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 \( n \log n \).

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 [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 \( m \to \infty \). This technique also shows that capacity-achieving subcodes exist for all codes \( RM(r, m) \) of rate \( R \to 1 \), 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 \( g_1g_2 \) and \( h_1h_2 \) on any degrading and upgrading channel, respectively.

Section [IV] introduces two new parameters, \( A \) and \( B \), which are related to the power moments of rv \( g \) and \( h \). Both are also related to the Bhattacharyya parameter \( Z \). Section [V] shows that parameters \( A \) and \( B \) square up on the upgrading and degrading channels, respectively. Section [VI] validates polarization property of SC decoding, by proving that the product \( AB \) tends to 0 for almost all sequences (except a vanishing fraction) of \( m \to \infty \) channel transformations. We also describe polarization as a set of random transformations of some angle \( \theta \), almost all of which lead \( \theta \) 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 \( (e.g. \text{path or channel}) \) \( \xi = (a_1, \ldots, a_m) \in \mathbb{F}_2^m \), we define the monomial

\[
x^\xi = x_1^{a_1} \cdots x_m^{a_m}
\]

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

\[
f(x) = \sum_{a_1=0}^1 x_1^{a_1} f_{a_1}(x_2, \ldots, x_m) = \cdots = \sum_{a_1, \ldots, a_t} x_1^{a_1} \cdots x_t^{a_t} f_{a_1, \ldots, a_t}(x_{t+1}, \ldots, x_m) = \cdots = \sum_{\xi} f_\xi x^\xi
\]

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 \( x_1^{a_1} \cdots x_\ell^{a_\ell} \). Finally, step \( m \) defines each bit \( f_\xi \) associated with a monomial \( x^\xi \).

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

\[
c = c(f) = \sum_\xi f_\xi c(x^\xi).
\]

Here any vector \( c(x^\xi) \) is generated by a monomial \( x^\xi \) and has weight \( 2^m - w(\xi) \), where \( w(\xi) \) is the Hamming weight of \( \xi \). Note that for \( a_1 = 0, 1 \), two polynomials \( x_1^{a_1} f_{a_1}(x_2, \ldots, x_m) \) generate the codewords \( (c_0, c_0) \) and \( (0, c_1) \) formed by two RM codes \( \{c_0\} \) and \( \{c_1\} \) of length \( 2^{m-1} \). Then the codewords \( c = (c_0, c_0 + c_1) \) of the code \( RM(r, m) \) form the Plotkin \( c = (u, u + v) \) construction. Similarly, each step \( \ell = 2, \ldots, m - 1 \) decomposes codewords \( c_0, c_0 + c_1 \) into RM codes of length \( 2^{m-\ell} \). This construction also yields the Arikan’s 2×2 kernel [5].

Decomposition [I] 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

\[
x^\xi = x_2 x_3.
\]
Recursive decoding of the Plotkin construction. Below, we consider transmission over a discrete memoryless channel \( W \) with inputs \( \pm 1 \). 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 \( \pm 1 \) has the form of \( c = (u, uv) \), where vector \( uv \) is the component-wise product of vectors \( u \) and \( v \). For any received symbol \( y \), let \( q = g(y) \) be the posterior probability (PP) that a symbol \( c = 1 \) is transmitted. Then we define two interrelated quantities, the offset \( g \) and the likelihood \( h \):

\[
q = g(y) = \Pr\{c = 1 \mid y\}
\]

\[
g = 2q - 1, \quad h = q/(1 - q)
\]

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

\[
g(y) = \varepsilon y, \quad h(y) = (1 + \varepsilon y)/(1 - \varepsilon y).
\]

Let \( c = (c_j) \) be any code vector and \( y = (y_j) \) be the received vector corrupted by noise. We then use vectors \( q_i = (q_{ij}) \) and \( h_i = (h_{ij}) \) with symbols defined in Eq. 5. 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 [3], and performs SC decoding of any recursive \((u, uv)\) construction. We first wish to evaluate vector \( v \) of length \( n/2 \) in the \((u, uv)\) construction. To do so, we find PP

\[
g(v_i) \equiv \Pr\{v_i = 1 \mid q_i, q_{i'}\}
\]

