On the Pointwise Threshold Behavior of the Binary Erasure Polarization Subchannels

Erik Ordentlich, *Fellow, IEEE* Ron M. Roth, *Fellow, IEEE*

***Abstract*—It is shown that when Arıkan's n-level polarization transformation is applied to the binary erasure channel, each of the resulting individual 2^n subchannels has a sharp threshold, for sufficiently large n.**

***Index Terms*—Binary erasure channel, Channel polarization, Polar codes, Sharp threshold.**

I. Introduction

For a word $\boldsymbol{w} = w_1 w_2 \ldots w_n \in \{0,1\}^n$, we define the polynomial function $x \mapsto \mathrm{B}_{\boldsymbol{w}}(x)$ on the real interval $[0,1]$ recursively as follows. When $n = 1$,

$$\mathrm{B}_0(x) = x^2 \ , \quad \mathrm{B}_1(x) = 1 - (1-x)^2 = 2x - x^2 \ ,$$

and, for $n > 1$,

$$\begin{aligned} \mathrm{B}_{w_1 w_2 \ldots w_n}(x) &= \mathrm{B}_{w_n}(\mathrm{B}_{w_1 w_2 \ldots w_{n-1}}(x)) \\ &= \mathrm{B}_{w_n} \circ \mathrm{B}_{w_{n-1}} \circ \ldots \circ \mathrm{B}_{w_1}(x) \ . \end{aligned}$$

Since both $x \mapsto \mathrm{B}_0(x)$ and $x \mapsto \mathrm{B}_1(x)$ are (strictly) increasing on $[0,1]$, it follows that so is $x \mapsto \mathrm{B}_{\boldsymbol{w}}(x)$, for every $\boldsymbol{w} \in \{0,1\}^n$, with $\mathrm{B}_{\boldsymbol{w}}(0) = 0$ and $\mathrm{B}_{\boldsymbol{w}}(1) = 1$. In particular, there is a unique "half-way" point $\alpha_{\boldsymbol{w}}$ such that $\mathrm{B}_{\boldsymbol{w}}(\alpha_{\boldsymbol{w}}) = 1/2$. Figure 1 depicts the eight functions $x \mapsto \mathrm{B}_{\boldsymbol{w}}(x)$, for $n = 3$. For notational simplicity in the sequel, we formally define $\mathrm{B}_{\boldsymbol{w}}(x)$ to be 0 (respectively, 1) when $x < 0$ (respectively, $x > 1$).

For $\mathsf{p} \in [0,1]$ the values $\mathrm{B}_{\boldsymbol{w}}(\mathsf{p})$ for $\boldsymbol{w} \in \{0,1\}^n$ are known to be the error probabilities of the (binary erasure) subchannels obtained when Arıkan's n-level polarization transformation is applied to a binary erasure channel, BEC(p), with raw error probability p [1]. The channel polarization phenomenon discovered in [1] results in $\mathrm{B}_{\boldsymbol{w}}(\mathsf{p})$ concentrating around 0 or 1 for an increasing fraction of subchannels as n increases. The "good" subchannels $\boldsymbol{w}$, namely those for which $\mathrm{B}_{\boldsymbol{w}}(\mathsf{p})$ concentrate around 0, can be used to transmit a corresponding number of bits reliably and computationally efficiently, as described in detail in [1].

In general, the set of useful or "good" subchannels for a target rate R (or for a target decoding error probability) depends in a complicated manner on the raw channel erasure probability p, and must be recomputed with any change in this channel parameter. A somewhat more universal construction would involve a sequence of channel-parameter-independent rankings or orderings (one for each n) of the subchannels and would retain (as "good") the $\lfloor R \cdot 2^n \rfloor$ highest ranking subchannels. The existence of such rankings was first (implicitly) considered in [1, §X] where it was shown that ranking subchannels by the Hamming weight of the corresponding row of the end-to-end generator matrix and retaining the highest ranked subchannels (corresponding to the Reed-Muller code for rate R and block length 2^n) results in the inclusion of some bad subchannels for the BEC.

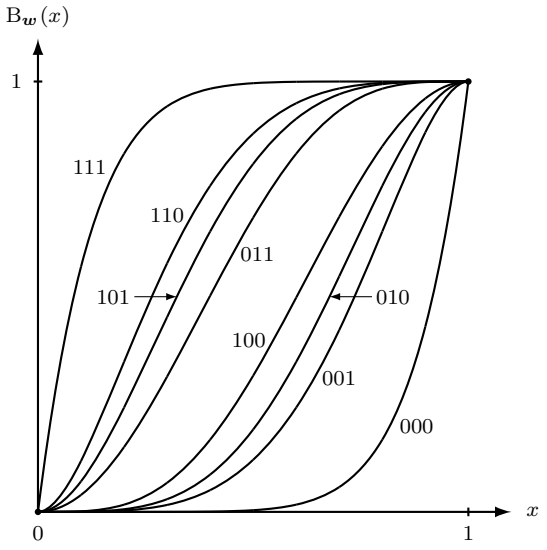

Fig. 1. Functions $x \mapsto \mathrm{B}_{\boldsymbol{w}}(x)$ for $\boldsymbol{w} \in \{0,1\}^3$.

To the best of our knowledge, it has been an open question, even for the case of the BEC, as to whether there exists a sequence of good channel-parameter-agnostic polarization subchannel ranking functions, in the sense that the error probability under successive cancellation decoding of the retained subchannels tends to 0 as long as $R < 1 - \mathsf{p}$. In this paper, we show that after excluding $\boldsymbol{w}$ with sufficiently low and sufficiently high Hamming weights (specified below), ranking subchannels by increasing $\alpha_{\boldsymbol{w}}$ is a good channel-parameter-agnostic subchannel ranking in the above sense.

We shall demonstrate this by establishing a sharp threshold behavior of the BEC error probability functions $x \mapsto \mathrm{B}_{\boldsymbol{w}}(x)$ at $x = \alpha_{\boldsymbol{w}}$, uniformly over $\boldsymbol{w} \in \{0,1\}^n$, as n goes to infinity. Namely, we find a mapping $(\delta, n) \mapsto \epsilon = \epsilon(\delta, n)$ from $\mathbb{Z}^+ \times (0,1]$ to $(0,1]$, where $\epsilon(\delta, n)$ tends to 0 (sufficiently

This work was supported in part by Grant 2015816 from the United-States–Israel Binational Science Foundation (BSF), and by NSF Grant CCF-BSF-1619053. Parts of this paper were presented at the International Symposium on Information Theory (ISIT), Aachen, Germany, June 2017.

Erik Ordentlich is with Verizon Media Group, Sunnyvale, CA 94089. Email: eordentlich@gmail.com

Ron M. Roth is with the Computer Science Dept., Technion, Haifa 3200003, Israel. Email: ronny@cs.technion.ac.il

fast) with n for every fixed δ, such that for every $\boldsymbol{w} \in \{0,1\}^n$,

$$\mathrm{B}_{\boldsymbol{w}}(\alpha_{\boldsymbol{w}} - \delta) \leq \epsilon \quad \text{and} \quad \mathrm{B}_{\boldsymbol{w}}(\alpha_{\boldsymbol{w}} + \delta) \geq 1 - \epsilon \; . \tag{1}$$

The existence of such a mapping from (δ, n) to ϵ is equivalent to the existence of infinite sequences $(\delta_n)_{n=1}^{\infty}$ and $(\epsilon_n)_{n=1}^{\infty}$ over $(0,1]$ such that (1) holds for every $(\epsilon, \delta) = (\epsilon_n, \delta_n)$ (see Example 1 below).

A more precise statement of this strong threshold result is as follows.

Denoting by $\overline{\boldsymbol{w}}$ the bitwise complement of $\boldsymbol{w}$, it is easy to verify by induction on n that

$$\mathrm{B}_{\overline{\boldsymbol{w}}}(x) = 1 - \mathrm{B}_{\boldsymbol{w}}(1 - x) \tag{2}$$

and, so, $\alpha_{\overline{\boldsymbol{w}}} = 1 - \alpha_{\boldsymbol{w}}$. Hence, to establish the threshold behavior for every $\boldsymbol{w} \in \{0,1\}^n$, it suffices to show that, say,

$$\mathrm{B}_{\boldsymbol{w}}(\alpha_{\boldsymbol{w}} + \delta) \geq 1 - \epsilon \tag{3}$$

for $\epsilon = \epsilon(\delta, n)$ tending to 0 as $n \to \infty$, since this implies that $\mathrm{B}_{\overline{\boldsymbol{w}}}(\alpha_{\overline{\boldsymbol{w}}} - \delta) \leq \epsilon$.

In our main analysis, we will obtain relations between δ and ϵ which will be parametric also in the Hamming weight $\|\boldsymbol{w}\|$ of $\boldsymbol{w}$. Specifically, the following relations between δ and $\epsilon = \epsilon(\delta, n, \|\boldsymbol{w}\|)$ will be a consequence of Proposition 2 in Section II and Theorem 17 in Section III-C. (Hereafter, all logarithms are taken to base 2, except when we specifically use the natural logarithm function $\ln(\cdot)$; the notation $\exp z$ stands for the natural exponential function e^z.)

Theorem 1 (Sharp threshold of $x \mapsto \mathrm{B}_{\boldsymbol{w}}(x)$)**.** *Eq. (1) holds for any $\boldsymbol{w} \in \{0,1\}^n$, where ϵ is related to $\delta \in [0,1]$ by*

$$\epsilon = \beta^{2^{\tau \cdot \delta^{\omega} \cdot \sqrt{\min(\|\boldsymbol{w}\|, \|\overline{\boldsymbol{w}}\|)}}} \; , \tag{4}$$

for real constants $\beta = \sqrt[8]{\ln 2}$ (< 0.955), $\omega < 1.537$, and $\tau > 1.758$ (where $\|\overline{\boldsymbol{w}}\| = n - \|\boldsymbol{w}\|$).[1]

In the special case that $\min(\|\boldsymbol{w}\|, \|\overline{\boldsymbol{w}}\|) \leq \log n - \log\log n$, Eq. (1) holds with

$$\epsilon = \exp\left\{ -\frac{2^n \log n}{n} \cdot \delta^{n/\log n} \right\} \; . \tag{5}$$

Note that ϵ in (4) tends to zero only when $\min\{\|\boldsymbol{w}\|, \|\overline{\boldsymbol{w}}\|\}$ goes to infinity. The case where either $\|\boldsymbol{w}\|$ or $\|\overline{\boldsymbol{w}}\|$ is of the order $\log n$ is covered by (5), which we will treat first, in Section II.

