An elementary description of polarization process

Ilya Dumer

Abstract—We analyze successive cancellation (SC) decoder by using two random functions. The first function is related to the likelihoods of 0 and 1 in each code position, while the second gives the difference between their posterior probabilities. We then study the second power moments of both functions. We show that these moments are being squared in channel transformations, while their product tends to 0 for growing lengths n. This gives an elementary proof of polarization properties of SC decoding. We also derive a simple ordering of decoding channels with construction complexity of order $n \log n$.

Index terms: Polar codes; Reed-Muller codes; successive cancellation decoding; polarization.

I. INTRODUCTION

In this paper, we analyze one algorithm of successive cancellation (SC) decoding and give an elementary proof of its polarizing behavior. This SC algorithm was first applied in [1] to the general Reed-Muller (RM) codes $RM(r, m)$ and showed that RM codes yield vastly different output bit error rates (BER) for different information bits. This disparity was addressed in [1] by eliminating some information bits with the highest BERs. Simulation results of [2] showed that the optimal selection of the eliminated (frozen) bits drastically improves decoding of the original RM codes. However, the analytical tools used in these and later publications [3]-[4] do not reveal polarization properties or capacity-reaching performance of these bit-frozen subcodes.

A major breakthrough in this area was achieved by E. Arikan [5], who proposed a new analytical technique, which reveals some novel properties of generic recursive processing, such as bit polarization. The main result of [5] shows that the optimal bit-frozen subcodes of the full codes $RM(m, m)$ - now well known as polar codes - achieve the channel capacity of any symmetric memoryless channel as $m \rightarrow \infty$. This technique also shows that capacity-achieving subcodes exist for all codes $RM(r, m)$ of rate $R \rightarrow 1$, such as RM codes with $\lim r/m > 1/2$.

In this paper, we wish to simplify the results on polarization properties of SC decoding [5]-[7] and the results on its fast polarization [8] - [10]. We also present a simple ordering technique for decoding channels. Section II gives a polynomial description of the Plotkin $(u, u + v)$-construction. Section III describes SC decoding of [1] that uses two different random variables (rv). For any received symbol y, one rv is the likelihood $h = P(0|y)/P(1|y)$. The other rv $g = P(0|y) - P(1|y)$ measures the variation between the posterior probabilities of the transmitted symbols 0 and 1. We show that quantities g and h are transformed as the products $g_1 g_2$ and $h_1 h_2$ on any degrading and upgrading channel, respectively.

I. Dumer is with the College of Engineering, University of California, Riverside, CA 92521, USA; email: dumer@ece.ucr.edu

Section IV introduces two new parameters, $\mathcal{A}$ and $\mathcal{B}$, which are related to the power moments of rv g and h. Both are also related to the Bhattacharyya parameter $\mathcal{Z}$. Section V shows that parameters $\mathcal{A}$ and $\mathcal{B}$ square up on the upgrading and degrading channels, respectively. Section VI validates polarization property of SC decoding, by proving that the product $\mathcal{A}\mathcal{B}$ tends to 0 for almost all sequences (except a vanishing fraction) of $m \rightarrow \infty$ channel transformations. We also describe polarization as a set of random transformations of some angle θ, almost all of which lead θ to 0 or $\pi/2$. Section VII addresses fast polarization techniques. Section VIII describes simple ordering of decoding channels, which yields construction complexity $O(n \log n)$ for any polar code.

II. RECURSIVE PLOTKIN CONSTRUCTION

RM codes and polar codes can be designed using polynomial constructions. Consider any boolean polynomial $f(x) \equiv f(x_1, \ldots, x_m)$ for any $x \in \mathbb{F}_2^m$. For any sequence (*e.g. path or channel*) $\xi = (a_1, \ldots, a_m) \in \mathbb{F}_2^m$, we define the monomial

$$x^\xi \equiv x_1^{a_1} \cdot \ldots \cdot x_m^{a_m}$$

Then any polynomial $f(x)$ is decomposed as follows

$$f(x) = \sum_{a_1 = 0,1} x_1^{a_1} f_{a_1}(x_2, \ldots, x_m) = \ldots = \sum_{a_1, \ldots, a_\ell} x_1^{a_1} \cdot \ldots \cdot x_\ell^{a_\ell}$$

$$\cdot f_{a_1, \ldots, a_\ell}(x_{\ell+1}, \ldots, x_m) = \ldots = \sum_\xi f_\xi x^\xi \qquad (1)$$

Any step $\ell = 1, \ldots, m - 1$ ends with the incomplete paths $\xi(\ell) \equiv (a_1, \ldots, a_\ell)$ that decompose the polynomial $f(x)$ with respect to monomials $x_1^{a_1} \cdot \ldots \cdot x_\ell^{a_\ell}$. Finally, step m defines each bit f_ξ associated with a monomial x^ξ.