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

\[
g(v_i) = g_i g_{i'}
\]

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

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

\[
h(u_i) = h_i \tilde{h}_{i'}
\]

Vector \( q(u) \) with symbols \( q(u_i) = h(u_i)/(1 + h(u_i)) \) is then decoded into some vector \( \bar{u} \in RM(r, m - 1) \). 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 \( u \) and \( v \) correspond to two paths \( q_1, 0 \) in Fig. 1. We then slightly change our notation and use vectors \( q(1) = q(v) \) and \( q(0) = q(u) \), which represent the corrupted versions of vectors \( v \) and \( u \). We proceed similarly at any level \( \ell = 2, ..., m \). Any current path \( \xi = \xi(\ell) \) receives a vector \( q(\xi) \) of length \( \mu = 2^{m-\ell} \) that consists of PP \( q_j(\xi) \). Then we derive the \( v \)-extension \((\xi, 1)\) using recalculations [4] on the two halves of vector \( g(\xi) \):

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

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

\[
h_i(\xi, 0) = h_i(\xi) \tilde{h}_{i'}(\xi) \quad (7)
\]

Then vector \( q(\xi, 0) \) can be decoded into some vector \( \bar{u}(\xi) \).

Thus, the \( v \)-extensions (marked with 1s on Fig. 1) always precede the \( u \)-extensions in each decoding step.

Finally, the last step gives the likelihood \( q(\xi) = Pr\{f_\xi = 0 \mid y\} \) of one information bit \( f_\xi \) associated with a path \( \xi \). We then choose the more reliable bit \( f_\xi \). Thus, the decoder recursively retrieves every information symbol \( f_\xi \) 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 $\xi^{(1)}, \ldots, \xi^{(k)}$ 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 $q = (q_i)$ of PP.
Take $\tau = 1, \ldots, k$ and $\ell = 1, \ldots, m$.
For a path $\xi^{(\tau)} = a_1^{(\tau)}, \ldots, a_m^{(\tau)}$ in step $\ell$ 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 [4] show that bit-frozen subcodes substantially outperform original RM codes in SC list decoding and also results of [2]–[4] show that bit-frozen subcodes substantially increase decoding complexity.

### IV. CHANNEL VARIABLES

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

$$h_i(\xi, 0) = h_i(\xi)h_i(\xi)$$

(8)

Recalculations (6) and (8) essentially form a new DMS channel $W_\xi: X \rightarrow Y_\xi$ that outputs a rv $h(\xi)$ or $g(\xi)$ starting from the original rv $g_j$ or $h_\ell$. Following [12], we consider the compound channel $W_\xi$ as an ensemble of some number $\omega$ 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

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

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

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

$$E(\varepsilon_t^2) = \sum_{t=1}^{\omega} \beta_t \varepsilon_t^2 \triangleq A_\xi$$

As noted by E. Arikan [14], parameter $E(\varepsilon_t)$ is also studied in statistics as the variational distance [14]. Next, for any channel $W_\xi$, 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

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

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

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

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

$$E(z_t) = \sum_{t=1}^{\omega} \beta_t \left(1 - \varepsilon_t\right)^2 \triangleq B_\xi = 1 - A_\xi$$

Then we can use Markov inequality to bound the output error probability $P_\xi$ of any path $\xi$:

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

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

(10)

Note that $B_\xi \leq Z_\xi$, since $z_t \leq 1$. Some other inequalities are also given in Appendix.

Next, consider an ensemble of $2^L$ equiprobable paths $\xi = (a_1, \ldots, a_L)$ with $L$ (the exception of a vanishing fraction) achieve polarization, in which case we have two possible options:

$$(A_\xi, B_\xi) \rightarrow (0, 1)$$

$$(A_\xi, B_\xi) \rightarrow (1, 0)$$

(11)

To prove this, we introduce the function

$$\nu_\xi = \sqrt{A_\xi B_\xi} = \sqrt{A_\xi (1 - A_\xi)} = \sqrt{B_\xi (1 - B_\xi)}$$

(12)

**Lemma 1:** For any channel $W_\xi$, asymptotic equalities [11] hold if and only if (iff) $\nu_\xi \rightarrow 0$.

**Proof.** Indeed, $\nu_\xi \rightarrow 0$ if $A_\xi \rightarrow 0$ (then $B_\xi \rightarrow 1$) or vice versa (then $A_\xi \rightarrow 1$). \[ \square \]