It follows from Theorem 1 that (1) is satisfied with an ϵ vanishing (much) faster than 2^{-n} for any given δ, whenever $\min\{\|\boldsymbol{w}\|, \|\overline{\boldsymbol{w}}\|\} \gg \log^2 n$. Moreover, ϵ decays doubly-exponentially with n for any $\boldsymbol{w}$ in the set

$$\mathcal{M}(n) = \left\{ \boldsymbol{w} \in \{0,1\}^n \; : \; \left| \|\boldsymbol{w}\| - \frac{n}{2} \right| \leq \gamma n \right\} \; ,$$

where γ is some fixed real in $(0, 1/2)$. Note that $1 - (|\mathcal{M}(n)|/2^n)$ vanishes exponentially with n.

Next, we demonstrate how Theorem 1 implies a capacity-achieving coding scheme for BEC(p), consisting of a pre-processing ranking phase which is universal in the sense that it depends only on n: in that phase, the words $\boldsymbol{w}$ in $\{0,1\}^n$ (or, more precisely, in $\mathcal{M}(n)$) are ordered from the smallest value of $\alpha_{\boldsymbol{w}}$ (corresponding to the lowest rank) to the largest. The dependence on the channel parameter p and on the target rate (or on the target decoding error probability) then amounts to just selecting the high-ranking subchannels down to a prescribed cut-off point.

Specifically, for $\vartheta \in [0,1]$, define the set

$$\mathcal{C}(\vartheta, n) = \{ \boldsymbol{w} \in \{0,1\}^n \; : \; \alpha_{\boldsymbol{w}} \geq \vartheta \} \; . \tag{6}$$

Thus, if we envision a list consisting of all words in $\{0,1\}^n$ sorted in descending order of $\alpha_{\boldsymbol{w}}$, the set $\mathcal{C}(\vartheta, n)$ consists of all the high-ranking words down to the cut-off value ϑ. Given the channel parameter $\mathsf{p} \in [0,1]$, the coding scheme is defined by taking all subchannels $\boldsymbol{w}$ in the intersection $\mathcal{C}(\mathsf{p}{+}\delta, n) \cap \mathcal{M}(n)$, for a fixed (with n) $\delta > 0$ that is determined by the target rate or the designed decay (with n) of the decoding error probability. Namely, we show that as n goes to infinity, the coding scheme has vanishing error probability under successive cancellation decoding, as well as rate R approaching (no less than) $1 - \mathsf{p} - 2\delta$.

Indeed, from Theorem 1 we get that the decoding error probability, which is given by the sum

$$\sum_{\boldsymbol{w} \in \mathcal{C}(\mathsf{p}+\delta, n) \cap \mathcal{M}(n)} \mathrm{B}_{\boldsymbol{w}}(\mathsf{p}) \; ,$$

is bounded from above by $2^n \cdot \epsilon$, for $\epsilon = \epsilon(\delta, n)$ that decays doubly-exponentially with $\sqrt{n}$ to 0 (where the exponents depend on δ). Hence, the decoding error probability decays doubly-exponentially with $\sqrt{n}$ as well.

To establish a lower bound on the rate R, we use the following equality concerning the polynomials $\mathrm{B}_{\boldsymbol{w}}(x)$ (see [1, Proposition 7]):

$$\frac{1}{2^n} \sum_{\boldsymbol{w} \in \{0,1\}^n} \mathrm{B}_{\boldsymbol{w}}(x) = x \; . \tag{7}$$

We then have:

$$\begin{aligned} \mathsf{p} + 2\delta \; &\overset{(7)}{=} \; \frac{1}{2^n} \sum_{\boldsymbol{w} \in \{0,1\}^n} \mathrm{B}_{\boldsymbol{w}}(\mathsf{p}{+}2\delta) \\ &\geq \; \frac{1}{2^n} \sum_{\boldsymbol{w} \in \overline{\mathcal{C}(\mathsf{p}+\delta, n)} \cap \mathcal{M}(n)} \mathrm{B}_{\boldsymbol{w}}(\mathsf{p}{+}2\delta) \; , \end{aligned}$$

where $\overline{\mathcal{C}(\vartheta, n)}$ stands for the complement set $\{0,1\}^n \setminus \mathcal{C}(\vartheta, n)$. Again, by Theorem 1 we get that $\mathrm{B}_{\boldsymbol{w}}(\mathsf{p}{+}2\delta) \geq 1 - \epsilon$ for every $\boldsymbol{w} \in \overline{\mathcal{C}(\mathsf{p}+\delta, n)} \cap \mathcal{M}(n)$, where $\epsilon = \epsilon(\delta, n)$ decays doubly-exponentially with $\sqrt{n}$ to 0. Hence,

$$\begin{aligned} \mathsf{p} + 2\delta \; &\geq \; \frac{1}{2^n} \left| \overline{\mathcal{C}(\mathsf{p}+\delta, n)} \cap \mathcal{M}(n) \right| - \epsilon \\ &= \; \frac{1}{2^n} \left| \mathcal{M}(n) \right| - \frac{1}{2^n} \left| \mathcal{C}(\mathsf{p}+\delta, n) \cap \mathcal{M}(n) \right| - \epsilon \\ &= \; 1 - 2^{-\Omega(n)} - R - \epsilon \; , \end{aligned}$$

where $\Omega(n)$ stands for an expression which grows (at least) linearly with n, for sufficiently large n. We readily conclude that

$$R \geq 1 - \mathsf{p} - 2\delta - \epsilon - 2^{-\Omega(n)} \; .$$

Although we believe that this is the first channel-parameter-agnostic ranking of subchannels whose highest ranked subchannels yield vanishing error probability under successive

[1] Our analysis will use loose bounds at some points, and therefore, by no means is (4) meant to present the smallest possible ϵ for given δ, n, and $\|\boldsymbol{w}\|$.

cancellation decoding, it does suffer from several drawbacks. One stems from the fact that the relation (4) is worse than the best decoding error probability, $\exp\left\{-2^{n/2-o(n)}\right\}$, attainable by polar codes [2] and, so, the above design approach, *per se*, guarantees a sub-optimal—but still super-polynomial in the code length 2^n—error decay (this sub-optimality is also reflected in yielding a worse scaling exponent behavior [4], [3], [7], [8], [5], [10]). One redeeming aspect here is that once the subchannels have been selected, one can use (explicit) enhancement techniques, as in [6], to push the decoding error probability down to the best asymptotic decay. A more challenging deficiency, however, is that as of yet, we lack a concise description of the set $\mathcal{C}(\vartheta, n)$ in (6), for any given $\vartheta \in [0,1]$. More generally, we lack a simple way of describing the list of words $\boldsymbol{w} \in \{0,1\}^n$ when ordered according to increasing $\alpha_{\boldsymbol{w}}$, e.g., through a function that is computable in polynomial-time in n which maps an index $i \in \{1,2,\ldots,2^n\}$ to the ith ranking word on the list (or an inverse function that maps $\vartheta \in [0,1]$ to an index of the lowest-ranking $\boldsymbol{w} \in \{0,1\}^n$ such that $\alpha_{\boldsymbol{w}} \geq \vartheta$). As of now, this problem is still open.

II. Low-weight $\boldsymbol{w}$

We first note that among all $\boldsymbol{w} \in \{0,1\}^n$ of the same Hamming weight m, the value of $\alpha_{\boldsymbol{w}}$ is minimized for

$$\boldsymbol{w}^* = \boldsymbol{w}^*(n,m) = \underbrace{1\,1\,\ldots\,1}_{m \text{ times}}\,\underbrace{0\,0\,\ldots\,0}_{n-m \text{ times}} ,$$

in which case

$$\mathrm{B}_{\boldsymbol{w}^*}(x) = \left(1-(1-x)^{2^m}\right)^{2^{n-m}} \tag{8}$$

(see [6]). This follows from the fact that for every $x \in (0,1)$,

$$\begin{aligned} \mathrm{B}_{01}(x) &= 1-(1-x^2)^2 = x^2(2-x^2) \\ &< x^2(2-x)^2 = (2x-x^2)^2 = \mathrm{B}_{10}(x) \end{aligned}$$

and, so, every transposition of 01 into 10 in $\boldsymbol{w}$ strictly increases $\mathrm{B}_{\boldsymbol{w}}(x)$; thus, $\mathrm{B}_{\boldsymbol{w}^*}(x)$ strictly dominates any other function $\mathrm{B}_{\boldsymbol{w}}(x)$ with $\|\boldsymbol{w}\| = \|\boldsymbol{w}^*\|$ (see also [11] and references therein).

The next proposition deals with the threshold behavior of functions $\mathrm{B}_{\boldsymbol{w}}(x)$, where $\|\boldsymbol{w}\| \leq \log n - \log\log n$.

Proposition 2. *For $\delta \in (0,1]$ and $n > 1$, let*

$$\epsilon = \epsilon^*(\delta, n) = \exp\left\{-\frac{2^n \log n}{n} \cdot \delta^{n/\log n}\right\} . \tag{9}$$

Then, for every $\boldsymbol{w} \in \{0,1\}^n$, $n > 1$, such that $\|\boldsymbol{w}\| \leq \log n - \log\log n$,

$$\mathrm{B}_{\boldsymbol{w}}(\alpha_{\boldsymbol{w}} - \delta) \leq \epsilon \quad \textit{and} \quad \mathrm{B}_{\boldsymbol{w}}(\alpha_{\boldsymbol{w}} + \delta) \geq 1-\epsilon .$$

Proof. Writing $m = \|\boldsymbol{w}\|$ and noting that $z \mapsto (1-\delta^z)^{1/z}$ is increasing in $z > 0$, we have:

$$\begin{aligned} \mathrm{B}_{\boldsymbol{w}}(1-\delta) &\leq \mathrm{B}_{\boldsymbol{w}^*}(1-\delta) \\ &\overset{(8)}{=} \left(1-\delta^{2^m}\right)^{2^n/2^m} \\ &\leq \left(1-\delta^{n/\log n}\right)^{(2^n \log n)/n} \\ &\leq \exp\left\{-\frac{2^n \log n}{n} \cdot \delta^{n/\log n}\right\} \\ &= \epsilon , \end{aligned}$$

i.e.,

$$1-\delta \leq \mathrm{B}_{\boldsymbol{w}}^{-1}(\epsilon) . \tag{10}$$

Since the proposition obviously holds when $\epsilon \geq 1/2$, we assume from now on in the proof that $\epsilon < 1/2$, in which case

$$1-\delta \overset{(10)}{\leq} \mathrm{B}_{\boldsymbol{w}}^{-1}(\epsilon) < \mathrm{B}_{\boldsymbol{w}}^{-1}(1/2) = \alpha_{\boldsymbol{w}} < 1 .$$

So,

$$\mathrm{B}_{\boldsymbol{w}}(\underbrace{\alpha_{\boldsymbol{w}} + \delta}_{>1}) = 1 > 1-\epsilon ,$$

and, by monotonicity,

$$\mathrm{B}_{\boldsymbol{w}}(\alpha_{\boldsymbol{w}} - \delta) < \mathrm{B}_{\boldsymbol{w}}(1-\delta) \leq \epsilon .$$

□

It is easy to see that for every fixed $\delta \in (0,1]$, the mapping $(\delta, n) \mapsto \epsilon^*(\delta, n)$ in (9) goes to 0 as $n \to \infty$.

