

Article

# The Fractality of Polar and Reed–Muller Codes <sup>†</sup>

Bernhard C. Geiger 

Signal Processing and Speech Communication Laboratory, Graz University of Technology, 8010 Graz, Austria; geiger@ieee.org; Tel.: +43-316-873-4367

<sup>†</sup> This paper is an extended version of our paper published in the 2015 NEWCOM# Emerging Topics in Modulation and Coding Workshop and in the 2016 International Zürich Seminar on Communications, Zurich, Switzerland, 2–4 March 2016.

Received: 16 October 2017 ; Accepted: 15 January 2018; Published: 17 January 2018

**Abstract:** The generator matrices of polar codes and Reed–Muller codes are submatrices of the Kronecker product of a lower-triangular binary square matrix. For polar codes, the submatrix is generated by selecting rows according to their Bhattacharyya parameter, which is related to the error probability of sequential decoding. For Reed–Muller codes, the submatrix is generated by selecting rows according to their Hamming weight. In this work, we investigate the properties of the index sets selecting those rows, in the limit as the blocklength tends to infinity. We compute the Lebesgue measure and the Hausdorff dimension of these sets. We furthermore show that these sets are finely structured and self-similar in a well-defined sense, i.e., they have properties that are common to fractals.

**Keywords:** polar codes; Reed–Muller codes; fractals; self-similarity

## 1. Introduction

In his book on fractal geometry, Falconer characterizes a set  $\mathcal{F}$  as a fractal if it has some of the following properties [1] (p. xxviii):

- $\mathcal{F}$  has a fine structure, i.e., there is detail on arbitrarily small scales
- $\mathcal{F}$  does not admit a description in traditional geometrical language, neither locally nor globally; it is irregular in some sense
- $\mathcal{F}$  has some form of self-similarity, at least approximate or statistical
- The fractal dimension of  $\mathcal{F}$  exceeds its topological dimension
- $\mathcal{F}$  is defined in a simple, often recursive way

In this work, we investigate whether polar codes and Reed–Muller are fractal in above sense. For a blocklength of  $2^n$ , these codes are based on the  $n$ -fold Kronecker product  $G(n) := F^{\otimes n}$ , where

$$F := \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad (1)$$

i.e., on a simple, recursive operation. Based on this, it has long been suspected that Kronecker product-based codes possess a fractal nature. For example, the authors of [2] observed that  $G(n)$ , when converted to a picture, resembles the Sierpinski triangle. In a personal communication [3], Abbe expressed his suspicion that the set of “good” polarized channels is fractal. Nevertheless, to the best of the author’s knowledge, a definite statement regarding this fractal nature has not been presented yet.

A rate- $K/2^n$  Kronecker product-based code is uniquely defined by a set  $\mathcal{F}$  of  $K$  indices: Its generator matrix is the submatrix of  $G(n)$  consisting of the rows indexed by  $\mathcal{F}$ . Letting  $\mathcal{F}$  index the  $K$  rows of  $G(n)$  with the largest Hamming weight defines a Reed–Muller code. Alternatively, one can

fix the order  $r$  of a Reed–Muller code, which defines  $\mathcal{F}$  as the index set of all rows with a Hamming weight at least as large as  $r$  (see Section 4). For polar codes, the rows of  $G(n)$  can be interpreted as a communication channels. Then, a rate- $K/2^n$  polar code is defined by the set  $\mathcal{F}$  indexing the  $K$  channels with the lowest Bhattacharyya parameters [4] (the “good” channels, see Section 2).

Although the sets  $\mathcal{F}$  are important for the construction of polar and Reed–Muller codes, surprisingly little is known about their fractal properties. Recently, Renes, Sutter, and Hassani stated conditions under which the good (bad) channels derived from one binary-input memoryless channel are also good (bad) for another channel [5]. Moreover, the authors of [6,7] observed the self-similar structure of  $\mathcal{F}$  by showing that polar and Reed–Muller codes are decreasing monomial codes.

In this paper, we analyze the fractal properties of  $\mathcal{F}$  for polar codes (Section 3) and Reed–Muller codes (Section 5). In contrast to [6,7], we study the properties of  $\mathcal{F}$  for infinite blocklengths, i.e., for  $n \rightarrow \infty$ . Specifically, we compute the Hausdorff dimension of  $\mathcal{F}$ , show that it is self-similar, and that it has detail on arbitrarily small scales (e.g.,  $\mathcal{F}$  is symmetric and dense in some well-defined containing set). Each of these results is relatively easy to obtain once appropriate definitions have been put in place. Taken as a whole, however, they paint an interesting picture and make a convincing point for the claim that polar and Reed–Muller codes are fractal.