Codes $RM(r, m)$ consist of the maps $f(x) : \mathbb{F}_2^m \rightarrow \mathbb{F}_2$, where we evaluate all polynomials $f(x)$ of degree r on all positions $x \in \mathbb{F}_2^m$. Each map generates a codeword

$$\mathbf{c} = \mathbf{c}(f) = \sum_\xi f_\xi \, \mathbf{c}\left(x^\xi\right).$$

Here any vector $\mathbf{c}(x^\xi)$ is generated by a monomial x^ξ and has weight $2^{m-w(\xi)}$, where $w(\xi)$ is the Hamming weight of ξ. Note that for $a_1 = 0, 1$, two polynomials $x_1^{a_1} f_{a_1}(x_2, \ldots, x_m)$ generate the codewords $(\mathbf{c}_0, \mathbf{c}_0)$ and $(\mathbf{0}, \mathbf{c}_1)$ formed by two RM codes $\{\mathbf{c}_0\}$ and $\{\mathbf{c}_1\}$ of length 2^{m-1}. Then the codewords $\mathbf{c} = \mathbf{c}_0, \mathbf{c}_0 + \mathbf{c}_1$ of the code $RM(r, m)$ form the Plotkin $\mathbf{c} = (\mathbf{u}, \mathbf{u} + \mathbf{v})$ construction. Similarly, each step $\ell = 2, \ldots, m - 1$ decomposes codewords $\mathbf{c}_0, \mathbf{c}_0 + \mathbf{c}_1$ into RM codes of length $2^{m-\ell}$. This construction also yields the Arikan's 2×2 kernel [5]. Decomposition (1) is shown in Fig. 1 for the code $RM(4, 4)$. Each decomposition step $\ell = 1, \ldots, 4$ is marked by the splitting monomial $x_\ell^{a_\ell}$. For example, path $\xi = 0110$ gives the information bit f_{0110} associated with the monomial $x^\xi \equiv x_2 x_3$.
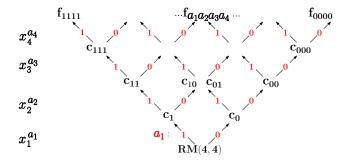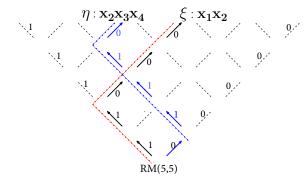
Fig. 1. Decomposition of $RM(4,4)$

Fig. 2. Subcode $C(m,T)$ of code RM(5,5)

Now consider some subset of k paths

$$T = \{\xi^{(\tau)}, \ \tau = 1, ..., k\} \subset \mathbb{F}_2^m$$

Then we encode k information bits via their paths and obtain codewords $\mathbf{c}(T) = \sum_{\xi \in T} \mathbf{c}(x^\xi)$. These codewords form a linear code $C(m,T)$.

Fig. 2 presents such a code $C(m,T)$. Here we use all paths ξ' bounded on the left by the path $\xi = 11000$ and all paths η' bounded by the path $\eta = 01110$. These two paths generate monomials x_1x_2 and $x_2x_3x_4$. In turn, code $C(m,T)$ is the sum of the codes generated by the boundaries ξ and η.

Construction $C(m,T)$ also leads to polar codes, which use subsets $T \subset \mathbb{F}_2^m$ optimized for the recursive SC decoding. This algorithm is considered in the next section.

III. SC DECODING

Recursive decoding of the Plotkin construction. Below, we consider transmission over a discrete memoryless channel W with inputs ± 1. To do so, we map a binary input $a = 0, 1$ onto the symbols $(-1)^a$. In particular, all-zero codeword 0^n is mapped onto 1^n. The Plotkin construction with symbols ± 1 has the form of $\mathbf{c} = (\mathbf{u}, \mathbf{uv})$, where vector $\mathbf{uv}$ is the component-wise product of vectors $\mathbf{u}$ and $\mathbf{v}$. For any received symbol y, let $q = q(y)$ be the posterior probability (PP) that a symbol $c = 1$ is transmitted. Then we define two interrelated quantities, the offset g and the likelihood h :

$$q = q(y) = \Pr\{c = 1 \mid y\}$$
$$g = 2q - 1, \ h = q/(1 - q) \quad (2)$$

For example, let W be a binary symmetric channel BSC(ε) with transition error probability $p = (1 - \varepsilon)/2$, where $\varepsilon \in [0, 1]$. Then any output $y = \pm 1$ gives quantities

$$g(y) = \varepsilon y, \quad h(y) = (1 + \varepsilon y)/(1 - \varepsilon y). \quad (3)$$

Let $\mathbf{c} = (c_j)$ be any code vector and $\mathbf{y} = (y_j)$ be the received vector corrupted by noise. We then use vectors $\mathbf{q} = (q_j)$, $\mathbf{g} = (g_j)$ and $\mathbf{h} = (h_j)$ with symbols defined in (2). Also, for vectors of an even length n, let $i \leq n/2$ and $i' = i + n/2$ be the two matching positions in the left and right halves.