Example 1. Let the sequence $(\delta_n)_{n=2}^{\infty}$ be defined by

$$\delta_n = \frac{4^{(\log^2 n)/n}}{n} \quad \left(= \frac{1+o(1)}{n}\right)$$

(where $o(1)$ stands for an expression that goes to 0 as $n \to \infty$). Substituting $\delta = \delta_n$ in (9) yields

$$\begin{aligned} &\epsilon^*(\delta_n, n) \\ &\quad = \exp\left\{-\frac{2^n \log n}{n} \cdot \left(2^{(2\log^2 n)/n - \log n}\right)^{n/\log n}\right\} \\ &\quad = \exp\left\{-\frac{2^n \log n}{n} \cdot 2^{2\log n - n}\right\} \\ &\quad = n^{-n\log e} . \end{aligned}$$

Defining $(\epsilon_n)_{n=2}^{\infty}$ by

$$\epsilon_n = n^{-n\log e} ,$$

it follows from Proposition 2 that for every $\boldsymbol{w} \in \{0,1\}^n$, $n > 1$, such that $\|\boldsymbol{w}\| \leq \log n - \log\log n$,

$$\mathrm{B}_{\boldsymbol{w}}(\alpha_{\boldsymbol{w}} - \delta_n) \leq \epsilon_n \quad \text{and} \quad \mathrm{B}_{\boldsymbol{w}}(\alpha_{\boldsymbol{w}} + \delta_n) \geq 1-\epsilon_n .$$

□

By (2) it follows that Proposition 2 holds also when $\|\boldsymbol{w}\| \geq n - (\log n - \log\log n)$. For such $\boldsymbol{w}$ we have $0 < \alpha_{\boldsymbol{w}} < \delta$.

We note that the value $\log n - \log\log n$ for m is (almost) the largest possible to have $\alpha_{\boldsymbol{w}^*} = 1 - o(1)$. Indeed, it can be readily verified that when $m = \log n$ we already have

$\alpha_{\boldsymbol{w}^*} = 1/2 - o(1)$. Furthermore, when $m \geq \log n + \log\ln n$ we get, for $x = 1/\log n$:

$$\begin{aligned}
\mathrm{B}_{\boldsymbol{w}^*}\Big(\frac{1}{\log n}\Big) &= \left(1 - \left(1 - \frac{1}{\log n}\right)^{2^m}\right)^{2^n/2^m} \\
&\geq \left(1 - \left(1 - \frac{1}{\log n}\right)^{n\ln n}\right)^{2^n/(n\ln n)} \\
&\geq \left(1 - 2^{-n}\right)^{2^n/(n\ln n)} \\
&= 1 - o(1)\ .
\end{aligned}$$

Namely, $\alpha_{\boldsymbol{w}^*} < 1/\log n$.

III. The general case

Given $\boldsymbol{w} = w_1 w_2 \ldots w_n \in \{0,1\}^n$, let $(x_0, x_1, x_2, \ldots, x_n)$ be the sequence over $(0,1)$ defined by $x_0 = \alpha_{\boldsymbol{w}}$ and

$$x_i = \mathrm{B}_{w_i}(x_{i-1})\ , \quad i = 1, 2, \ldots, n\ ,$$

where $x_n = \mathrm{B}_{\boldsymbol{w}}(x_0) = \mathrm{B}_{\boldsymbol{w}}(\alpha_{\boldsymbol{w}}) = 1/2$. Also, for some $y_0 \in (x_0, 1)$ let $(y_0, y_1, \ldots, y_n)$ be a sequence over $(0,1)$ that is defined similarly, i.e., $y_i = \mathrm{B}_{w_i}(y_{i-1})$, for $i = 1, 2, \ldots, n$; thus, $y_n = \mathrm{B}_{\boldsymbol{w}}(y_0)$. Since $x \mapsto B_0(x)$ and $x \mapsto B_1(x)$ are both increasing functions on $[0,1]$, it follows by induction on i that $y_i > x_i$ for $i = 1, 2, \ldots, n$.

For $i = 0, 1, \ldots, n$, write $y_i = x_i^{p_i}$, where $p_i \in (0,1)$ (namely, $p_i = (\log y_i)/(\log x_i)$). Our analysis of the threshold behavior of $x \mapsto \mathrm{B}_{\boldsymbol{w}}(x)$ will be carried out through studying the evolution of p_i, as i ranges from 0 to n. This study, in turn, will lead to a relationship between δ, n, and ϵ for which (1) is satisfied while having $\epsilon \to 0$ when $n \to \infty$.

For an index $i > 0$ such that $w_i = 0$, we have:

$$p_i = \frac{\log y_i}{\log x_i} = \frac{\log y_{i-1}^2}{\log x_{i-1}^2} = p_{i-1}\ .$$

Otherwise, when $w_i = 1$:

$$\begin{aligned}
p_i = \frac{\log y_i}{\log x_i} &= \frac{\log\left(2y_{i-1} - y_{i-1}^2\right)}{\log\left(2x_{i-1} - x_{i-1}^2\right)} \\
&= \frac{\log y_{i-1} + \log\left(2 - y_{i-1}\right)}{\log x_{i-1} + \log\left(2 - x_{i-1}\right)} \\
&= \frac{p_{i-1}\log x_{i-1} + \log\left(2 - x_{i-1}^{p_{i-1}}\right)}{\log x_{i-1} + \log\left(2 - x_{i-1}\right)} \\
&= p_{i-1} \cdot \frac{\log x_{i-1} + (1/p_{i-1})\log\left(2 - x_{i-1}^{p_{i-1}}\right)}{\log x_{i-1} + \log\left(2 - x_{i-1}\right)}\ .
\end{aligned}$$

That is,

$$p_i = \begin{cases} p_{i-1} & \text{if } w_i = 0 \\ p_{i-1} \cdot f(x_{i-1}, p_{i-1}) & \text{if } w_i = 1 \end{cases}\ , \tag{11}$$

where $f : (0,1) \times (0,1) \to \mathbb{R}$ is defined by

$$f(x,p) = \frac{\log x + (1/p)\log\left(2 - x^p\right)}{\log x + \log\left(2 - x\right)}\ . \tag{12}$$

The rest of this section is organized as follows. In Section III-A, we present several properties of the function $f(\cdot,\cdot)$ in (12). Then, in Section III-B, we utilize those properties to show that if p_0 is (close to, yet) sufficiently bounded away from 1 (from below), the sequence of p_i's evolve by (11) to a value p_n that is close to 0. The resulting convergence to 0 is quite slow, yet the main purpose of the analysis is to show that at some point in the sequence, p_i gets below a certain value, which, in turn, guarantees a much faster convergence to 0 from that point onward. That convergence will be analyzed in Section III-C.

A. Analysis of the function $(x,p) \mapsto f(x,p)$

We extend the function f to the domain $([0,1] \times [0,1]) \setminus \{(0,0)\}$ as follows. For $x = 0$ and $p \in (0,1]$,

$$f(0,p) = \lim_{x \to 0^+} f(x,p) = 1\ ,$$

and for $x = 1$ and $p \in [0,1]$ we have, by L'Hôpital's rule:

$$\begin{aligned}
f(1,p) &= \lim_{x \to 1} f(x,p) \\
&= \lim_{x \to 1} \frac{(1/x) - (1/p)\cdot p \cdot x^{p-1}/(2 - x^p)}{(1/x) - 1/(2-x)} \\
&= \lim_{x \to 1} \frac{(2-x)(1-x^p)}{(2-x^p)(1-x)} \\
&= \lim_{x \to 1} \frac{1 - x^p}{1 - x} = \lim_{x \to 1} p \cdot x^{p-1} = p\ .
\end{aligned}$$

For $p = 0$ and $x \in (0,1]$,

$$\begin{aligned}
f(x,0) &= \lim_{p \to 0^+} f(x,p) \\
&= \lim_{p \to 0^+} \frac{\log x - (x^p \log x)/(2 - x^p)}{\log x + \log\left(2 - x\right)} = 0\ ,
\end{aligned}$$

and for $p = 1$ and $x \in [0,1]$,

$$f(x,1) = \frac{\log x + (1/p)\log\left(2 - x^p\right)}{\log x + \log\left(2 - x\right)}\ \Big|_{p=1} = 1\ .$$

Figure 2 shows the curves $p \mapsto f(x,p)$ for various values of x.

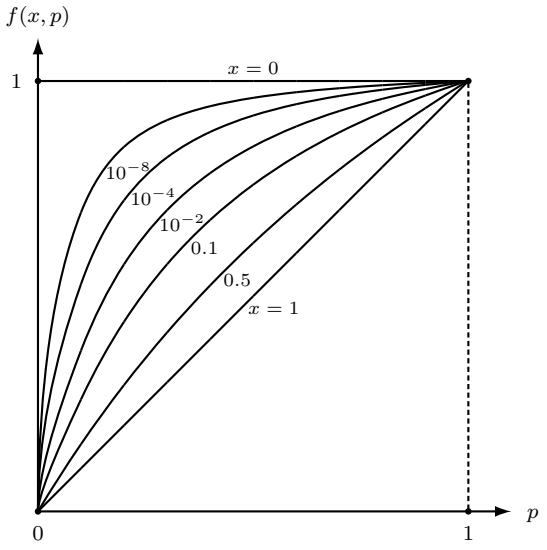

Fig. 2. Function $p \mapsto f(x,p)$ for various values of $x \in [0,1]$.