### V. CHANNEL TRANSFORMATIONS

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

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

(13)

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

$$A_{\xi,1} = E(\varepsilon_t^2) = \sum_{t,s} \beta_{t,s} (\varepsilon_t \varepsilon_s)^2 = A_\xi^2$$

(14)

For a $BSC(\beta_t, \varepsilon_t)$ with $z_t = \sqrt{1 - \varepsilon_t^2}$, we will also use notation $BSC(\beta_t \varepsilon_t^2)$. Similarly to (13), we see from (9) that the upgrading channel $W_{\xi,0}: X \rightarrow Y_{\xi,0}$ forms the ensemble

$$W_{\xi,0} = \bigcup_{t,s} BSC(\beta_{t,s} \varepsilon_t \varepsilon_s)$$
This gives equality
\[ B_{\xi,0} = \mathbb{E}(z_{1,0}^2 z_{2,0}^2) = \sum_{t,s} \beta_{t,s} (z_t z_s)^2 = B_\xi^2 \] (15)

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

**Lemma 2:** For any channel \( W_\xi = \bigcup_{t=1}^2 BSC(\beta_t, \varepsilon_t) \), its extensions \( W_{\xi,1} \) and \( W_{\xi,0} \) satisfy equalities
\[
V_{\xi,1} = \sqrt{A_{\xi,1} B_{\xi,1}} = A_\xi \sqrt{1 - A_\xi^2} = r(A_\xi^2)
\]
\[
V_{\xi,0} = \sqrt{A_{\xi,0} B_{\xi,0}} = B_\xi \sqrt{1 - B_\xi^2} = r(B_\xi^2)
\]

**Proof.** Substitute equalities (14) and (15) in definition (12). \( \square \)

**VI. PROOF OF POLARIZATION PROPERTY**

Consider the ensemble \( \mathcal{N} \) of equiprobable paths \( \xi = (a_1, \ldots, a_\ell) \). For each \( \xi \), the bit \( a_{\ell+1} \) takes values 0 and 1 equally likely. Then for any given path \( \xi \), its extension \( \xi, a_{\ell+1} \) yields the random variable \( V_{\xi, a_{\ell+1}} \) with the expected value
\[
\mathbb{E} V_{\xi, a_{\ell+1}} = (V_{\xi,0} + V_{\xi,1})/2.
\]

For brevity, let \( B_\xi = \varepsilon \), \( A_\xi = 1 - \varepsilon \). Given any path \( \xi \) with parameters \( A_\xi \) and \( B_\xi \), consider the ratio
\[
R(x) = \frac{\mathbb{E} V_{\xi, a_{\ell+1}}}{V_\xi} = \frac{r(x^2) + r((1-x)^2)}{2r(x)}
\]

The following theorem proves the polarization property.

**Theorem 1:** For the ensemble of paths \( \xi = (a_1, \ldots, a_\ell) \) of length \( \ell \), random functions \( V_\xi \) have the expected value
\[
\mathbb{E} V_{\xi(\ell)} \leq \sqrt{3/2}^\ell
\] (16)

**Proof.** Function \( R(x) \) is plotted below. It is easy to verify that \( R(x) \leq \sqrt{3/2} \) for any \( x \in [0,1] \). Thus, function \( V_\xi \) satisfies (16) on the paths of length \( \ell \). \( \square \)

**Corollary 1:** Most paths \( \xi(\ell) = (a_1, \ldots, a_\ell) \), except the fraction \( \sqrt{3/2}^{-\ell/2} \) of them, satisfy inequality
\[
V_{\xi(\ell)} < \sqrt{3/2}^{\ell/2}
\] (17)

**Geometric interpretation.** Let \( A_\xi = \cos^2 \theta \) and \( B_\xi = \sin^2 \theta \) for some angle \( \theta \in [0, \pi/2] \). Then one-bit extensions of a path \( \xi \) give parameters \( B_{\xi,0} = \sin^2 \theta \) and \( A_{\xi,1} = \cos^2 \theta \). Thus, angle \( \theta \) is equally likely transformed into one of two angles:
\[
\begin{align*}
\theta^{(0)} &= \arcsin(\sin^4 \theta) & \text{if} & a_{\ell+1} = 0 \\
\theta^{(1)} &= \arccos(\cos^4 \theta) & \text{if} & a_{\ell+1} = 1
\end{align*}
\] (18)