The following recursive algorithm of [1] is identical to the conventional decoder of [5], and performs SC decoding of any recursive $(\mathbf{u}, \mathbf{uv})$ construction. We first wish to evaluate vector $\mathbf{v}$ of length $n/2$ in the $(\mathbf{u}, \mathbf{uv})$ construction. To do so, we find PP

$$q(v_i) \equiv \Pr\{v_i = 1 \mid q_i, q_{i'}\}$$

using PP q_i, $q_{i'}$ of symbols u_i and $u_{i'}v_{i'}$. Namely, it can be readily verified that the corresponding offset $g(v_i) = 2q(v_i) - 1$ can be calculated as

$$g(v_i) = g_i g_{i'} \quad (4)$$

We now decode vector $\mathbf{q}(v)$ with symbols $q(v_i) = (1 + g(v_i))/2$ into some vector $\widetilde{\mathbf{v}} \in RM(r - 1, m - 1)$ of length $n/2$.

Given vector $\widetilde{\mathbf{v}}$, note that two symbols y_i and $y_{i'}\widetilde{v}_i$ represent two corrupted versions of symbol u_i in the $(\mathbf{u}, \mathbf{uv})$ construction. Then symbol u_i has likelihoods h_i and $\widetilde{h}_{i'} = (h_{i'})^{\widetilde{v}_i}$ in the left and right halves. This gives its overall likelihood

$$h(u_i) = h_i \widetilde{h}_{i'} \quad (5)$$

Vector $\mathbf{q}(u)$ with symbols $q(u_i) = h(u_i)/(1 + h(u_i))$ is then decoded into some vector $\widetilde{\mathbf{u}} \in RM(r, m - 1)$. We will see that recalculations (4) degrade the original channel, whereas recalculations (5) upgrade it.

In the general setting, recalculations (4) and (5) form the level $\ell = 1$ of SC decoding. Recall, that vectors $\mathbf{u}$ and $\mathbf{v}$ correspond to two paths $a_1 = 0, 1$ in Fig. 1. We then slightly change our notation and use vectors $\mathbf{q}(1) = \mathbf{q}(v)$ and $\mathbf{q}(0) = \mathbf{q}(u)$, which represent the corrupted versions of vectors $\mathbf{v}$ and $\mathbf{u}$. We proceed similarly at any level $\ell = 2, ..., m$. Any current path $\xi = \xi(\ell)$ receives a vector $\mathbf{q}(\xi)$ of length $\mu = 2^{m-\ell}$ that consists of PP $q_j(\xi)$. Then we derive the $\mathbf{v}$-extension $(\xi, 1)$ using recalculations (4) on the two halves of vector $\mathbf{g}(\xi)$:

$$g_i(\xi, 1) = g_i(\xi) g_{i'}(\xi) \quad (6)$$

Note that each path $(\xi, 1)$ returns its output $\widetilde{\mathbf{v}} = \widetilde{\mathbf{v}}(\xi)$ to the node ξ. We use recalculations (5) with likelihoods $h_i(\xi)$ and $\widetilde{h}_{i'}(\xi) = [h_{i'}(\xi)]^{\widetilde{v}_i}$ for the $\mathbf{u}$-extension

$$h_i(\xi, 0) = h_i(\xi)\widetilde{h}_{i'}(\xi) \quad (7)$$

Then vector $\mathbf{q}(\xi, 0)$ can be decoded into some vector $\widetilde{\mathbf{u}}(\xi)$. Thus, the $\mathbf{v}$-extensions (marked with 1s on Fig. 1) always precede the $\mathbf{u}$-extensions in each decoding step.

Finally, the last step gives the likelihood $q(\xi) = \Pr\{f_\xi = 0 \mid \mathbf{y}\}$ of one information bit f_ξ associated with a path ξ. We then choose the more reliable bit f_ξ. Thus, the decoder recursively retrieves every information symbol f_ξ moving back

and forth along the paths of Fig. 1 or Fig. 2. It is easy to verify [1] that the overall complexity has the order of $n \log n$.

Recursive decoding of polar codes. Any subcode $C(m,T) \subset RM(m,m)$ with k paths $\xi^{(1)}, ..., \xi^{(k)}$ is decoded similarly. Here we simply drop all frozen paths $\xi \notin T$ that give information bits $f_\xi \equiv 0$. This gives the following algorithm.

Algorithm $\Psi(m,T)$ for code $C(m,T)$.

Given: a vector $\mathbf{q} = (q_j)$ of PP.

Take $\tau = 1, ..., k$ and $\ell = 1, ..., m$.

For a path $\xi^{(\tau)} = a_1^{(\tau)}, ..., a_m^{(\tau)}$ in step ℓ do:

  Apply recalculations (6) if $a_\ell^{(\tau)} = 1$

  Apply recalculations (7) if $a_\ell^{(\tau)} = 0$.

  Output information bit $f_{\xi^{(\tau)}}$ if $\ell = m$.