The proofs of the next three lemmas can be found in Appendix A.

Lemma 3. *For every $x \in (0,1]$, the function $p \mapsto f(x,p)$ is increasing and concave at every $p \in [0,1]$.*

Since $f(x,0) = 0$ and $f(x,1) = 1$, it follows by monotonicity that the function $p \mapsto f(x,p)$ is onto $[0,1]$, for every $x \in (0,1]$.

Lemma 4. *For every $x \in (0,1]$,*

$$\frac{\partial}{\partial p} f(x,p)\Big|_{p=1} \geq \frac{1}{1-\log x} \ .$$

Lemma 5. *For every $p \in (0,1)$, the function $x \mapsto f(x,p)$ is decreasing at every $x \in [0,1]$.*

Our analysis of the evolution of p_i, as i ranges from 0 to n, will use the properties of the function $(x,p) \mapsto f(x,p)$ (in conjunction with (11)), and will lead to a mapping $(\delta, m) \mapsto \epsilon^+(\delta, m)$ such that $\lim_{m\to\infty} \epsilon^+(\delta,m) = 0$ and

$$\mathrm{B}_{\boldsymbol{w}}(\alpha_{\boldsymbol{w}} + \delta) \geq 1 - \epsilon^+(\delta, \|\boldsymbol{w}\|) \ ,$$

for every $\boldsymbol{w} \in \{0,1\}^n$; that is, ϵ^+ is a function of the Hamming weight of $\boldsymbol{w}$ (rather than of n), and the superscript "+" indicates that we require only the inequality (3) (i.e., the right inequality in (1)) to hold. Once we have such a mapping, both inequalities in (1) will be satisfied for $\epsilon = \epsilon(\delta, n)$ given by

$$\epsilon(\delta, n) = \max\left\{\epsilon^*(\delta,n), \max_{m \in \mathcal{S}(n)} \epsilon^+(\delta,m)\right\} \ ,$$

where $\epsilon^*(\delta,n)$ is defined in (9) and

$$\mathcal{S}(n) = \left\{m \in \mathbb{Z}^+ \ : \ \min\{m, n-m\} > \log n - \log\log n\right\} .$$

Since both $\epsilon^*(\delta,n)$ and $\epsilon^+(\delta,m)$ tend to 0 as n (or m) $\to \infty$, then so is $\epsilon(\delta,n)$.

Given $\boldsymbol{w} \in \{0,1\}^n$, we next identify for any $p_0 = (\log y_0)/(\log x_0)$ (and for the respective p_n obtained from p_0 by (11)) a pair (δ, ϵ) for which (3) holds. Consider the difference

$$y_0 - x_0 = x_0^{p_0} - x_0 \ .$$

For a given p_0, the right-hand side is maximized when $x_0^{p_0-1} = 1/p_0$, i.e., $x_0 = p_0^{1/(1-p_0)}$. So,

$$\begin{aligned} y_0 - \alpha_{\boldsymbol{w}} &= y_0 - x_0 = x_0\left(x_0^{p_0-1} - 1\right) \\ &\leq p_0^{1/(1-p_0)} \cdot \left(\frac{1}{p_0} - 1\right) \\ &< \frac{1-p_0}{e \cdot p_0} \ . \end{aligned}$$

Hence, a given $p_0 > 1/e$ corresponds to taking, say,

$$\delta = 1 - p_0 \ . \tag{13}$$

As for the respective p_n,

$$y_n = x_n^{p_n} = \left(\frac{1}{2}\right)^{p_n} \geq 1 - (\ln 2) \cdot p_n \ .$$

Thus, if p_0 evolves into p_n and δ is selected according to (13), then (3) holds for any

$$\epsilon \geq (\ln 2) \cdot p_n \ . \tag{14}$$

B. Crude analysis of the evolution

Given $\boldsymbol{w} \in \{0,1\}^n$, write $m = \|\boldsymbol{w}\|$, and let $i_1 < i_2 < \ldots < i_m$ be the indexes i for which $w_i = 1$. Since $p_i < p_{i-1}$ only when $w_i = 1$ (and $p_i = p_{i-1}$ otherwise), we will find it notationally convenient to define the following subsequence $(q_j)_{j=0}^m$ of $(p_i)_{i=0}^n$:

$$q_0 = p_0 \ , \quad \text{and} \quad q_j = p_{i_j} \ , \quad j = 1,2,\ldots,m \ .$$

From (11):

$$q_j = q_{j-1} \cdot f(x_{i_j - 1}, q_{j-1}) \ , \quad j = 1,2,\ldots,m \ . \tag{15}$$

Let $(\xi_0, \xi_1, \ldots, \xi_m)$ be defined by a (backward) recursion as follows: $\xi_m = 1/2$, and

$$\xi_{j-1} = \mathrm{B}_1^{-1}(\xi_j) = 1 - \sqrt{1-\xi_j} \ , \quad j = m, m-1, \ldots, 1 \ .$$

Thus, $\xi_j = \mathrm{B}_1(\xi_{j-1}) = 2\xi_{j-1} - \xi_{j-1}^2$, for $j = 1,2,\ldots,m$.

Lemma 6. *For $j = 0,1,\ldots,m$,*

$$\xi_j \geq 2^{j-m-1} \ ,$$

with equality only if $j = m$.

Proof. For every $j = 1,2,\ldots,m$,

$$\xi_j = 2\xi_{j-1} - \xi_{j-1}^2 < 2\xi_{j-1}$$

and, so, by backward induction on j, with the induction base being $\xi_m = 1/2$, we have

$$\xi_{j-1} > \xi_j/2 \geq 2^{j-m-1}/2 = 2^{j-m-2} \ .$$

□

Lemma 7. *For $j = 1,2,\ldots,m$,*

$$x_{i_j} \geq \xi_j \quad \textit{and} \quad x_{i_j - 1} \geq \xi_{j-1} \ .$$

Proof. For $j = 1,2,\ldots,m$, we have

$$x_{i_j} = \mathrm{B}_1(x_{i_j - 1}) \leq \mathrm{B}_1(x_{i_{j-1}}) \ , \tag{16}$$

where the inequality follows from $x_{i_j - 1} \leq x_{i_{j-1}}$. Also, $x_{i_m} \geq x_n = 1/2 = \xi_m$, which serves as the base for a backward induction on j, with the following induction step:

$$x_{i_{j-1}} \overset{(16)}{\geq} \mathrm{B}_1^{-1}(x_{i_j}) \geq \mathrm{B}_1^{-1}(\xi_j) = \xi_{j-1} \ .$$

Hence, $x_{i_j} \geq \xi_j$ for $j = 0,1,\ldots,m$. It also follows that for $j = 1,2,\ldots m$,

$$x_{i_j - 1} = \mathrm{B}_1^{-1}(x_{i_j}) \geq \mathrm{B}_1^{-1}(\xi_j) = \xi_{j-1} \ .$$

□

From Lemmas 3 and 4 we get that for every $p \in [0,1)$:

$$\begin{aligned} \frac{1-f(x,p)}{1-p} &= \frac{f(x,1) - f(x,p)}{1-p} \\ &\geq \frac{\partial}{\partial p} f(x,p)\Big|_{p=1} \\ &\geq \frac{1}{1-\log x} \ , \end{aligned}$$

namely,

$$f(x,p) \leq 1 - \frac{1-p}{1-\log x} \ .$$

Hence, by Lemmas 6 and 7,

$$f(x_{i_j-1}, q_{j-1}) \le 1 - \frac{1-q_{j-1}}{1-\log \xi_{j-1}} \le 1 - \frac{1-q_{j-1}}{m-j+3} \ .$$

Combining with (15), we obtain

$$q_j \le q_{j-1}\Big(1 - \frac{1-q_{j-1}}{m-j+3}\Big) \ , \quad j = 1, 2, \ldots, m \ . \tag{17}$$

Let $U = (u_0, u_1, \ldots, u_m)$ be a sequence that satisfies (17) with equality, namely, for a given $u_0 \in (0,1)$ (to be determined later on):

$$u_j = u_{j-1}\Big(1 - \frac{1-u_{j-1}}{m-j+3}\Big) \ , \quad j = 1, 2, \ldots, m \ . \tag{18}$$

Equivalently,

$$1-u_j = (1-u_{j-1})\Big(1+\frac{u_{j-1}}{m-j+3}\Big) \ , \quad j = 1, 2, \ldots, m \ . \tag{19}$$

It is easy to see that the sequence U is decreasing and its elements are all in $(0,1)$. The following lemma is immediate.

Lemma 8. *Let $(q_0, q_1, \ldots, q_m)$ be any sequence that satisfies (17), where $q_0 = u_0$. Then $q_j \le u_j$, for all $j = 1, 2, \ldots, m$.*

Next, we turn to analyzing the sequence U. To this end, we fix a positive index $k \in \{1, 2, \ldots, m\}$ and assume that u_k equals some fixed real $\theta \in (0,1)$ to be determined later on. Lemma 9 below presents a lower bound on u_0, in terms of u_k, and Lemma 11 (whose proof makes use of Lemma 10) presents an upper bound on u_j, for $k \le j \le m$. The proofs of the lemmas are given in Appendix B.

Lemma 9. *Assuming that $u_k = \theta$,*

$$u_0 > 1 - \rho(\theta) \cdot \Big(\frac{m-k+3}{m+3}\Big)^{\theta} \ ,$$

where $\rho(\theta) = (1-\theta) \cdot e^{\theta^2/4}$ (< 1).

Lemma 10. *Given $(u_k =)\ \theta$, let $\ell \le m$ be an integer such that*

$$k \le \ell \le k - (m-k+3) \cdot \ln \theta + \ln\Big(\frac{m-\ell+3}{m-k+3}\Big) \ . \tag{20}$$

Then, for $k \le j \le \ell$,

$$u_j < \frac{m-j+3}{m-k+3} \ .$$

Lemma 11. *If $(u_k =)\ \theta \le \theta_0 = (3/8)^{1/8}/e \approx 0.325$, then, for $k \le j \le m$,*

$$u_j < \frac{m-j+3}{m-k+3} \ . \tag{21}$$

Remark 1. The bound (21) corresponds in fact to the case where $\theta = \theta_0$. For smaller θ, one can decrease the right-hand side of (21) by a constant factor (which depends on θ). □

Lemmas 8, 9, and 11 lead to the following result.