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

**Remark.** Function \( V_\xi(\lambda) = (A_\xi B_\xi)^\lambda \) tightens (16) to \( \mathbb{E} V_\xi(\lambda) \leq e^{\lambda^2/2} \), where we take any \( c > 1/2 \) as \( \lambda \to 0 \).

**VII. FAST POLARIZATION**

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

**Theorem 2:** For \( m \to \infty \), sequences \( \xi \in \mathbb{F}_m^\ell \), except their fraction less than \( \exp\{-m^{1/4}\} \), yield channel polarization parameters \( A_\xi \) and \( B_\xi \) such that
\[
\log_2 (A_\xi B_\xi) \leq -2m^{2-\theta(m)} \text{, } \theta(m) = O(m^{3/4})
\] (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 \( \xi = (a_1, \ldots, a_m) \), let \( u_0 = (a_1, \ldots, a_\ell) \) be its initial segment of length \( \ell \geq 5\lambda \ln \lambda \) and \( u_i = (a_{\ell+i-(1)-(\lambda+1)}, \ldots, a_{\ell+i+)}) \) be an \( i \)th segment of length \( \lambda \) for all \( i = 1, \ldots, (m - \ell) / \lambda \). Consider the set \( \Theta \) of paths such that for every path \( \xi \):

(A) each segment \( u_i, i \geq 1 \), has weight \( s \leq w_i \leq \rho \).

Note that \( \Theta \) contains most paths \( \xi \) in \( \mathbb{F}_m^\ell \). Indeed, \( (\sqrt{3/2})^\delta < 1/2 \). 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
\[
A_{\xi(\ell)} < 2^{-\lambda \ln \lambda}, \ B_{\xi(\ell)} > 1 - 2^{-\lambda \ln \lambda}
\]
\[
A_{\xi(\ell)} < 2^{-\lambda \ln \lambda}, \ A_{\xi(\ell)} > 1 - 2^{-\lambda \ln \lambda}
\]

Consider the first case. We now estimate the rate of decline of parameter \( A_\xi \) for a path \( \xi \) of any length \( \mu \). Here (14) and (15) give parameters
\[
A_{\xi,0} = 1 - B_\xi^2 \leq 2A_\xi, \ A_{\xi,1} = A_\xi^2
\]
which at most double on the extension \( (\xi,0) \) but decline exponentially \( A_\xi \) times on \( (\xi,1) \). Thus, \( A_\xi \) achieves its maximum on the first segment \( u_1 \) if \( u_1 = (0^\rho, 1^s) \). Indeed, then \( A_{\xi(\ell)} \) undergoes the maximum possible increase of order \( 2^\rho \) on the sequence \( 0^\rho \) and gets the lowest decline possible on \( 1^s \). Next, note that \( (1-x)^t \geq 1 - xt \) for any integer \( t \geq 1 \) 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
\[
B_{\xi(\ell+\rho)} > [B_{\xi(\ell)}]^2 > 1 - 2^{-\rho \ln \lambda} A_{\xi(\ell+\lambda)} < 2^{-\rho \ln \lambda} [2^s < 2^{-2^{-1} \lambda \ln \lambda}
\]

We now use recursion and assume that the sequence \( \xi(\mu) = u_0, \ldots, u_\ell \) of length \( \mu \) gives
\[
A_{\xi(\mu)} < 2^{-2^{\ell-1} \lambda \ln \lambda}
\]

Then adding \( u_{\ell+1} = (0^\rho, 1^s) \) gives our recursive estimate:
\[
B_{\xi(\mu+\rho)} > [B_{\xi(\mu)}]^2 > 1 - 2^{-2^{\ell-1} \lambda \ln \lambda} A_{\xi(\mu+\lambda)} < 2^{-2^{\ell-1} \lambda \ln \lambda} [2^s < 2^{-2^{\ell-1} \lambda \ln \lambda}
\]