SC decoding can include list decoding [2] that tracks L most probable code candidates throughout the process. SC list decoding has complexity order of $Ln \log n$. Simulation results of [2]-[4] show that bit-frozen subcodes substantially outperform original RM codes in SC list decoding and also require much smaller lists for a similar performance. SC list decoding can also be combined with precoding techniques, which can further reduce the output BERs, as shown in [11].

IV. CHANNEL VARIABLES

Consider a code $C(m,T) = C(m,\xi)$ defined by a single path $\xi = (a_1, ..., a_m)$ and let it be used over a discrete memoryless symmetric (DMS) channel W. We now consider a codeword $\mathbf{1}^n$ transmitted over this path and assume that all preceding (frozen) paths η give correct outputs $\widetilde{\mathbf{v}}^{(\eta)} = \mathbf{1}$ in recursive recalculations (6) and (7). Then recalculations (7) are simplified for every prefix $\xi = (a_1, ..., a_\ell)$ as

$$h_i(\xi, 0) = h_i(\xi) h_{i'}(\xi) \tag{8}$$

Recalculations (6) and (8) essentially form a new DMS channel $W_\xi : X \to Y_\xi$ that outputs a rv $h(\xi)$ or $g(\xi)$ starting from the original rv g_j or h_j. Following [12], we consider the compound channel W_ξ as an ensemble of some number ω of binary symmetric channels $W_\xi(t) = BSC(\beta_t, \varepsilon_t)$ that have transition error probabilities $p_t = (1 - \varepsilon_t)/2$ and occur with some probability distribution $\{\beta_t\}$. We use notation

$$W_\xi = \bigcup_{t=1}^{\omega} BSC(\beta_t, \varepsilon_t), \ \sum_{t=1}^{\omega} \beta_t = 1$$

Here the new parameters ω, ε_t and β_t depend on a specific path ξ. We now introduce the expectations and the second moments of the offsets ε_t over the distribution $\{\beta_t\}$:

$$\mathbb{E}(\varepsilon_t) = \sum_{t=1}^{\omega} \beta_t \varepsilon_t$$

$$\mathbb{E}(\varepsilon_t^2) = \sum_{t=1}^{\omega} \beta_t \varepsilon_t^2 \triangleq \mathcal{A}_\xi$$

As noted by E. Arikan [13], parameter $\mathbb{E}(\varepsilon_t)$ is also studied in statistics as the variational distance [14]. Next, for any channel W_ξ, we consider the Bhattacharyya parameter [5]:

$$\sum_{y \in Y_\xi} \sqrt{W_\xi(y|0)} \sqrt{W_\xi(y|1)}$$

For a binary channel $BSC(\beta_t, \varepsilon_t)$, this parameter is

$$z_t = (1 + \varepsilon_t)^{1/2} (1 - \varepsilon_t)^{1/2} = \left(1 - \varepsilon_t^2\right)^{1/2}$$

Note also that $BSC(\beta_t, \varepsilon_t)$ yields only two values of likelihood $h = (1 \pm \varepsilon_t)/(1 \mp \varepsilon_t)$. Then parameter z_t gives the expectation $\mathbb{E} h^{-1/2}$ of the quantity $h^{-1/2}$:

$$z_t = \left(\frac{1 + \varepsilon_t}{1 - \varepsilon_t}\right)^{1/2} \left(\frac{1 - \varepsilon_t}{2}\right) + \left(\frac{1 - \varepsilon_t}{1 + \varepsilon_t}\right)^{1/2} \left(\frac{1 + \varepsilon_t}{2}\right)$$

From now on, we represent the Bhattacharyya parameter of the compound channel W_ξ as the expectation $\mathbb{E}(z_t)$. We also consider the second moment $\mathbb{E}(z_t^2)$:

$$\mathbb{E}(z_t) = \sum_t \beta_t \sqrt{1 - \varepsilon_t^2} = \mathcal{Z}_\xi \tag{9}$$

$$\mathbb{E}(z_t^2) = \sum_t \beta_t \left(1 - \varepsilon_t^2\right) \triangleq \mathcal{B}_\xi = 1 - \mathcal{A}_\xi$$

Then we can use Markov inequality to bound the output error probability P_ξ of any path ξ :

$$P_\xi = \Pr\{h_\xi^{-1/2} > 1\} \leq \mathcal{Z}_\xi$$

$$P_\xi = \Pr\{h_\xi^{-1} > 1\} \leq \mathcal{B}_\xi \tag{10}$$

Note that $\mathcal{B}_\xi \leq \mathcal{Z}_\xi$, since $z_t \leq 1$. Some other inequalities are also given in Appendix.