Proposition 12. *If p_0 is selected so that*

$$p_0 = q_0 = u_0 \ \left(> 1 - \rho(\theta) \cdot \Big(\frac{m-k+3}{m+3}\Big)^{\theta} \right)$$

for some $\theta \le \theta_0$, then, for $k < j \le m$,

$$q_j \le u_j < \frac{m-j+3}{m-k+3} \ .$$

The last proposition already implies a mapping $(\delta, m) \mapsto \epsilon^+(\delta, m)$ that is vanishing (albeit slowly) as $m \to \infty$, for which (3) is satisfied.

Theorem 13. *Given $\delta \in (0, 1-(1/e))$, let k be the smallest integer κ satisfying*

$$\delta \ge \rho_0 \cdot \Big(\frac{m-\kappa+3}{m+3}\Big)^{\theta_0} \ ,$$

where θ_0 (> 0.325) is as in Lemma 11 and $\rho_0 = (1-\theta_0) \cdot e^{\theta_0^2/4}$ (< 0.693). Then, for every $\boldsymbol{w} \in \{0,1\}^n$ such that $\|\boldsymbol{w}\| = m$,

$$\mathrm{B}_{\boldsymbol{w}}(\alpha_{\boldsymbol{w}} + \delta) \ge 1 - \epsilon \ ,$$

where

$$\epsilon = \epsilon^+(\delta, m) = \frac{3 \cdot \ln 2}{m-k+3} = \frac{1}{\Omega(m \cdot \delta^{1/\theta_0})} \ . \tag{22}$$

Proof. Selecting $p_0 = q_0 = 1 - \delta$ (as in (13)), we get from Proposition 12 that[2]

$$p_n = q_m < \frac{3}{m-k+3} \ .$$

The result then follows by noting that ϵ in (22) satisfies (14) and that $m-k+3 = \lfloor (m+3) \cdot (\delta/\rho_0)^{1/\theta_0} \rfloor$. □

In Section III-C below, we present a much better decay of $\epsilon^+(\delta, m)$ than (22).

In Appendix B, we prove the following (somewhat stronger) version of Lemma 9.

Lemma 14. *If $(u_k =)\ \theta \ge 1 - (2/3)\, e^{-5/4} \approx 0.809$, then*

$$u_0 > 1 - \frac{2}{3} \cdot \frac{m-k+3}{m+3} \ .$$

Note, however, that this lemma, as stated, cannot be used in conjunction with Lemma 11, since the two lemmas apply to disjoint ranges of values of u_k. For sufficiently larger m, the value θ_0 in Lemma 11 can be lowered at the expense of some scaling of the term $(m-j+3)/(m-k+3)$ in (21). We demonstrate this in Appendix C.

C. Fine analysis of the evolution

We use the notation $(q_j)_{j=0}^m$ and $(\xi_j)_{j=0}^m$ as defined in Section III-B.

The next proposition serves as our tool for getting a sharper threshold behavior of the functions $x \mapsto \mathrm{B}_{\boldsymbol{w}}(x)$.

Proposition 15 (Log-squaring rule)**.** *For $j = 1, 2, \ldots, m$:*

$$|q_j \ln \xi_j| \ \le \ |q_{j-1} \ln \xi_{j-1}|^2 \ .$$

The proof of the proposition makes use of the following lemma, the proof of which can be found in Appendix D.

Lemma 16. *For every $z \in (0, 1]$,*

$$-\ln\left(z(2-z)\right) \le \ln^2 z \ .$$

[2]The requirement $\delta < 1 - 1/e$ guarantees that $p_0 > 1/e$, which was assumed to obtain the relation (13).

Proof of Proposition 15. We will use (15) where we bound $f(x_{i_j-1}, q_{j-1})$ from above using Lemmas 5 and 16. Writing $\upsilon_{j-1} = \xi_{j-1}^{q_{j-1}}$, we have, for $j = 1, 2, \ldots, m$:

$$\begin{aligned} f(x_{i_j-1}, q_{j-1}) &\leq f(\xi_{j-1}, q_{j-1}) \\ &= \frac{1}{q_{j-1}} \cdot \frac{\log\left(\upsilon_{j-1}(2-\upsilon_{j-1})\right)}{\log\left(\xi_{j-1}(2-\xi_{j-1})\right)} \\ &\leq \frac{1}{q_{j-1}} \cdot \frac{\ln^2 \upsilon_{j-1}}{-\ln \xi_j} \\ &= q_{j-1} \cdot \frac{\ln^2 \xi_{j-1}}{-\ln \xi_j} \ , \end{aligned}$$

where the first inequality follows from Lemma 5 and the second from Lemma 16. Combining with (15) we obtain:

$$q_j = q_{j-1} \cdot f(x_{i_j-1}, q_{j-1}) \leq (q_{j-1})^2 \cdot \frac{\ln^2 \xi_{j-1}}{-\ln \xi_j} \ ,$$

thereby yielding the result. □

The next theorem is our main result.

Theorem 17. *Theorem 13 holds with (22) replaced by*

$$\epsilon = \epsilon^+(\delta, m) = \beta^{2^{\sqrt{m-k+4}}} < \beta^{2^{\tau \cdot \delta^\omega \cdot \sqrt{m}}} \ , \qquad (23)$$

where $\beta = \sqrt[8]{\ln 2}$ (< 0.955), $\omega = 1/(2\theta_0)$ (< 1.537), *and* $\tau = 1/\rho_0^\omega$ (> 1.758).

Proof. For k as defined in Theorem 13, let r be the unique integer which satisfies

$$(m-r+2)^2 \leq m-k+3 < (m-r+3)^2 \ . \qquad (24)$$

Selecting $p_0 = q_0 = 1 - \delta$ (as in (13)), we get from (24) and Proposition 12 that

$$q_r < \frac{m-r+3}{m-k+3} < \frac{1}{m-r+1} \ ,$$

which, with Lemma 6, yields

$$|q_r \ln \xi_r| < \ln 2 \ .$$

By Proposition 15 it follows that

$$\begin{aligned} p_n = q_m &\leq \frac{1}{|\ln \xi_m|} \cdot (\ln 2)^{2^{m-r}} \\ &= (\log e) \cdot (\ln 2)^{2^{m-r}} \\ &\overset{(24)}{\leq} (\log e) \cdot (\ln 2)^{2^{\sqrt{m-k+4}-3}} \\ &= (\log e) \cdot \beta^{2^{\sqrt{m-k+4}}} \ . \end{aligned}$$

Noting that ϵ in (23) satisfies (14) and that $m-k+4 = \lfloor (m+3) \cdot (\delta/\rho_0)^{2\omega} \rfloor + 1 > m \cdot (\delta/\rho_0)^{2\omega}$ yields

$$\begin{aligned} \epsilon &= \beta^{2^{\sqrt{m-k+4}}} \\ &< \beta^{2^{(\delta/\rho_0)^\omega \cdot \sqrt{m}}} \\ &= \beta^{2^{\tau \cdot \delta^\omega \cdot \sqrt{m}}} \ . \end{aligned}$$

□

For specific (small) n and m, one can enumerate over all words $\boldsymbol{w} \in \{0,1\}^n$ of a given Hamming weight m and compute a tight (i.e., best) mapping $\delta \mapsto \epsilon = \epsilon^+(\delta, n, m)$ for which (3) holds. We have done that for $n = 16$ and listed in Table I the values of δ that correspond to $\epsilon^+(\delta, n, m) = 10^{-4}$, for any given m, along with the maximizing (i.e., worst) word $\boldsymbol{w}$. Based on the results in the table, one may speculate whether, in general, for given n and m, the word $\boldsymbol{w}$ that exhibits the slowest threshold behavior is one for which the support indexes i_j are consecutive. In Table II, we have listed the words $\boldsymbol{w}$ that exhibit the smallest derivatives of $\mathrm{B}_{\boldsymbol{w}}(x)$ at $x = \alpha_{\boldsymbol{w}}$.

Proof of Theorem 1. The first part (Eq. (4)) is obtained by combining Theorem 17 with (2), and the second part (Eq. (5)) follows from Proposition 2. □

IV. Future work

We present below an alternate characterization of the functions $x \mapsto \mathrm{B}_{\boldsymbol{w}}(x)$; this characterization could lead to a different approach for deriving the threshold behavior of these functions.

For words $\boldsymbol{w} \in \{0,1\}^n$, consider subsets $\mathcal{A}_{\boldsymbol{w}} \subset \{0,1\}^{2^n}$ defined recursively as follows: $\mathcal{A}_0 = \{11\}$, $\mathcal{A}_1 = \{01, 10, 11\}$, and

$$\begin{aligned} \mathcal{A}_{\boldsymbol{w}0} &= \{\boldsymbol{u}\boldsymbol{v} \ : \ \boldsymbol{u}, \boldsymbol{v} \in \mathcal{A}_{\boldsymbol{w}}\} \\ \mathcal{A}_{\boldsymbol{w}1} &= \{\boldsymbol{u}\boldsymbol{v} \ : \ \boldsymbol{u} \in \mathcal{A}_{\boldsymbol{w}} \text{ or } \boldsymbol{v} \in \mathcal{A}_{\boldsymbol{w}}\} \ . \end{aligned}$$

It follows inductively from the recursive definition that each $\mathcal{A}_{\boldsymbol{w}}$ is a monotone set for all n and all $\boldsymbol{w}$, in the sense that $\boldsymbol{v} \in \mathcal{A}_{\boldsymbol{w}}$ implies that $\boldsymbol{u} \in \mathcal{A}_{\boldsymbol{w}}$ for every $\boldsymbol{u} \in \{0,1\}^{2^n}$ that satisfies the inequality $\boldsymbol{u} \geq \boldsymbol{v}$ componentwise.