The presented results will improve our understanding of polar and Reed–Muller codes, even though we have to admit that their practical implication (e.g., in code construction) still eludes us. Nevertheless, our results may apply in areas beyond channel coding: Arkan’s polarization technique was used to polarize Rényi information dimension [8] and to construct high-girth matrices [9]. Moreover, Nasser showed that a sufficient and necessary condition for a binary operation to be polarizing is that it is uniformity preserving and that its inverse is strongly ergodic [10,11]. We are convinced that fractality carries over to these applications as well and that an analysis similar to ours can deepen understanding.

Since we consider the case  $n \rightarrow \infty$ , the set  $\mathcal{F}$  indexes a subset of  $\Omega := \{0, 1\}^\infty$ , the set of infinite binary sequences. We let  $b := (b_1 b_2 \dots) \in \Omega$  and abbreviate  $b^n := (b_1 b_2 \dots b_n)$ . Let  $\bar{b} := (\bar{b}_1 \bar{b}_2 \dots)$  where  $\bar{b}_i := 1 - b_i$ . Let furthermore  $(\Omega, \mathfrak{A}, \mathbb{P})$  be a probability space with  $\mathfrak{A}$  the Borel field generated by the cylinder sets  $S(b^n) := \{w \in \Omega: w_1 = b_1, \dots, w_n = b_n\}$  and  $\mathbb{P}$  a probability measure satisfying  $\mathbb{P}(S(b^n)) = 1/2^n$ . In the following, we represent every infinite binary sequence  $b \in \Omega$  by a point in the unit interval  $[0, 1]$ . The mapping between  $\Omega$  and  $[0, 1]$  is given by

$$f(b) := \sum_{n=1}^{\infty} \frac{b_n}{2^n}. \tag{2}$$

**Lemma 1** ([12] (Exercises 7–10, p. 80)). *Let  $\mathfrak{B}_{[0,1]}$  be the Borel  $\sigma$ -algebra on  $[0, 1]$  and let  $\lambda$  be the Lebesgue measure. Let furthermore  $\mathbb{D} := [0, 1] \cap \{p/2^n: p \in \mathbb{Z}, n \in \mathbb{N}\}$  denote the set of dyadic rationals in the unit interval. Then, the function  $f: (\Omega, \mathfrak{A}) \rightarrow ([0, 1], \mathfrak{B}_{[0,1]})$  in (2) satisfies the following properties:*

1.  $f$  is measurable
2.  $f$  is bijective on  $\Omega \setminus f^{-1}(\mathbb{D})$
3. for all  $I \in \mathfrak{B}_{[0,1]}$ ,  $\mathbb{P}(f^{-1}(I)) = \lambda(I)$

**Example 1.** *Lemma 1 states that  $f$  is not injective in general. The reason is that dyadic rationals have a non-unique binary expansion. For example,  $f$  maps both  $b = (0111111 \dots)$  and  $b = (1000000 \dots)$  to 0.5, where we call the latter binary expansion terminating.*

## 2. Preliminaries for Polar Codes

Let  $W: \{0,1\} \rightarrow \mathcal{Y}$  be a binary-input memoryless channel with finite output alphabet  $\mathcal{Y}$ , (symmetric) capacity  $0 \leq I(W) \leq 1$ , and with Bhattacharyya parameter

$$Z(W) := \sum_{y \in \mathcal{Y}} \sqrt{W(y|0)W(y|1)}. \tag{3}$$

It can be shown using [13] (Proposition 1) that  $Z(W) = 0 \Leftrightarrow I(W) = 1$  and  $Z(W) = 1 \Leftrightarrow I(W) = 0$ . We say that the channel  $W$  is *symmetric* if there exists a permutation  $\pi: \mathcal{Y} \rightarrow \mathcal{Y}$  such that  $\pi^{-1} = \pi$  and, for every  $y \in \mathcal{Y}$ ,  $W(y|0) = W(\pi(y)|1)$ .

Arikan’s polarization technique [13] *combines and splits* two channel uses of  $W$  into one use of a “worse” channel

$$W_2^0(y_1^2|u_1) := \frac{1}{2} \sum_{u_2} W(y_1|u_1 \oplus u_2)W(y_2|u_2) \tag{4}$$

and one use of a “better” channel

$$W_2^1(y_1^2, u_1|u_2) := \frac{1}{2} W(y_1|u_1 \oplus u_2)W(y_2|u_2) \tag{5}$$

where  $u_1, u_2 \in \{0,1\}$  and  $y_1, y_2 \in \mathcal{Y}$ . The combining operation encodes two input bits by  $F$  in (1); transmitting them via two channel uses of  $W$  creates a vector channel. This vector channel can then be split into the two virtual binary-input memoryless channels indicated in (4) and (5). The better (worse) channel obtained by polarization has a larger (smaller) capacity than the original channel  $W$ , i.e.,  $I(W_2^0) \leq I(W) \leq I(W_2^1)$ —the inequalities are strict if  $0 < I(W) < 1$ . The sum capacity equals two times the capacity of the original channel, i.e.,  $I(W_2^0) + I(W_2^1) = 2I(W)$  [13] (Proposition 4). Similarly, polarization has an effect on the Bhattacharyya parameter:

**Lemma 2** (Bounds on the Bhattacharyya Parameter).

$$Z(W_2^1) = g_1(Z(W)) := Z^2(W) \leq Z(W) \tag{6}$$

$$Z(W) \leq Z(W) \sqrt{2 - Z^2(W)} =: h_0(Z(W)) \leq Z(W_2^0) \stackrel{(a)}{\leq} g_0(Z(W)) := 2Z(W) - Z^2(W) \tag{7}$$

with equality in (a) if  $W$  is a binary erasure channel (BEC).

**Proof.** The equality and inequality in (6) follow from [13] (Proposition 7) and the fact that  $Z(W) \leq 1$ , respectively. The inequalities in (7) follow from the fact that  $Z(W) \leq 1$ , from [14] (Lemma 20), and from [13] (Proposition 7). The last inequality becomes an equality if  $W$  is a BEC [13] (Proposition 7).  $\square$

For larger blocklengths  $2^n$ ,  $n > 1$ , we apply the polarization procedure recursively and obtain, for  $b^n \in \{0,1\}^n$ ,

$$\left( W_{2^n}^{b^n}, W_{2^n}^{b^n} \right) \rightarrow \left( W_{2^{n+1}}^{b^n 0}, W_{2^{n+1}}^{b^n 1} \right) \tag{8}$$

where  $b^n 0$  and  $b^n 1$  denote the sequences of zeros and ones obtained by appending 0 and 1 to  $b^n$ , respectively. Note that the functions  $g_1$ ,  $g_0$ , and  $h_0$  from Lemma 2 are non-negative and non-decreasing and map the unit interval onto itself. Hence, the inequalities in (7) are preserved under composition:

$$Z(W_{2^n}^{b^n}) \leq p_{b^n}(Z(W)) := g_{b_n}(g_{b_{n-1}}(\dots g_{b_1}(Z(W)) \dots)) \tag{9}$$

$$Z(W_{2^n}^{b^n}) \geq q_{b^n}(Z(W)) := h_{b_n}(h_{b_{n-1}}(\dots h_{b_1}(Z(W)) \dots)) \tag{10}$$

where  $h_1 \equiv g_1$ .

Applying this recursive polarization infinitely often leads to a situation in which almost all channels are either perfect or useless, i.e., either  $I(W_\infty^b) = 1$  or  $I(W_\infty^b) = 0$  for  $b \in \{0, 1\}^\infty$ . This is the assertion of Arkan’s polarization theorem:

**Proposition 1** ([13] (Proposition 10)). *With probability one, the limit RV  $I_\infty(b) := I(W_\infty^b)$  takes values in the set  $\{0, 1\}$ :  $\mathbb{P}(I_\infty = 1) = I(W)$  and  $\mathbb{P}(I_\infty = 0) = 1 - I(W)$ .*

If we stop the polarization procedure at a finite blocklength  $2^n$  for  $n$  large enough, then still most of the resulting  $2^n$  channels are either almost perfect or almost useless (i.e., the channel capacities are close to one or to zero). The idea of polar coding is to transmit data only on those channels that are almost perfect. The generator matrix of a blocklength- $2^n$  polar code is thus the submatrix of  $G(n)$  consisting of rows indexed by  $\mathcal{F}$ , where  $\mathcal{F}$  contains the indices corresponding to the  $K$  virtual channels with the largest capacities. Determining this set  $\mathcal{F}$  is inherently difficult, since (whenever  $W$  is not a BEC) the cardinality of the output alphabet increases exponentially in  $2^n$  [15] (Chapter 3.3), [16] (p. 36). Tal and Vardy proposed an approximate construction method based on reducing the output alphabet and showing that the resulting channels are either upgraded or degraded w.r.t. the channel of interest [17] (see also Korada’s PhD thesis [16] (Definition 1.7 & Lemma 1.8)). These upgrading/degrading properties are important tools in our proofs.