Next, consider an ensemble of 2^ℓ equiprobable paths $\xi = (a_1, ..., a_\ell)$. Our main goal is to prove that for $\ell \to \infty$, most paths ξ (with the exception of a vanishing fraction) achieve polarization, in which case we have two possible options:

$$\begin{aligned} (\mathcal{A}_\xi, \mathcal{B}_\xi) &\to (0, 1) \\ (\mathcal{A}_\xi, \mathcal{B}_\xi) &\to (1, 0) \end{aligned} \tag{11}$$

To prove this, we introduce the function

$$\mathcal{V}_\xi = \sqrt{\mathcal{A}_\xi \mathcal{B}_\xi} = \sqrt{\mathcal{A}_\xi (1 - \mathcal{A}_\xi)} = \sqrt{\mathcal{B}_\xi (1 - \mathcal{B}_\xi)} \tag{12}$$

Lemma 1: For any channel W_ξ, asymptotic equalities (11) hold if and only if (iff) $\mathcal{V}_\xi \to 0$.

Proof. Indeed, $\mathcal{V}_\xi \to 0$ iff $\mathcal{A}_\xi \to 0$ (then $\mathcal{B}_\xi \to 1$) or vice versa (then $\mathcal{A}_\xi \to 1$). $\square$

V. CHANNEL TRANSFORMATIONS

Recall from (3) that the $\mathbf{v}$-extension of any channel $BSC(\beta_t, \varepsilon_t)$ yields the pairwise products (6) of the offsets $g(y) = y\varepsilon_t$. Thus, any degrading channel $W_{\xi,1} : X \to Y_{\xi,1}$ can be considered as the ensemble of the new BSC channels:

$$W_{\xi,1} = \bigcup_{t,s} BSC(\beta_{t,s}, \varepsilon_t \varepsilon_s), \ \beta_{t,s} = \beta_t \beta_s \tag{13}$$

where $t, s = 1, ..., \omega$. Then

$$\mathcal{A}_{\xi,1} = \mathbb{E}(\varepsilon_t^2 \varepsilon_s^2) = \sum_{t,s} \beta_{t,s} (\varepsilon_t \varepsilon_s)^2 = \mathcal{A}_\xi^2 \tag{14}$$

For a $BSC(\beta_t, \varepsilon_t)$ with $z_t = \sqrt{1 - \varepsilon_t^2}$, we will also use notation $BSC(\beta_t \smallsetminus z_t)$. Similarly to (13), we see from (8) that the upgrading channel $W_{\xi,0} : X \to Y_{\xi,0}$ forms the ensemble

$$W_{\xi,0} = \bigcup_{t,s} BSC(\beta_{t,s} \smallsetminus z_t z_s)$$

This gives equality

$$\mathcal{B}_{\xi,0} = \mathbb{E}(z_t^2 z_s^2) = \sum_{t,s} \beta_{t,s} (z_t z_s)^2 = \mathcal{B}_\xi^2 \qquad (15)$$

Now let $r(x) = \sqrt{x(1-x)}$ for $x \in [0,1]$.

Lemma 2: For any channel $W_\xi = \bigcup_{t=1}^\omega BSC(\beta_t, \varepsilon_t)$, its extensions $W_{\xi,1}$ and $W_{\xi,0}$ satisfy equalities

$$\mathcal{V}_{\xi,1} = \sqrt{\mathcal{A}_{\xi,1}\mathcal{B}_{\xi,1}} = \mathcal{A}_\xi\sqrt{1-\mathcal{A}_\xi^2} = r(\mathcal{A}_\xi^2)$$

$$\mathcal{V}_{\xi,0} = \sqrt{\mathcal{A}_{\xi,0}\mathcal{B}_{\xi,0}} = \mathcal{B}_\xi\sqrt{1-\mathcal{B}_\xi^2} = r(\mathcal{B}_\xi^2)$$

Proof. Substitute equalities (14) and (15) in definition (12). □

VI. Proof of polarization property

Consider the ensemble $\mathcal{N}$ of equiprobable paths $\xi = (a_1,...,a_\ell)$. For each ξ, the bit $a_{\ell+1}$ takes values 0 and 1 equally likely. Then for any given path ξ, its extension $\xi, a_{\ell+1}$ yields the random variable $\mathcal{V}_{\xi,a_{\ell+1}}$ with the expected value

$$\mathbb{E}\,\mathcal{V}_{\xi,a_{\ell+1}} = (\mathcal{V}_{\xi,0} + \mathcal{V}_{\xi,1})/2$$

For brevity, let $\mathcal{B}_\xi = x \in [0,1]$ and $\mathcal{A}_\xi = 1-x$. Given any path ξ with parameters $\mathcal{A}_\xi$ and $\mathcal{B}_\xi$, consider the ratio

$$R(x) = \frac{\mathbb{E}\,\mathcal{V}_{\xi,a_{\ell+1}}}{\mathcal{V}_\xi} = \frac{r(x^2) + r[(1-x)^2]}{2r(x)}$$

The following theorem proves the polarization property.