**Fig. 3.** The ratio \( R(x) \)
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. □

VIII. Simple ordering of paths

Power moments \( A_\xi = \mathbb{E}(\varepsilon_\xi^2) \) and \( B_\xi = \mathbb{E}(\eta_\xi^2) \) 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 \( Z_\xi \). Also, we can use \( B_\xi \) to estimate error rate \( P_\xi \) in (10). One important consequence is that we can void all recalculations of probability distributions on each path \( \xi = (a_1, \ldots, a_\ell) \). Instead, for any \( \ell \) we use simple recursive recalculations:
\[
B_{\xi,0} = B_{\xi}^2, \quad B_{\xi,1} = 1 - (1 - B_{\xi})^2 \tag{20}
\]

Corollary 2: Polar codes of length \( n \) have construction complexity \( O(n \log n) \) with respect to the moments \( B_\xi \).

Proof. For each path \( \xi \) of any polar code, moments \( A_\xi \) and \( B_\xi \) can be calculated with complexity \( O(m) \). Then all paths can be ordered with complexity \( O(nm) \). □

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

1. “Swap-bits-upgrade” (SBU). Consider two paths \( \xi = (0_1, \ldots, a_m) \) and \( \eta = (1_1, \ldots, a_m) \) with the first two bits swapped. Then \( B_\xi \leq B_\eta \). Indeed, let \( x \) be the original channel moment \( B \). For the paths \( \xi' = 01 \) and \( \eta' = 10 \), recalculations (20) give
\[
B_\xi' = 1 - (1 - x^2)^2, \quad B_\eta' = \left[ 1 - (1 - x^2)^2 \right]^2
\]
Both functions are plotted in Fig. 4. Inequality \( B_\xi' \leq B_\eta' \) yields \( B_\xi \leq B_\eta \) for any \( \varepsilon \in (0, 1) \) due to the monotonic behavior of both transformations (20).

2. “Center-move upgrade” (CMU). Now let \( \xi = (0_110, \ldots, a_m) \) and \( \eta = (1_001, \ldots, a_m) \). Straightforward calculations show that prefixes \( \xi' = 0110 \) and \( \eta' = 1001 \) yield functions \( B_\xi ' \) and \( B_\eta ' \) that are similar to those plotted in Fig. 4. Thus, \( B_\xi ' < B_\eta ' \) for any \( \varepsilon \in (0, 1) \).

3. We combine the previous examples and consider \( \xi' = (10) (01) (01) \) and \( \eta' = (01) (10) (10) \). Here \( \xi' \) undergoes SBU in two center bits but reverses CMU in the outer bits of \( \eta' \). Calculations again show a similar behavior with \( B_\xi ' \leq B_\eta ' \) for any \( \varepsilon \in (0, 1) \).

More generally, we raise

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

Acknowledgment. The author is grateful to E. Arikan, B. Shuval and I. Tal for helpful comments.

Appendix.

Lemma 3: For any channel \( W_\xi \), the Bhattacharyya parameter \( Z_\xi \) satisfies inequalities
\[
1 - \mathbb{E}(\varepsilon_\xi) \leq Z_\xi \leq \sqrt{1 - \mathbb{E}^2(\varepsilon_\xi)} \tag{21}
\]
\[
1 - \mathbb{E}(\varepsilon_\eta^2) \leq Z_\xi \leq \sqrt{1 - \mathbb{E}(\varepsilon_\eta^2)} \tag{22}
\]

Proof. We apply the Jensen inequality for \( Z_\xi = \sum \beta_i \sqrt{1 - \varepsilon_i^2} \). Here \( \sqrt{1 - \varepsilon_i^2} \) is a concave function of the variables \( x = \varepsilon_i \) and \( y = \varepsilon_i^2 \). Taking \( x = \varepsilon_i \) 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 \( y = \varepsilon_i^2 \), 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, \( B_\xi \leq Z_\xi \leq \sqrt{B_\xi} \), which is identical to (22). □

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. 1250-1266, 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. Arikan 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. Savoglu, 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] I. Duchi, “Lecture Notes for Statistics 311/Electrical Engineering 377,” Stanford University, 2016, https://stanford.edu/class/stats311/Lectures/full_notes.pdf