It is not hard to see that $\mathrm{B}_{\boldsymbol{w}}(x)$ can be expressed as

$$\mathrm{B}_{\boldsymbol{w}}(x) = \mathsf{Prob}\{\boldsymbol{V} \in \mathcal{A}_{\boldsymbol{w}}\} \ ,$$

where $\boldsymbol{V}$ is a random word taking values in $\{0,1\}^{2^n}$ whose entries are independent, identically distributed Bernoulli-x random variables. We conjecture that it should be possible to derive the aforementioned, or perhaps sharper, threshold behavior for $\mathrm{B}_{\boldsymbol{w}}(x)$ using isoperimetric based methods that have established the threshold behavior of the probabilities of

TABLE I
Values of δ, per Hamming weight m, that correspond to $\epsilon^+(\delta, n, m) = 10^{-4}$, and attaining words $\boldsymbol{w}$, for $n = 16$.

m	δ	Maximizing $\boldsymbol{w}$
0	0.00001	0000000000000000
1	0.00454	1000000000000000
2	0.07181	1100000000000000
3	0.20717	1110000000000000
4	0.24676	1111000000000000
5	0.18810	1111100000000000
6	0.14995	0111111000000000
7	0.12081	0011111110000000
8	0.11101	0011111111000000
9	0.10138	0011111111100000
10	0.09771	0001111111111000
11	0.10170	0001111111111100
12	0.10809	0001111111111110
13	0.11820	0001111111111111
14	0.07332	0011111111111111
15	0.01216	0111111111111111
16	0.00013	1111111111111111

some famous monotone sets under Bernoulli measures. The threshold behavior of the error probability of decoding linear codes on binary symmetric and erasure channels, for example, was established using such tools in [12]. We remark that a straightforward application of the main result (Theorem 2.2) of [12] does not seem possible, however, as the implied threshold behavior depends on a property of the boundary of the monotone sets in question, that, in the case of the sets $\mathcal{A}_{\boldsymbol{w}}$ above, fails to imply any threshold behavior.

APPENDIX A
ADDITIONAL PROOFS FOR SECTION III-A

For the upcoming proofs, we find it convenient to define the following function $Q : [0,1] \to \mathbb{R}$: for every $z \in (0,1]$,

$$Q(z) = z \log z + (2-z) \log (2-z) \; ,$$

and $Q(0) = \lim_{z \to 0^+} Q(z) = 2$. It is easy to see that $Q(z) = 2(1 - H(z/2))$, where $H(z) = -z \log z - (1-z) \log (1-z)$ is the binary entropy function. Thus, for $z \in [0,1]$, the function $Q(z)$ takes on $[0,2]$, is decreasing and convex, and $(d/dz)\, Q(z) = \log (z/(2-z))$.

Proof of Lemma 3. For $x = 1$ we have $f(1,p) = p$ and the lemma is immediate; hence we fix x from now on to be in $(0,1)$. Let

$$g(x,p) = -f(x,p) \cdot \log (x(2-x)) = -\log x - \frac{1}{p} \log (2 - x^p) \; .$$

Since $\log (x(2-x)) < 0$ for $x \in (0,1]$, it suffices to show that for any such x, the function $p \mapsto g(x,p)$ is increasing and concave on $[0,1]$.

Write $y = y(x,p) = x^p$. Then

$$\frac{\partial}{\partial p}\, y(x,p) = y \ln x \; .$$

TABLE II
SMALLEST DERIVATIVES OF $\mathrm{B}_{\boldsymbol{w}}(x)$ AT $x = \alpha_{\boldsymbol{w}}$ PER HAMMING WEIGHT m, AND ATTAINING WORDS $\boldsymbol{w}$, FOR $n = 16$.

m	$(d/dx)\, \mathrm{B}(x)\vert_{x=\alpha_{\boldsymbol{w}}}$	Minimizing $\boldsymbol{w}$
0	32768.34658	0000000000000000
1	150.71087	1000000000000000
2	17.18959	1100000000000000
3	8.95317	1110000000000000
4	9.54279	1111000000000000
5	12.58686	0111110000000000
6	15.31910	1110000000000111
7	16.56761	0011111110000000
8	17.50690	0011111111000000
9	16.56761	1100000001111111
10	15.31910	0001111111111000
11	12.58686	1000001111111111
12	9.54279	0000111111111111
13	8.95317	0001111111111111
14	17.18959	0011111111111111
15	150.71087	0111111111111111
16	32768.34658	1111111111111111

Computing the partial derivative of $g(x,p)$ with respect to p, we get:

$$\begin{aligned}
\frac{\partial}{\partial p}\, g(x,p) &= -\frac{\partial}{\partial p}\Big(\frac{1}{p} \log \big(2 - y(x,p)\big)\Big) \\
&= -\frac{\partial}{\partial p}\Big(\frac{1}{p} \log (2-y)\Big) \\
&\quad - \frac{\partial}{\partial y}\Big(\frac{1}{p} \log (2-y)\Big) \cdot \frac{\partial}{\partial p}\, y(x,p) \\
&= \frac{1}{p^2} \cdot \log (2-y) + \frac{1}{p} \cdot \frac{y \log x}{2-y} \\
&= \frac{1}{p^2 (2-y)} \Big((2-y) \log (2-y) + y \log y \Big) \; ,
\end{aligned}$$

i.e.,

$$\frac{\partial}{\partial p}\, g(x,p) = \frac{Q(y)}{p^2 (2-y)} \; . \tag{25}$$

For $p \in (0,1]$ and $x \in (0,1)$ we have $y = x^p \in (0,1)$ thereby implying that $(\partial/\partial p)\, g(x,p) > 0$, namely, that $p \mapsto f(x,p)$ is increasing.

As our next step, we compute the second partial derivative, $(\partial^2/\partial p^2)\, g(x,p)$, from (25). Recalling that $(d/dz)\, Q(z) = \log (z/(2-z))$ and that $(\partial/\partial p)\, y(x,p) = (y \ln y)/p$, we obtain:

$$\begin{aligned}
&\frac{\partial^2}{\partial p^2}\, g(x,p) \\
&\quad = \frac{\partial}{\partial p}\Big(\frac{Q(y)}{p^2(2-y)}\Big) \\
&\qquad + \frac{\partial}{\partial y}\Big(\frac{Q(y)}{p^2(2-y)}\Big) \cdot \frac{\partial}{\partial p}\, y(x,p) \\
&\quad = -\frac{2\, Q(y)}{p^3 (2-y)} \\
&\qquad + \frac{1}{p^2(2-y)} \left(\frac{Q(y)}{2-y} + \log\Big(\frac{y}{2-y}\Big) \right) \cdot \frac{y \ln y}{p} \\
&\quad = \frac{2\, \varphi(y)}{p^3 (2-y)^2} \; ,
\end{aligned}$$

where

$$\varphi(y) = (y-2) \cdot Q(y) + (\ln 2) \cdot y \log^2 y \; .$$

Thus, to establish the concavity of $p \mapsto f(x,p)$, it remains to show that $\varphi(y) < 0$ for $y \in (0,1)$. To this end, since $\varphi(1) = 0$, it suffices to show that $y \mapsto \varphi(y)$ is increasing at every $y \in [0,1)$. Indeed,

$$\begin{aligned}
\frac{d}{dy}\, \varphi(y) &= Q(y) + (y-2) \log\Big(\frac{y}{2-y}\Big) \\
&\qquad + (\ln 2) \log^2 y + 2 \log y \\
&= 2\, Q(y) + (\ln 2) \log^2 y \\
&> 0 \; .
\end{aligned}$$

This completes the proof. □

Proof of Lemma 4. It follows from (25) that

$$\begin{aligned}
\frac{\partial}{\partial p} f(x,p)\Big|_{p=1} &= -\frac{Q(x)}{(2-x) \log (x(2-x))} \\
&= -\frac{Q(x)}{Q(x) + 2(1-x) \log x} \; .
\end{aligned}$$

Hence, the lemma will be proved once we show that

$$-\frac{2(1-x)\log x}{Q(x)} - 2 \le -\log x \ . \tag{26}$$

We prove (26) separately for the neighborhood $(0, \varepsilon_0]$ of $x = 0$ and for the neighborhood $[1-\varepsilon_0, 1]$ of $x = 1$, where ε_0 is a fixed small positive real for which the $O(\cdot)$ terms that appear in our analysis become sufficiently small. For x in the interval $(\varepsilon_0, 1-\varepsilon_0)$, the inequality (26) can be verified numerically, or "by inspection" through Figure 3; in fact, the figure suggests that on that interval, the (negative) derivative (with respect to x) of the left-hand side of (26) is always greater than the derivative of the right-hand side.

The neighborhood of $x = 0$. Starting with the left-hand side of (26), and letting $O(x)$ denote a term that grows at most linearly with x, we have:

$$\begin{aligned} -\frac{2(1-x)\log x}{Q(x)} - 2 &= -\frac{2(1-x)\log x}{2 + x\log x - O(x)} - 2 \\ &= -\frac{2\log x + 4 - O(x)}{2 + x\log x - O(x)} \\ &= -\log x - 2 + O(x\log^2 x) \\ &\le -\log x \ , \end{aligned}$$

for sufficiently small $x > 0$.

The neighborhood of $x = 1$. Both sides of (26) converge to 0 when $x \to 1$. Substituting $x = 1 - \varepsilon$, the derivative with respect to ε of the left-hand side can be verified to be 1, while the derivative of the right-hand side equals $\log e \ (> 1)$. Hence, for small $\varepsilon > 0$, the left-hand side equals $\varepsilon + O(\varepsilon^2)$, while the right-hand side equals $\varepsilon \cdot \log e + O(\varepsilon^2)$. □

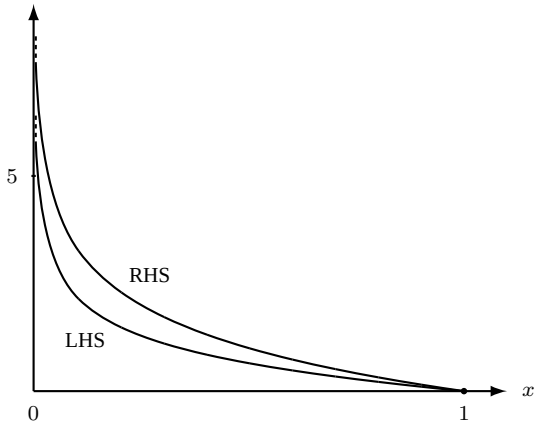

Fig. 3. Left-hand side and right-hand side of (26).