Theorem 1: For the ensemble of paths $\xi = (a_1,...,a_\ell)$ of length ℓ, random functions $\mathcal{V}_\xi$ have the expected value

$$\mathbb{E}\,\mathcal{V}_{\xi(\ell)} \le \left(\sqrt{3}/2\right)^\ell \qquad (16)$$

Proof. Function $R(x)$ is plotted below. It is easy to verify that $R(x) \le \sqrt{3}/2$ for any $x \in [0,1]$. Thus, function $\mathcal{V}_\xi$ satisfies (16) on the paths of length ℓ. □
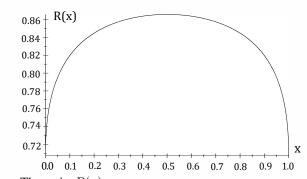

Fig. 3. The ratio $R(x)$

Corollary 1: Most paths $\xi(\ell) = (a_1,...,a_\ell)$, except the fraction $\left(\sqrt{3}/2\right)^{\ell/2}$ of them, satisfy inequality

$$\mathcal{V}_{\xi(\ell)} < \left(\sqrt{3}/2\right)^{\ell/2} \qquad (17)$$

Geometric interpretation. Let $\mathcal{A}_\xi = \cos^2\theta$ and $\mathcal{B}_\xi = \sin^2\theta$ for some angle $\theta \in [0, \pi/2]$. Then one-bit extensions of a path ξ give parameters $\mathcal{B}_{\xi,0} = \sin^4\theta$ and $\mathcal{A}_{\xi,1} = \cos^4\theta$. Thus, angle θ is equally likely transformed into one of two angles:

$$\begin{aligned} \theta^{(0)} &= \arcsin\left(\sin^4\theta\right) & \text{if } a_{\ell+1}=0 \\ \theta^{(1)} &= \arccos\left(\cos^4\theta\right) & \text{if } a_{\ell+1}=1 \end{aligned} \qquad (18)$$

Then Theorem 1 shows that angles $\theta(\xi)$ tend to 0 or $\pi/2$ for most sequences ξ of $m \to \infty$ transformations (18).

Remark. Function $\mathcal{V}_\xi(\lambda) = (\mathcal{A}_\xi\mathcal{B}_\xi)^\lambda$ tightens (16) to $\mathbb{E}\,\mathcal{V}_\xi(\lambda) \le c^{\ell/2}$, where we take any $c > 1/2$ as $\lambda \to 0$.

VII. Fast polarization

Theorem 2 below uses elementary recalculations to simplify some known results [8]-[10] on fast polarization.

Theorem 2: For $m \to \infty$, sequences $\xi \in \mathbb{F}_2^m$, except their fraction less than $\exp\{-m^{1/4}\}$, yield channel polarization parameters $\mathcal{A}_\xi$ and $\mathcal{B}_\xi$ such that

$$\log_2\,(\mathcal{A}_\xi\mathcal{B}_\xi) < -2^{m/2-\theta(m)}, \ \theta(m) = O(m^{3/4}) \qquad (19)$$

Proof. Let $\lambda = m^{3/4}$, $\delta = m^{1/2}$, $s = (\lambda-\delta)/2$ and $\rho = (\lambda+\delta)/2$ be integer parameters. For any path $\xi = (a_1,...,a_m)$, let $u_0 = (a_1,...,a_\ell)$ be its initial segment of length $\ell = 5\lambda\ln\lambda$ and $u_i = (a_{\ell+(i-1)\lambda+1},...,a_{\ell+i\lambda})$ be an ith segment of length λ for all $i = 1,...,(m-\ell)/\lambda$. Consider the set Θ of paths such that for every path ξ:

(A) each segment u_i, $i \ge 1$, has weight $s \le w_i \le \rho$.

Note that Θ contains most paths ξ in $\mathbb{F}_2^m$. Indeed, $\left(\sqrt{3}/2\right)^5 < 1/2$. Then condition (17) fails on the segment u_0 with probability less than $2^{-m^{3/4}}$, while condition (A) fails on any (one or more) segment u_i with probability less than $\exp\{-m^{1/4}\}$. We also replace inequality (17) with two (weaker) asymptotic cases

$$\mathcal{A}_{\xi(\ell)} < 2^{-\lambda\ln\lambda}, \ \mathcal{B}_{\xi(\ell)} > 1-2^{-\lambda\ln\lambda}$$

$$\mathcal{B}_{\xi(\ell)} < 2^{-\lambda\ln\lambda}, \ \mathcal{A}_{\xi(\ell)} > 1-2^{-\lambda\ln\lambda}$$