**Definition 1** (Channel Upgrading and Degrading). *A channel  $W^-: \{0, 1\} \rightarrow \mathcal{Z}$  is degraded w.r.t. the channel  $W$  (short:  $W^- \preceq W$ ) if there exists a channel  $Q: \mathcal{Y} \rightarrow \mathcal{Z}$  such that*

$$W^-(z|u) = \sum_{y \in \mathcal{Y}} W(y|u)Q(z|y). \tag{11}$$

*A channel  $W^+: \{0, 1\} \rightarrow \mathcal{Z}$  is upgraded w.r.t. the channel  $W$  (short:  $W^+ \succeq W$ ) if there exists a channel  $P: \mathcal{Z} \rightarrow \mathcal{Y}$  such that*

$$W(y|u) = \sum_{z \in \mathcal{Z}} W^+(z|u)P(y|z). \tag{12}$$

Moreover,  $W^+ \succeq W$  if and only if  $W \preceq W^+$ .

Upgraded (degraded) channels remains upgraded (degraded) during polarization:

**Lemma 3** ([16] (Lemma 4.7) & [17] (Lemmas 3 & 5)). *Suppose that  $W^- \preceq W \preceq W^+$ . Then,*

$$\begin{aligned} I(W^-) &\leq I(W) \leq I(W^+) \\ Z(W^-) &\geq Z(W) \geq Z(W^+) \\ (W^-)_2^1 &\preceq W_2^1 \preceq (W^+)_2^1 \\ (W^-)_2^0 &\preceq W_2^0 \preceq (W^+)_2^0. \end{aligned} \tag{13}$$

**Lemma 4** ([15] (p. 9) & [6] (Lemma 3)).  *$W \preceq W_2^1$ . If  $W$  is symmetric, then  $W_2^0 \preceq W \preceq W_2^1$ .*

**Proof.** By choosing

$$P(y|y_1^2, u_1) = \begin{cases} 1, & \text{if } y = y_2 \\ 0, & \text{else.} \end{cases} \tag{14}$$

one can show that  $W \preceq W_2^1$ . To show that also  $W_2^0 \preceq W$  for symmetric channels, take [6] (Lemma 3)

$$Q(y_1^2|y) = \begin{cases} \frac{1}{2}W(y_2|0) & \text{if } y_1 = y \\ \frac{1}{2}W(y_2|1) & \text{if } y_1 = \pi(y) \\ 0 & \text{else.} \end{cases} \tag{15}$$

□

### 3. Fractal Properties of the Sets of Good and Bad Channels

We next investigate the behavior of the set  $\mathcal{F}$  as we let the blocklength tend to infinity, i.e., as  $n \rightarrow \infty$ . This set indexes all sequences  $b$  for which we obtain  $I(W_\infty^b) = 1$ . With the help of (2), we map these sequences to a subset of the unit interval, which we will call the set of *good channels*.

**Definition 2** (The Good and the Bad Channels). *Let  $\mathcal{G}$  denote the set of good channels, i.e.,*

$$x \in \mathcal{G} \Leftrightarrow \exists b \in f^{-1}(x): I(W_\infty^b) = 1. \quad (16)$$

*Let  $\mathcal{B}$  denote the set of bad channels, i.e.,*

$$x \in \mathcal{B} \Leftrightarrow \exists b \in f^{-1}(x): I(W_\infty^b) = 0. \quad (17)$$

If  $I(W) = 0$ , then all polarized channels are useless and we have  $\mathcal{B} = [0, 1]$ . Similarly, if  $I(W) = 1$ , then all polarized channels are perfect and we have  $\mathcal{G} = [0, 1]$ . We hence assume throughout this section that the channel  $W$  is nontrivial, i.e., that  $0 < I(W) < 1$ .

**Proposition 2** (Denseness).  *$\mathcal{G} \cap \mathcal{B} = \mathbb{D}$ , i.e., the good and bad channels are dense in the unit interval. Moreover,  $\mathcal{G} \setminus \mathbb{D}$  and  $\mathcal{B} \setminus \mathbb{D}$  are dense in  $[0, 1]$ .*

