission of C_α over the BAWGNC(σ^2), where $\sigma^2 = 0.6309$, the SNR varies in $\{1, 1.25, \dots, 3\}$ and $\alpha \in \{0.4, 0.6, 0.8, 1\}$. The block length is $N = 2^{11}$ and the rate is $R = 0.5$. For the target error probability $P_e = 10^{-3}$ an improvement ≥ 0.5 dB with respect to the original polar code $C_\alpha|_{\alpha=1}$ can be noticed using the codes $C_\alpha|_{\alpha=0.8}$ and $C_\alpha|_{\alpha=0.6}$.

similar to those represented in Figure 5 for the BEC and testify the remarkable performance gain achievable by codes of the form C_α with respect to the original polar code $C_\alpha|_{\alpha=1}$.

V. CONCLUDING REMARKS

As pointed out in [12], the error probability of polar codes at practical block lengths can be reduced by acting both on the decoder and on the code itself. Unfortunately, an improvement only in the decoding algorithm does not seem to be enough to change the scaling exponent [13]. In this work we address the issue of boosting the finite-length performance of polar codes by modifying jointly the code and the SC

decoding algorithm. In particular, we construct a family of codes $\mathcal{C}_{\text{inter}} = \{C_\alpha : \alpha \in [0, 1]\}$ of fixed block length and rate which interpolates from the original polar code $C_\alpha|_{\alpha=1}$ to the RM code $C_\alpha|_{\alpha=0}$. Numerically, the error probability under MAP decoding decreases as α goes from 1 to 0. Since MAP decoding is not practical for transmission over general channels, we develop a trade-off between complexity and performance by considering low-complexity decoders (e.g., BP, SCL), thus showing the significant benefit coming from the adoption of codes in $\mathcal{C}_{\text{inter}}$. This improvement in the finite-length performance of polar codes can be substantial: we provide experimental evidence of the fact that the error probability under MAP decoding for the transmission over the BEC of C_α for α sufficiently small is very close to that of random codes, which achieve a better scaling exponent than polar codes.

ACKNOWLEDGEMENT

The authors would like to thank M. B. Parizi for providing the code which simulates the SCL decoder for transmission over the BAWGNC. This work was supported by grant No. 200020_146832/1 of the Swiss National Foundation.

REFERENCES

- [1] E. Arıkan, "Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.
- [2] E. Arıkan and I. E. Telatar, "On the rate of channel polarization," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, July 2009, pp. 1493–1495.
- [3] S. H. Hassani and R. Urbanke, "On the scaling of polar codes: I. The behavior of polarized channels," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, June 2010, pp. 874–878.
- [4] S. B. Korada, A. Montanari, I. E. Telatar, and R. Urbanke, "An empirical scaling law for polar codes," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, June 2010, pp. 884–888.
- [5] A. Goli, S. H. Hassani, and R. Urbanke, "Universal bounds on the scaling behavior of polar codes," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, July 2012, pp. 1957–1961.
- [6] S. H. Hassani, K. Alishahi, and R. Urbanke, "Finite-length scaling of polar codes," *submitted to IEEE Trans. Inf. Theory*, Apr. 2013.
- [7] E. Arıkan, H. Kim, G. Markarian, U. Ozgur, and E. Poyraz, "Performance of short polar codes under ml decoding," in *ICT-Mobile Summit Conf. Proc.*, 2009.
- [8] S. Kahrman and M. E. Celebi, "Code based efficient maximum-likelihood decoding of short polar codes," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, July 2012, pp. 1967–1971.
- [9] N. Goela, S. B. Korada, and M. Gastpar, "On LP decoding of polar codes," in *IEEE Inf. Theory Workshop (ITW)*, Sept. 2010, pp. 1–5.
- [10] N. Hussami, S. B. Korada, and R. Urbanke, "Performance of polar codes for channel and source coding," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, July 2009, pp. 1488–1492.
- [11] A. Eslami and H. Pishro-Nik, "On finite-length performance of polar codes: stopping sets, error floor, and concatenated design," *IEEE Trans. Commun.*, vol. 61, no. 3, pp. 919–929, Mar. 2013.
- [12] I. Tal and A. Vardy, "List decoding of polar codes," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, Aug. 2011, pp. 1–5.
- [13] M. Mondelli, S. H. Hassani, and R. Urbanke, "Scaling exponent of list decoders with applications to polar codes," *accepted at IEEE Inf. Theory Workshop (ITW)*, 2013.

- [14] D. E. Muller, "Application of boolean algebra to switching circuit design and to error detection," *IRE Trans. Electronic Computers*, vol. EC-3, no. 3, pp. 6–12, 1954.
- [15] I. Reed, "A class of multiple-error-correcting codes and the decoding scheme," *IRE Trans. Electronic Computers*, vol. 4, no. 4, pp. 38–49, 1954.
- [16] E. Arıkan, "A performance comparison of polar codes and reed-muller codes," *Communications Letters, IEEE*, vol. 12, no. 6, pp. 447–449, June 2008.
- [17] —, "A survey of reed-muller codes from polar coding perspective," in *IEEE Inf. Theory Workshop (ITW)*, Jan. 2010, pp. 1–5.
- [18] I. Dumer, "Recursive decoding and its performance for low-rate reed-muller codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 811–823, May 2004.
- [19] S. B. Korada, "Polar codes for channel and source coding," Ph.D. dissertation, EPFL, 2009.
- [20] D. J. Costello and J. Forney, G. D., "Channel coding: The road to channel capacity," *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1150–1177, June 2007.
- [21] M. B. Parizi and I. E. Telatar, "On correlation between polarized BECs," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, July 2013, pp. 784–788.
- [22] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008.