Consider the first case. We now estimate the rate of decline of parameter $\mathcal{A}_\xi$ for a path ξ of any length μ. Here (14) and (15) give parameters

$$\mathcal{A}_{\xi,0} = 1-\mathcal{B}_\xi^2 \le 2\mathcal{A}_\xi, \ \mathcal{A}_{\xi,1} = \mathcal{A}_\xi^2$$

which at most double on the extension $(\xi,0)$ but decline exponentially $\mathcal{A}_\xi$ times on $(\xi,1)$. Thus, $\mathcal{A}_\xi$ achieves its maximum on the first segment u_1 if $u_1 = (0^\rho, 1^s)$. Indeed, then $\mathcal{A}_{\xi(\ell)}$ undergoes the maximum possible increase of order 2^ρ on the sequence 0^ρ and gets the lowest decline possible on 1^s. Next, note that $(1-x)^t \ge 1-xt$ for any integer $t \ge 1$ and any $x < 1/t$. This can be verified by expanding the term $(1-x)^t$. Also, $\rho < (\lambda\ln\lambda)/2$. Then for $u_1 = (0^\rho, 1^s)$, the above arguments yield

$$\mathcal{B}_{\xi(\ell+\rho)} = \left[\mathcal{B}_{\xi(\ell)}\right]^{2^\rho} > 1-2^{\rho-\lambda\ln\lambda}$$

$$\mathcal{A}_{\xi(\ell+\lambda)} < 2^{[\rho-\lambda\ln\lambda]2^s} < 2^{-2^{s-1}\lambda\ln\lambda}$$

We now use recursion and assume that the sequence $\xi(\mu) = u_0,...,u_i$ of length μ gives

$$\mathcal{A}_{\xi(\mu)} < 2^{-2^{i(s-1)}\lambda\ln\lambda}$$

Then adding $u_{i+1} = (0^\rho, 1^s)$ gives our recursive estimate :

$$\mathcal{B}_{\xi(\mu+\rho)} = \left[\mathcal{B}_{\xi(\mu)}\right]^{2^\rho} > 1-2^{\rho-2^{i(s-1)}\lambda\ln\lambda}$$

$$\mathcal{A}_{\xi(\mu+\lambda)} < 2^{[\rho-2^{i(s-1)}\lambda\ln\lambda]2^s} < 2^{-2^{(i+1)(s-1)}\lambda\ln\lambda}$$

Finally, the last segment $i = (m - \ell) / \lambda$ gives

$$i(s - 1) \geq \left(\frac{m - \ell}{\lambda}\right)\left(\frac{\lambda - \delta}{2}\right) - i = \frac{m}{2} - O(m^{3/4})$$

This leads to (19). The second case is identical. $\square$

VIII. SIMPLE ORDERING OF PATHS

Power moments $\mathcal{A}_\xi = \mathbb{E}(\varepsilon_t^2)$ and $\mathcal{B}_\xi = \mathbb{E}(z_t^2)$ offer one clear advantage in the analysis of polarization process. Here we use exact equalities in channel transformations (14) and (15), unlike inequalities for the Bhattacharyya parameter $\mathcal{Z}_\xi$. Also, we can use $\mathcal{B}_\xi$ to estimate error rate P_ξ in (10). One important consequence is that we can void all recalculations of probability distributions on each path $\xi = (a_1, ..., a_\ell)$. Instead, for any ℓ we use simple recursive recalculations:

$$\mathcal{B}_{\xi,0} = \mathcal{B}_\xi^2, \quad \mathcal{B}_{\xi,1} = 1 - (1 - \mathcal{B}_\xi)^2 \qquad (20)$$

Corollary 2: Polar codes of length n have construction complexity $O(n \log n)$ with respect to the moments $\mathcal{B}_\xi$.
Proof. For each path ξ of any polar code, moments $\mathcal{A}_\xi$ and $\mathcal{B}_\xi$ can be calculated with complexity $O(m)$. Then all paths can be ordered with complexity $O(nm)$. $\square$

Next we consider some examples of path (channel) ordering with respect to equalities (20).

1. "Swap-bits-upgrade" (SBU). Consider two paths $\xi = (01, ..., a_m)$ and $\eta = (10, ..., a_m)$ with the first two bits swapped. Then $\mathcal{B}_\xi \leq \mathcal{B}_\eta$. Indeed, let x be the original channel moment $\mathcal{B}$. For the paths $\xi' = 01$ and $\eta' = 10$, recalculations (20) give

$$\mathcal{B}_{\xi'} = 1 - (1 - x^2)^2, \quad \mathcal{B}_{\eta'} = \left[1 - (1 - x)^2\right]^2$$

Both functions are plotted in Fig. 4. Inequality $\mathcal{B}_{\xi'} \leq \mathcal{B}_{\eta'}$ yields $\mathcal{B}_\xi \leq \mathcal{B}_\eta$ for any $\varepsilon \in (0, 1)$ due to the monotonic behavior of both transformations (20).