**Proof.** See Appendix A.  $\square$

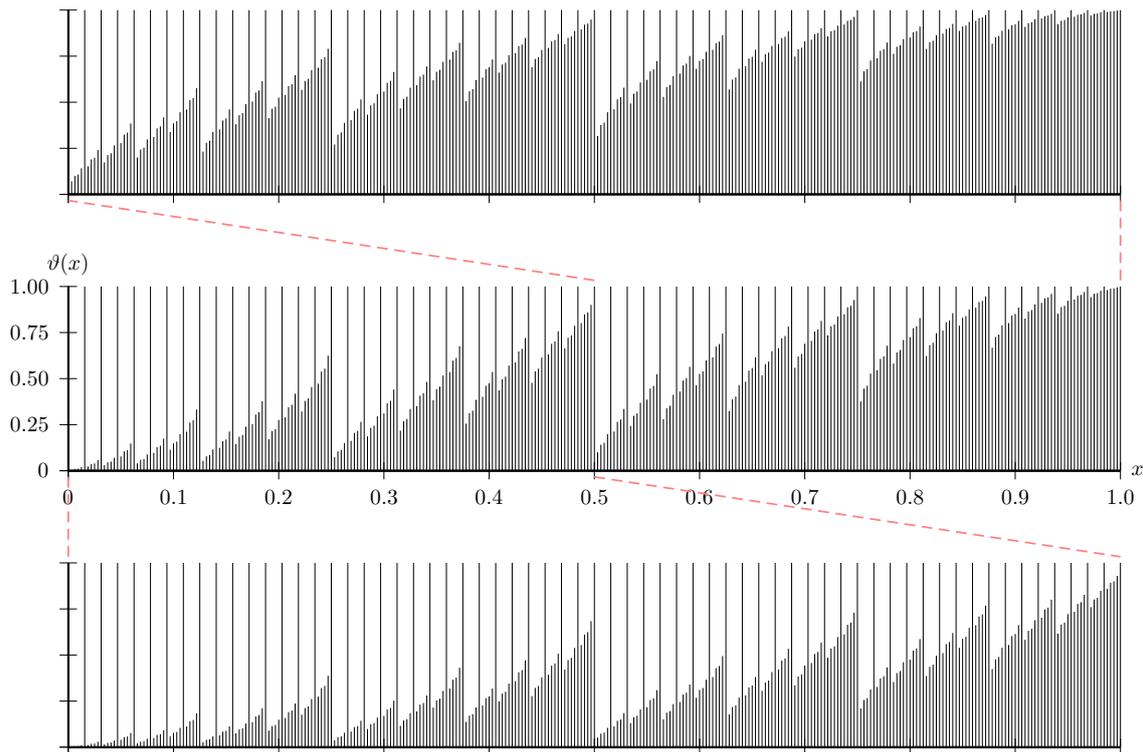
It is not really surprising that  $\mathcal{G}$  and  $\mathcal{B}$  are not disjoint; this is a direct consequence of the fact that  $f$  is not injective. It is not obvious, however, that the intersection exhausts the set on which  $f$  is non-injective. A consequence of this proposition is that there is no interval that contains only good channels. This has implications for code construction techniques. Indeed, the authors of [18,19] suggest that, for a polar code of a given blocklength, one may stop polarizing channels at shorter blocklengths and use copies of these channels rather than their polarization. For example, they suggest to use the channels  $(W_{2^n}^{b_n}, W_{2^n}^{b_n})$  rather than the channels  $(W_{2^{n+1}}^{b_n 0}, W_{2^{n+1}}^{b_n 1})$  if  $I(W_{2^n}^{b_n})$  is sufficiently large. Such a procedure can be justified if further polarizing  $W_{2^n}^{b_n}$  to the desired blocklength will lead to including all channels polarized from  $W_{2^n}^{b_n}$  in the code. Such a justification can never appear for polar codes with unbounded blocklength: Stopping polarizing at a given blocklength  $2^n$  for a given polarization sequence and using copies of the resulting channel  $W_{2^n}^{b_n}$  is equivalent to including a dyadic interval in the index set. This dyadic interval contains, by Proposition 2, bad channels, which shows that this choice is suboptimal.

**Proposition 3** (Symmetry). *There exists a function  $\vartheta$ , defined for almost all values in  $[0, 1]$ , that is independent of  $W$  and satisfies  $0 \leq \vartheta(x) \leq 1$  and  $\vartheta(1-x) = 1 - \vartheta(x)$ . Let  $x \in [0, 1]$  be such that  $\vartheta(x)$  is defined. Then,  $\vartheta(x) > Z(W)$  implies  $x \in \mathcal{G}$ . If  $W$  is a BEC, then  $\vartheta(x) < Z(W)$  implies  $x \in \mathcal{B}$ .*

**Proof.** See Appendix B.  $\square$

Proposition 3 has two implications. The first implication concerns the *alignment* of the sets  $\mathcal{G}$  and  $\mathcal{G}'$  for two different channels  $W$  and  $W'$ . Specifically, it is connected to the question whether  $Z(W) \geq Z(W')$  implies  $\mathcal{G} \subseteq \mathcal{G}'$ . In general, the answer is negative [5]. Indeed, it may happen that for some  $b \in \Omega$ , we have  $I(W_\infty^b) = 1$  despite  $Z(W) > \vartheta(f(b))$ , i.e., that the polarized channel turns out to be good even though the sufficient condition from Proposition 3 is not fulfilled. Such a situation cannot occur for BECs, as Proposition 3 shows. Hence, the set of good channels for a BEC is also good for any binary-input memoryless channel with a smaller Bhattacharyya parameter [20].