Proof of Lemma 5. Fixing $p \in (0,1)$, it can be verified that at every $x \in (0,1)$,

$$\frac{\partial}{\partial x} f(x,p) = \frac{(1-x)(1-y)\cdot \ln 2}{p \cdot x(2-x)(2-y)(\log(x(2-x)))^2} \cdot \mu(x) \ ,$$

where $y = y(x,p) = x^p \ (> x)$ and

$$\mu(x) = p \cdot \frac{Q(x)}{1-x} - \frac{Q(y)}{1-y} \ .$$

Hence, we need to show that $\mu(x)$ is negative for $x \in (0,1)$. Since $\mu(1) = \lim_{x\to 1}\mu(x) = 0$, it suffices to show that $\mu(x)$ is increasing at every $x \in (0,1)$. Now,

$$\frac{d}{dx}\frac{Q(x)}{1-x} = \frac{\log(x(2-x))}{(1-x)^2}$$

and, so,

$$\begin{aligned} \frac{d}{dx}\mu(x) &= p \cdot \frac{\log(x(2-x))}{(1-x)^2} \\ &\quad - \frac{\log(y(2-y))}{(1-y)^2} \cdot \underbrace{p \cdot x^{p-1}}_{(\partial/\partial x)\, y(x,p)} \\ &= \frac{p}{x \cdot \ln 2} \cdot \Big(\nu(x) - \nu(y)\Big) \ , \end{aligned}$$

where

$$\nu(x) = \frac{x \cdot \ln(x(2-x))}{(1-x)^2} \ .$$

Thus, the proof reduces to showing that $x \mapsto \nu(x)$ is decreasing on $(0,1)$, which, by taking derivatives, is equivalent to having

$$\eta(x) = \frac{2(1-x)^2}{(2-x)(1+x)} + \ln(x(2-x)) < 0 \ .$$

Since $\eta(1) = 0$, it therefore suffices to show that $x \mapsto \eta(x)$ is increasing on $(0,1)$. Indeed,

$$\frac{d}{dx}\eta(x) = \frac{2(x^2+2)(1-x)^2}{x(2-x)^2(1+x)^2} > 0 \ .$$

□

Appendix B
Additional proofs for Section III-B

Proof of Lemma 9. From (19) and the monotonicity of U we have

$$1 - u_j \ge (1-u_{j-1})\Big(1 + \frac{\theta}{m-j+3}\Big) \ , \quad j = 1, 2, \ldots, k \ ,$$

and, so,

$$1 - \theta = 1 - u_k \ge (1-u_0) \cdot \prod_{s=1}^{k}\Big(1 + \frac{\theta}{m-s+3}\Big) \ .$$

Therefore,

$$\begin{aligned} 1 - u_0 &\le (1-\theta) \cdot \prod_{s=1}^{k}\Big(1 + \frac{\theta}{m-s+3}\Big)^{-1} \\ &< (1-\theta) \cdot \prod_{s=1}^{k} \exp\left\{-\frac{\theta}{m-s+3} + \frac{\theta^2}{2(m-s+3)^2}\right\} \\ &< (1-\theta) \cdot \exp\left\{-\theta \cdot \sum_{s=1}^{k}\frac{1}{m-s+3} \right. \\ &\qquad \left. + \frac{\theta^2}{2} \cdot \sum_{s=-\infty}^{m}\frac{1}{(m-s+3)^2}\right\} \\ &< (1-\theta) \cdot \exp\left\{-\theta \cdot \ln\Big(\frac{m+3}{m-k+3}\Big) + \frac{\theta^2}{4}\right\} \ , \end{aligned}$$

where the second inequality follows from $\ln(1+z) > z - z^2/2$ (for $z \in (0,1)$) and the last inequality follows from

$$\sum_{s=a}^{b} \frac{1}{s} > \int_{a}^{b+1} \frac{d\,z}{z} = \ln\Big(\frac{b+1}{a}\Big)$$

and

$$\sum_{s=a}^{b} \frac{1}{s^2} < \sum_{s=a}^{b} \Big(\frac{1}{s-1} - \frac{1}{s}\Big) = \frac{1}{a-1} - \frac{1}{b}\,,$$

for integers $b \geq a > 1$. Thus, we conclude that

$$u_0 > 1 - \underbrace{(1-\theta)\cdot e^{\theta^2/4}}_{\rho(\theta)} \cdot \Big(\frac{m-k+3}{m+3}\Big)^{\theta}\,.$$

□

Proof of Lemma 10. We prove the lemma by induction on j. The lemma trivially holds for the induction base ($j = k$), where $u_k = \theta < 1$. Assuming now that $j > k$, we have

$$\begin{aligned} u_j &\overset{(18)}{=} \theta \cdot \prod_{s=k+1}^{j} \Big(1 - \frac{1-u_{s-1}}{m-s+3}\Big) \\ &< \theta \cdot \prod_{s=k+1}^{j} \exp\Big\{-\frac{1-u_{s-1}}{m-s+3}\Big\} \\ &\leq \theta \cdot \prod_{s=k+1}^{j} \exp\left\{-\frac{1-(m-s+4)/(m-k+3)}{m-s+3}\right\}\,, \end{aligned}$$

where the last step follows from the induction hypothesis. Hence,

$$\begin{aligned} u_j &< \theta \cdot \exp\left\{-\Big(1-\frac{1}{m-k+3}\Big)\Big(\sum_{s=k+1}^{j}\frac{1}{m-s+3}\Big) + \frac{j-k}{m-k+3}\right\} \\ &< \theta \cdot \exp\left\{\frac{j-k}{m-k+3} - \Big(1-\frac{1}{m-k+3}\Big)\ln\Big(\frac{m-k+3}{m-j+3}\Big)\right\} \\ &= \theta \cdot \exp\left\{\frac{1}{m-k+3}\cdot\Big(j-k-\ln\Big(\frac{m-j+3}{m-k+3}\Big)\Big)\right\} \cdot \frac{m-j+3}{m-k+3} \\ &\leq \exp\left\{\ln\theta + \frac{1}{m-k+3}\cdot\Big(\ell-k-\ln\Big(\frac{m-\ell+3}{m-k+3}\Big)\Big)\right\} \cdot \frac{m-j+3}{m-k+3}\,. \end{aligned}$$

By (20), the argument of the last $\exp\{\cdot\}$ is nonpositive and, so, we conclude that

$$u_j < \frac{m-j+3}{m-k+3}\,.$$

□

Proof of Lemma 11. The minimum of $(\ln(3/s))/s$ over $s \in \mathbb{Z}^+$ is attained at $s = 8$; therefore, under the conditions of the lemma we have

$$\begin{aligned} m &\leq k - (m-k+3)\cdot \underbrace{\left(\ln\theta - \frac{1}{8}\ln\Big(\frac{3}{8}\Big)\right)}_{\leq -1} \\ &\leq k - (m-k+3)\cdot\left(\ln\theta - \frac{1}{m-k+3}\cdot\ln\Big(\frac{3}{m-k+3}\Big)\right) \\ &\leq k - (m-k+3)\cdot\ln\theta + \ln\Big(\frac{3}{m-k+3}\Big)\,, \end{aligned}$$

which means that (20) is satisfied by $\ell = m$. □

Proof of Lemma 14. We start by finding a lower bound on u_j, for $j = 1, 2, \ldots, k$, as a function of u_0. By (19),

$$\begin{aligned} 1 - u_j &= (1-u_0)\cdot\prod_{s=1}^{j}\Big(1+\frac{u_{s-1}}{m-s+3}\Big) \\ &< (1-u_0)\cdot\prod_{s=1}^{j}\exp\left\{\frac{u_{s-1}}{m-s+3}\right\} \\ &= (1-u_0)\cdot\exp\left\{\sum_{s=1}^{j}\frac{u_{s-1}}{m-s+3}\right\} \\ &\leq (1-u_0)\cdot\exp\left\{\sum_{s=1}^{j}\frac{1}{m-s+3}\right\} \\ &< (1-u_0)\cdot\exp\left\{\ln\Big(\frac{m+2}{m-j+2}\Big)\right\}\,, \end{aligned}$$

where the last inequality follows from

$$\sum_{s=a}^{b}\frac{1}{s} < \int_{a-1/2}^{b+1/2}\frac{d\,z}{z} = \ln\Big(\frac{2b+1}{2a-1}\Big) < \ln\Big(\frac{b}{a-1}\Big)\,,$$

for integers $b \geq a > 1$. Hence,

$$u_j > 1 - (1-u_0)\cdot\frac{m+2}{m-j+2}\,. \tag{27}$$

Next, we compute a lower bound on $1 - u_k$ $(= 1-\theta)$:

$$\begin{aligned} 1-\theta &\overset{(19)}{=} (1-u_0)\cdot\prod_{s=1}^{k}\Big(1+\frac{u_{s-1}}{m-s+3}\Big) \\ &> (1-u_0)\cdot\prod_{s=1}^{k}\exp\left\{\frac{u_{s-1}}{m-s+3} - \frac{u_{s-1}^2}{2(m-s+3)^2}\right\} \\ &> (1-u_0)\cdot\exp\left\{\sum_{s=1}^{k}\frac{1}{m-s+3} - \sum_{s=1}^{k}\frac{1-u_{s-1}}{m-s+3} - \frac{1}{2}\sum_{s=-\infty}^{m}\frac{1}{(m-s+3)^2}\right\} \\ &> (1-u_0)\cdot\exp\left\{\ln\Big(\frac{m+3}{m-k+3}\Big) - \sum_{s=1}^{k}\frac{1-u_{s-1}}{m-s+3} - \frac{1}{4}\right\} \\ &> (1-u_0)\cdot\exp\left\{\ln\Big(\frac{m+3}{m-k+3}\Big)\right. \end{aligned}$$

$$- (1-u_0)(m{+}2)\cdot\sum_{s=1}^{k}\frac{1}{(m{-}s{+}3)^2}-\frac{1}{4}\Bigg\}\ ,$$

where the last step follows from (27). We therefore get:

$$\begin{aligned}(1-\theta)\cdot e^{1/4} &> (1-u_0)\cdot\frac{m{+}3}{m{-}k{+}3}\\ &\quad\cdot\exp\Big\{-(1-u_0)\Big(\frac{m{+}2}{m{-}k{+}2}-1\Big)\Big\}\\ &> (1-u_0)\cdot\frac{m{+}3}{m{-}k{+}3}\\ &\quad\cdot\exp\Big\{-(1-u_0)\cdot\frac{3}{2}\cdot\frac{m{+}3}{m{-}k{+}3}\Big\}\ ,\end{aligned}$$

i.e.,

$$(1-\theta)\cdot e^{1/4} > \psi\Big((1-u_0)\cdot\frac{m{+}3}{m{-}k{+}3}\Big)\ , \tag{28}$$

where $\psi(z)=z{\cdot}e^{-(3/2)z}$. To complete the proof of the lemma, it remains to show that

$$(1-u_0)\cdot\frac{m{+}3}{m{-}k{+}3}<\frac{2}{3}\ , \tag{29}$$

whenever θ satisfies

$$(1-\theta)\cdot e^{1/4}\le 2/(3\,e)=\psi(2/3)\ . \tag{30}$$

Let $\lambda: u_0\mapsto u_k$ be the (unique) function that maps u_0 to u_k by applying (18) for $j=1,2,\ldots,k$; clearly, $u_0\mapsto\lambda(u_0)$ is a continuous (strictly) increasing function on $(0,1)$. Let u_0^* be the unique value $u_0\in(0,1)$ for which (29) holds with equality, and assume by contradiction that (29) does not hold, namely, that $u_0<u_0^*$ for some choice of θ $(=u_k)$ that satisfies (30) (we are ruling out the equality $u_0=u_0^*$ since otherwise (30) would contradict (28)). Note that under this assumption, the argument of $\psi(\cdot)$ in (28) is greater than $2/3$ and is therefore in the range where $z\mapsto\psi(z)$ is decreasing. We will also assume hereafter that the inequality in (30) is strict; otherwise, we can slightly increase θ so that the respective $u_0=\lambda^{-1}(\theta)$ is (increased but is) still smaller than u_0^*.