2. "Center-move upgrade" (CMU). Now let $\xi = (0110, ..., a_m)$ and $\eta = (1001, ..., a_m)$. Straightforward calculations show that prefixes $\xi' = 0110$ and $\eta' = 1001$ yield functions $\mathcal{B}_{\xi'}$ and $\mathcal{B}_{\eta'}$ that are similar to those plotted in Fig. 4. Thus, $\mathcal{B}_\xi < \mathcal{B}_\eta$ for any $\varepsilon \in (0, 1)$.
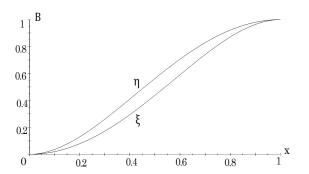

Fig. 4. The moments $\mathcal{B}_\xi$ and $\mathcal{B}_\eta$

3. We combine the previous examples and consider $\xi' = (10)(01)(01)$ and $\eta' = (01)(10)(10)$. Here ξ' undergoes SBU in two center bits but reverses CMU in the outer bits of

η'. Calculations again show a similar behavior with $\mathcal{B}_{\xi'} \leq \mathcal{B}_{\eta'}$ for any $\varepsilon \in (0, 1)$.

More generally, we raise

Open problem. Do all paths ξ of any given weight $w = w(\xi)$ have permanent ordering on a $BSC(\varepsilon)$ for any $\varepsilon \in (0, 1)$?

Appendix.

Lemma 3: For any channel W_ξ, the Bhattacharyya parameter $\mathcal{Z}_\xi$ satisfies inequalities

$$1 - \mathbb{E}(\varepsilon_t) \leq \mathcal{Z}_\xi \leq \sqrt{1 - \mathbb{E}^2(\varepsilon_t)} \qquad (21)$$

$$1 - \mathbb{E}(\varepsilon_t^2) \leq \mathcal{Z}_\xi \leq \sqrt{1 - \mathbb{E}(\varepsilon_t^2)} \qquad (22)$$

Proof. We apply the Jensen inequality for $\mathcal{Z}_\xi = \sum_t \beta_t \sqrt{1 - \varepsilon_t^2}$. Here $\sqrt{1 - \varepsilon_t^2}$ is a concave function of the variables $x = \varepsilon_t$ and $y = \varepsilon_t^2$. Taking $x = \varepsilon_t$ gives the upper bound in (21). Inequality $\sqrt{1 - x^2} \geq 1 - x$ for any $x \in [0, 1]$ gives the lower bound in (21). Taking $y = \varepsilon_t^2$, we obtain the upper bound in (22), while the lower bound in (22) follows from the inequality $\sqrt{1 - y} \geq 1 - y$ for $y \in [0, 1]$. Thus, $\mathcal{B}_\xi \leq \mathcal{Z}_\xi \leq \sqrt{\mathcal{B}_\xi}$, which is identical to (22). $\square$

REFERENCES

[1] I. Dumer, "Recursive decoding of Reed-Muller codes," *Proc. 37th Allerton Conf. on Commun., Cont., and Comp.,* Monticello, IL, USA, 1999, pp. 61-69 (http://arxiv.org/abs/1703.05303).

[2] I. Dumer and K. Shabunov, "Recursive constructions and their maximum likelihood decoding," *Proc. 38th Allerton Conf. on Commun., Cont., and Comp.,* Monticello, IL, USA, 2000, pp. 71-80 (http://arxiv.org/abs/1703.05302).

[3] I. Dumer and K. Shabunov, "Near-optimum decoding for subcodes of Reed-Muller codes," *2001 IEEE Intern. Symp. Info. Theory*, Washington DC, USA, June 24-29, 2001, p. 329.

[4] I. Dumer and K. Shabunov, "Soft decision decoding of Reed-Muller codes: recursive lists," *IEEE Trans. Info. Theory*, vol. 52, no. 3, pp. 1260-1266, 2006.

[5] E. Arikan, "Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Info. Theory,* vol. 55, no. 6, pp. 3051-3073, 2009.

[6] V. Guruswami and P. Xia, "Polar Codes: speed of polarization and polynomial gap to capacity," vol. 61, no. 1, pp. 3-16, 2015.

[7] M. Alsan and E. Telatar, "A simple proof of polarization and polarization for non-stationary memoryless channels," *IEEE Trans. Info. Theory,* vol. 62, no. 9, pp. 4873-4878, 2016.

[8] E. Arıkan and E. Telatar, "On the rate of channel polarization," in Proc. IEEE Intern. Symp. Inform. Theory (ISIT'2009), Seoul, South Korea, 2009, pp. 1493–1495.

[9] S. B. Korada, E. Sasoglu, and R. Urbanke, "Polar codes: characterization of exponent, bounds, and constructions," IEEE Trans. Inform. Theory, vol. 56, no. 12, pp. 6253–6264, 2010.

[10] I. Tal, "A simple proof of fast polarization," available at https://arxiv.org/abs/1704.07179, April 2017.

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

[12] S. B. Korada, "*Polar Codes for Channel and Source Coding,*" Ph.D. thesis, Ecole Polytechnique Federale De Lausanne, 2009.

[13] E. Arikan, Private communication, March 2017.

[14] J. Duchi, "Lecture Notes for Statistics 311/Electrical Engineering 377", Stanford University, 2016, https://stanford.edu/class/stats311/Lectures/full_notes.pdf