The second implication is that, at least for BECs, the sets  $\mathcal{G}$  and  $\mathcal{B}$  are symmetric. Indeed, if  $\vartheta(x) \neq Z(W)$ , then  $x \in \mathcal{G}$  implies  $1-x \in \mathcal{B}$ . This symmetry is visible in the polar fractal that we display in Figure 1.



**Figure 1.** The polar fractal. The center plot shows the thresholds  $\vartheta(x)$  for a finite set of values  $x \in [0, 1]$ ; the bottom and the top plots show thresholds for equally many values in the sets  $[0, 0.5]$  and  $[0.5, 1]$ , respectively. One can observe how the thresholds are ordered, i.e., thresholds in the top plot exceed those in the center plot, which exceed those in the bottom plot. For a binary erasure channel (BEC)  $W$ , the indicator function of  $\mathcal{G}$  is obtained by setting each value in the plot to one (zero) if the Bhattacharyya parameter  $Z(W)$  is smaller (larger) than the threshold. Note that this plot illustrates the behavior of  $\mathcal{G}$  in the limit  $n \rightarrow \infty$ . Note further that the figure illustrates the symmetry of  $\vartheta(x)$  claimed in Proposition 3. The MATLAB code to generate these thresholds is available as Supplementary Material.

It is possible to define  $\vartheta$  for  $x \in \mathbb{D}$ . We know from Proposition 2 that dyadic rationals are both good and bad, hence setting  $\vartheta(x) = 1$  for every  $x \in \mathbb{D}$  leads to  $\mathbb{D} \subseteq \mathcal{G}$ . (The fact that also  $\mathbb{D} \subseteq \mathcal{B}$  is not captured by nor in conflict with this setting.) The question whether the function  $\vartheta$  can be defined for  $x \in \mathbb{Q} \setminus \mathbb{D}$  is more interesting. In this case, the binary expansion is unique and *recurring*, i.e., there is a length- $k$  sequence  $a^k \in \{0, 1\}^k$  such that  $f(b^n a^k a^k \dots) = x$  for some  $b^n \in \{0, 1\}^n$ . It is straightforward to show that for every non-trivial sequence  $a_k$  (i.e.,  $a_k$  contains zeros and ones),  $p_{a^k}$  is from  $[0, 1]$  to  $[0, 1]$ , non-negative, and non-decreasing with vanishing derivatives at 0 and 1. Since this ensures that  $p_{a^k}(z) < z$  for  $z$  close to zero and  $p_{a^k}(z) > z$  for  $z$  close to one, the operation  $z_{i+1} = p_{a^k}(z_i)$  constitutes an iterated function system with attracting fixed points at  $z = 0$  and  $z = 1$ . Note further that, since  $p_{a^k}$  corresponds to the recurring part of the binary expansion of  $x$ ,  $Z(W_\infty^{b^n a^k a^k \dots})$  will be bounded from above by the value to which this iterated function system converges after being initialized with  $Z(W_{2^n}^{b^n})$ . To show that Proposition 3 holds for  $x \in \mathbb{Q} \setminus \mathbb{D}$  requires showing that  $p_{a^k}$  intersects the identity function only once on  $(0, 1)$ , i.e., that there is no attracting fixed point on this open interval. We leave this problem for future investigation.

**Example 2.** Let  $x = 2/3$ , hence  $f^{-1}(x) = 101010101 \dots$ . We determine the fixed points of the iterated function system corresponding to one period of the recurring sequence, i.e, the fixed points of  $p_{10}(z) = 2z^2 - z^4$ . These are given by the roots of  $p_{10}(z) - z$ , which are  $z = 0$ ,  $z = 1$ , and  $z = (\pm\sqrt{5} - 1)/2$ . One of these

latter nontrivial roots lies outside  $[0, 1]$  and is hence irrelevant. The remaining root determines the threshold,  $\vartheta(2/3) = (\sqrt{5} - 1)/2$ .

Now suppose that  $W$  is a BEC with Bhattacharyya parameter  $Z(W) = \vartheta(2/3)$ . Since  $\vartheta(2/3)$  is a fixed point, we get  $Z(W_\infty^{f^{-1}(2/3)}) = Z(W) \notin \{0, 1\}$ . This illustrates that Proposition 1 holds only almost surely.

**Proposition 4** (Lebesgue Measure & Hausdorff Dimension).  $\mathcal{G}$  is a Borel set and has Lebesgue measure  $\lambda(\mathcal{G}) = I(W)$ .  $\mathcal{B}$  is a Borel set and has Lebesgue measure  $\lambda(\mathcal{B}) = 1 - I(W)$ . Therefore, the Hausdorff dimensions of  $\mathcal{G}$  and  $\mathcal{B}$  satisfy  $d(\mathcal{G}) = d(\mathcal{B}) = 1$ .