Let ε be the unique positive real for which

$$(1-\lambda(u_0))\cdot e^{1/4}=\psi(2/3+\varepsilon)\ . \tag{31}$$

Under our assumptions, there exists $u_0'\in(u_0,u_0^*)$ such that

$$\frac{2}{3}<(1-u_0')\cdot\frac{m{+}3}{m{-}k{+}3}<\frac{2}{3}+\varepsilon\ .$$

We have,

$$\begin{aligned}\psi\Big((1-u_0')\cdot\frac{m{+}3}{m{-}k{+}3}\Big) &> \psi(2/3+\varepsilon)\\ &\overset{(31)}{=} (1-\lambda(u_0))\cdot e^{1/4}\\ &\overset{u_0'>u_0}{>} (1-\lambda(u_0'))\cdot e^{1/4}\ ,\end{aligned}$$

thereby contradicting (28). □

Appendix C
Extensions to Lemma 11

Let $d_0<d_1<d_2<\ldots<d_r$ be positive integers and $\theta_0,\theta_1,\ldots,\theta_r$ be respective positive reals in $(0,1)$ which are defined iteratively as follows: θ_0 is as in Lemma 11 and $d_0=m{-}k{+}3$ for the respective k therein. For $i=1,2,\ldots,r$, we let

$$d_i=\left\lceil\frac{d_{i-1}}{\theta_{i-1}}\right\rceil \tag{32}$$

and

$$\theta_i=\exp\left\{\frac{d_{i-1}}{d_i}-1+\frac{1}{d_i}\ln\Big(\frac{d_{i-1}}{d_i}\Big)\right\}\ . \tag{33}$$

Write $k_i=m{-}d_i{+}3$, for $i=0,1,2,\ldots,r$. It can be verified that the conditions of Lemma 10 are implied by (33) when we substitute $\theta\leftarrow\theta_i$, $k\leftarrow k_i$ and $\ell\leftarrow k_{i-1}$ therein. Hence, it follows by that lemma and (32) that for $i=1,2,\ldots,r$,

$$u_{k_{i-1}}\le\frac{m{-}k_{i-1}{+}3}{m{-}k_i{+}3}=\frac{d_{i-1}}{d_i}\le\theta_{i-1}\ .$$

In particular, $u_k\le\theta_0$ (as assumed in Lemma 11). Hence, if we assume that $u_{k_r}=\theta_r$, we get that for every $j\ge k=k_0$,

$$u_j\le\frac{m{-}j{+}3}{m{-}k_0{+}3}=\frac{m{-}j{+}3}{m{-}k_r{+}3}\cdot\frac{d_r}{d_0}\ . \tag{34}$$

By (32), the factor d_r/d_0, which equals $\prod_{i=1}^{r}(d_i/d_{i-1})$, is at least $1/\prod_{i=0}^{r-1}\theta_i$. On the other hand, we can now use Lemma 9 with k_r and θ_r replacing k and (the smaller) θ_0. This will lead to a smaller (i.e., better) power of δ in the right-hand side of (22), yet also to scaling by a factor of d_r/d_0.

Example 2. We demonstrate the improvement obtained by the above analysis when selecting $r=2$. Taking $\theta_0=0.3$ (say) and assuming that $d_0=m{-}k{+}3$ is a multiple of 3, we get $d_1=d_0/\theta_0=(10/3)\,d_0$. Then, we get a lower bound on θ_1 from (33) by replacing the term $(1/d_i)\ln(d_{i-1}/d_i)$ therein with the lower bound $(1/8)\ln(3/8)$, thereby yielding $\theta_1>0.439>7/16$. In the next iteration, we compute $d_2=\lceil d_1/\theta_1\rceil$, and, assuming that d_1 is a multiple of 7, we get $d_2=(16/7)\,d_1$. By (33), we finally get $\theta_2>0.504>1/2$. The factor d_2/d_0 in (34) in this example is

$$\frac{d_2}{d_0}=\frac{d_1}{d_0}\cdot\frac{d_2}{d_1}=\frac{10}{3}\cdot\frac{16}{7}\approx 7.62\ .$$

□

When m is very large so that we can assume large d_i's and ignore the ceiling in (32) and the rightmost term, $(1/d_i)\ln(d_{i-1}/d_i)$, in (33), then (33) becomes

$$\theta_i=e^{\theta_{i-1}-1}\ ,$$

in which case θ_i converges to 1 as $i\to\infty$.

Appendix D
Additional proofs for Section III-C

Proof of Lemma 16. Since both sides vanish at $z=1$, it suffices to show that the derivatives of both sides satisfy the inequality in the other direction, namely, that

$$-\frac{1}{z}+\frac{1}{2-z}\ge\frac{2\ln z}{z}\ ,$$

which simplifies to

$$\frac{1}{2-z}-1\ge\ln z\ .$$

Again, both sides vanish at $z = 1$, so we show that the derivatives satisfy the reverse inequality:

$$\frac{1}{(2-z)^2} \leq \frac{1}{z} \,.$$

Indeed, both sides are equal (to 1) at $z = 1$, yet the left-hand side is increasing on $(0, 1]$, while the right-hand side is decreasing. □

REFERENCES

[1] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, 55 (2009), 3051–3073.

[2] E. Arıkan, E. Telatar, "On the rate of channel polarization," *Proc. IEEE Int'l Symp. Inf. Theory*, Seoul, Korea (2009), 1493–1495.

[3] J. Błasiok, V. Guruswami, M. Sudan, "Polar codes with exponential small error at finite block length," *Proc. 22nd Int'l Workshop on Randomization and Approximation Techniques in Comp. Science (APPROX/RANDOM 2018)*, Princeton, New Jersey (2018), 34:1–34:17.

[4] J. Błasiok, V. Guruswami, P. Nakkiran, A. Rudra, M. Sudan, "General strong polarization," *Proc. 50th Annual ACM SIGACT Symp. Theory of Computing (STOC 2018)*, Los Angeles, California (2018), 471–484.

[5] D. Goldin, D. Burshtein, "Improved bounds on the finite length scaling of polar codes," *IEEE Trans. Inf. Theory*, 60 (2014), 6966–6978.

[6] V. Guruswami, P. Xia, "Polar codes: Speed of polarization and polynomial gap to capacity," *IEEE Trans. Inf. Theory*, 61 (2015), 3–16.

[7] A. Fazeli, H. Hassani, M. Mondelli, A. Vardy, "Binary linear codes with optimal scaling: Polar codes with large kernels," submitted to *IEEE Trans. Inf. Theory.*

[8] S.H. Hassani, K. Alishahi, R. Urbanke, "Finite-length scaling for polar codes," *IEEE Trans. Inf. Theory*, 60 (2014), 5875–5898.

[9] S. Kudekar, S. Kumar, M. Mondelli, H.D. Pfister, E. Şaşoğlu, R. Urbanke, "Reed–Muller codes achieve capacity on erasure channels," *IEEE Trans. Inf. Theory*, 63 (2017), 4298–4216.

[10] M. Mondelli, S.H. Hassani, R. Urbanke, "Unified scaling of polar codes: Error exponent, scaling exponent, moderate deviations, and error floors," *IEEE Trans. Inf. Theory*, 62 (2016), 6698–6712.

[11] M. Mondelli, S.H. Hassani, R. Urbanke, "Construction of polar codes with sublinear complexity," *Proc. IEEE Int'l Symp. Inf. Theory*, Aachen, Germany (2017), 1853–1857.

[12] J.P. Tillich, G. Zémor, "Discrete isoperimetric inequalities and the probability of a decoding error," *Combin. Probab. Comput.*, 9 (2000), 465–479.

Erik Ordentlich (S'92–M'96–SM'06–F'11) received the S.B. and S.M. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, MA, in 1990, and the Ph.D. degree, also in electrical engineering, from Stanford University, Stanford, CA, in 1996.

During the years 1996–1999 and 2002–2014 he was a research scientist at Hewlett-Packard Laboratories, Palo Alto, CA and during the years 1999–2002 he was a staff engineer at iCompression, Inc., Santa Clara. Since 2014 he as been a scientist at Yahoo, Inc., now a part of Verizon Media Group. His work has addressed multiple topics in signal processing, information theory, and machine learning. He is a coinventor on numerous U.S. Patents and contributed technology to the ISO JPEG 2000 image compression standard.

Dr. Ordentlich was a corecipient of the 2006 IEEE Joint Communications–Information Theory Paper Award and is a member of Phi Beta Kappa and Tau Beta Pi. He served as Associate Editor for Source Coding for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2007–2010.

Ron M. Roth (M'88–SM'97–F'03) received the B.Sc. degree in computer engineering, the M.Sc. in electrical engineering, and the D.Sc. in computer science from Technion—Israel Institute of Technology, Haifa, Israel, in 1980, 1984, and 1988, respectively. Since 1988 he has been with the Computer Science Department at Technion, where he now holds the General Yaakov Dori Chair in Engineering. During the academic years 1989–91 he was a Visiting Scientist at IBM Research Division, Almaden Research Center, San Jose, California, and during 1996–97, 2004–05, and 2011–2012 he was on sabbatical leave at Hewlett–Packard Laboratories, Palo Alto, California. He is the author of the book *Introduction to Coding Theory*, published by Cambridge University Press in 2006. Dr. Roth was an associate editor for coding theory in IEEE TRANSACTIONS ON INFORMATION THEORY from 1998 till 2001, and he is now serving as an associate editor in *SIAM Journal on Discrete Mathematics*. His research interests include coding theory, information theory, and their application to the theory of complexity.