**Proof.** See Appendix C.  $\square$

Loosely speaking, the Lebesgue measure of  $\mathcal{G}$  is the asymptotic equivalent of the rate of the “infinite-blocklength” polar code for the channel  $W$ . The fact that  $\lambda(\mathcal{G}) = I(W)$  states that the rate approaches the symmetric capacity of  $W$ . A positive Lebesgue measure and a Hausdorff dimension equal to one are not indicators of fractality.

The last fractal property we consider is self-similarity. As Falconer notes [1] (p. xxviii), self-similarity often occurs only approximately. What we show in the following proposition is that the set  $\mathcal{G}$  is quasi self-similar. Along the same lines, the quasi self-similarity of  $\mathcal{B}$  can be shown.

**Proposition 5** (Self-Similarity). Let  $\mathcal{G}_n(k) := \mathcal{G} \cap [(k - 1)2^{-n}, k2^{-n}]$  for  $k = 1, \dots, 2^n$ .  $\mathcal{G} = \mathcal{G}_0(1)$  is quasi self-similar in the sense that, for all  $n$  and all  $k$ ,  $\mathcal{G}_n(k) = \mathcal{G}_{n+1}(2k - 1) \cup \mathcal{G}_{n+1}(2k)$  is quasi self-similar to its right half:

$$\mathcal{G}_n(k) \subseteq 2\mathcal{G}_{n+1}(2k) - k2^{-n} \tag{18}$$

If  $W$  is symmetric,  $\mathcal{G}_n(k)$  is quasi self-similar:

$$2\mathcal{G}_{n+1}(2k - 1) - (k - 1)2^{-n} \subseteq \mathcal{G}_n(k) \subseteq 2\mathcal{G}_{n+1}(2k) - k2^{-n} \tag{19}$$

**Proof.** See Appendix D.  $\square$

In other words, at least for a symmetric channel,  $\mathcal{G}$  is composed of two similar copies of itself (see Figure 1). The self-similarity is closely related to the fact that polar codes are decreasing monomial codes [6] (Theorem 1).

**Example 3.** We want to determine whether  $1/3 \in \mathcal{G}$  for a given BEC  $W$ . This question translates the questions whether  $1/6 \in \mathcal{G}_1(1)$  and whether  $2/3 \in \mathcal{G}_1(2)$ . Along the lines of Example 2, we obtain  $\vartheta(1/6) \approx 0.214$ ,  $\vartheta(1/3) \approx 0.382$ , and  $\vartheta(2/3) \approx 0.618$ , i.e.,  $\vartheta(1/6) < \vartheta(1/3) < \vartheta(2/3)$ . Since  $W$  is a BEC, we can connect this with Proposition 3 and thus obtain the inclusion indicated in Proposition 5.

#### 4. Preliminaries for Reed–Muller Codes

An order- $r$ , length- $2^n$  Reed–Muller code is defined by having a generator matrix  $G_{RM}(r, n)$  composed of all length- $2^n$  sequences with a Hamming weight larger than  $2^{n-r}$ . For example, we have  $G_{RM}(n, n) = G(n)$ , while  $G_{RM}(0, n)$  is a single row vector containing only ones (length- $2^n$  repetition code). To make this more precise, let  $w(b^n) = \sum_{i=1}^n b_i$  be the Hamming weight of  $b^n \in \{0, 1\}^n$  and let  $s_i(n)$  be the  $i$ -th row of  $G(n)$ . Then, the generator matrix  $G_{RM}(r, n)$  of an order- $r$ , length- $2^n$  Reed–Muller code consists of the rows of  $G(n)$  indexed by [4]

$$\mathcal{F} = \{i \in \{1, \dots, 2^n\} : w(s_i(n)) \geq 2^{n-r}\}. \tag{20}$$

To analyze the effect of doubling the block length, note that

$$G(n + 1) := \begin{bmatrix} G(n) & 0 \\ G(n) & G(n) \end{bmatrix}. \tag{21}$$

Assume that we indicate the rows of  $G(n)$  by a sequence of binary numbers, i.e., let the  $i$ -th row be indexed by  $h_n(b^n) := 2^n \sum_{l=1}^n b_l 2^{-l}$ . Furthermore, let  $0b^n$  and  $1b^n$  denote the sequences of zeros and ones obtained by prepending 0 and 1 to  $b^n$ , respectively. Clearly,  $h_{n+1}(0b^n) = h_n(b^n)$  and  $h_{n+1}(1b^n) = h_n(b^n) + 2^n$ . Combining this with (21) yields

$$\begin{aligned} w(s_{h_{n+1}(0b^n)}(n + 1)) &= w(s_{h_n(b^n)}(n)) \\ w(s_{h_{n+1}(1b^n)}(n + 1)) &= 2w(s_{h_n(b^n)}(n)). \end{aligned}$$

Defining  $G(0) := 1$ , we thus get

$$w(s_{h_n(b^n)}(n)) = 2^{w(b^n)} \tag{22}$$

and

$$\mathcal{F} = h_n \left( \{b^n \in \{0, 1\}^n : 2^{w(b^n)} \geq 2^{n-r}\} \right). \tag{23}$$

In Section 5, we will analyze the properties of  $\mathcal{F}$  in the limit as  $n$  tends to infinity. An important ingredient in our proofs is the concept of *normal numbers*.

**Definition 3** (Normal Numbers). *A number  $x \in [0, 1]$  is called simply normal to base 2 ( $x \in \mathcal{N}$ ) if and only if*

$$\exists b \in f^{-1}(x) : \lim_{n \rightarrow \infty} \frac{w(b^n)}{n} = \frac{1}{2}. \tag{24}$$

In general, a number is simply normal to base  $M$  if the fraction of its digits used in its  $M$ -ary expansion is  $1/M$ . A number is called normal if this property not only holds for digits, but also for subsequences: a number is normal in base  $M$  if, for each  $k \geq 1$ , the fraction of each length- $k$  sequences used in its  $M$ -ary expansion is  $1/M^k$ . It immediately follows that a normal number is simply normal. The converse is in general not true:

**Example 4.** *Let  $x = 1/3$ , hence  $b = 010101 \dots$ .  $x$  is simply normal to base 2, but not normal (since the sequences 00 and 11 never occur). Let  $x = 1/7$ , hence  $b = 001001001 \dots$ .  $x$  is neither normal nor simply normal. Let  $x \in \mathbb{D}$ , hence  $b$  is either terminating ( $\lim_{n \rightarrow \infty} w(b^n)/n = 0$ ) or non-terminating ( $\lim_{n \rightarrow \infty} w(b^n)/n = 1$ ). Dyadic rationals are not simply normal.*

**Lemma 5** (Borel’s Law of Large Numbers, cf. [21] (Corollary 8.1, p. 70)). *Almost all numbers in  $[0, 1]$  are simply normal, i.e.,*

$$\lambda(\mathcal{N}) = 1. \tag{25}$$

Despite this result, there are uncountably many numbers in the unit interval which are not normal. Moreover, the set of numbers that are not normal is *superfractal*, i.e., it has a Hausdorff dimension equal to one although it has zero Lebesgue measure [22].

### 5. Fractal Properties of the Set of Heavy Codewords

If we let  $n$  tend to infinity, the definition of  $\mathcal{F}$  in (23) becomes problematic. Rather than looking at order- $r$ , length- $2^n$  Reed–Muller codes, we investigate order- $(1 - \rho)n$ , length- $2^n$  codes, where we assume that  $\rho n$  is integer. In other words, we assume that the threshold for the Hamming weight increases linearly with the blocklength. This gives rise to the definition of *heavy codewords*:

**Definition 4** (The Heavy Codewords). Let  $\mathcal{H}(\rho)$  denote the set of  $\rho$ -heavy codewords, i.e.,

$$x \in \mathcal{H}(\rho) \Leftrightarrow \exists b \in f^{-1}(x): \liminf_{n \rightarrow \infty} \frac{2^{w(b^n)}}{2^{n\rho}} \geq 1. \tag{26}$$

Loosely speaking, the set of heavy codewords corresponds to those rows of  $G(n)$  that asymptotically have a fractional Hamming weight larger than a given threshold.

**Example 5.**  $\mathcal{H}(1) = \{1\}$ . This follows from the fact that 1 is the only number in the unit interval with a binary expansion consisting only of ones.  $\mathcal{H}(0) = [0, 1]$ . This follows from the fact that  $w(b^n) \geq 0$ .

**Proposition 6** (Denseness). For all  $\rho \in [0, 1)$ ,  $\mathbb{D} \subset \mathcal{H}(\rho)$ . Moreover, for  $\rho \in (0, 1)$ ,  $\mathcal{H}(\rho) \setminus \mathbb{D}$  and its complement are dense in  $[0, 1]$ .

Similarly as for polar codes, also Reed–Muller codes are such that no interval is contained in either  $\mathcal{H}(\rho)$  or its complement (unless in the trivial cases  $\mathcal{H}(0)$  and  $\mathcal{H}(1)$ ). This is again in contrast with the intuition one obtains for Reed–Muller code with finite blocklength. Suppose we fix  $n$  to be even and set  $r = n/2$ , i.e., we require that at least one half of the bits in  $b^n$  are one. The matrix  $G(n)$  resembles a Sierpinski triangle, as depicted in [2] (Figure 2). In our notation, the set  $\mathcal{F}$  indexes none of the first  $2^{n/2} - 1$  rows of  $G(n)$ , since they cannot have sufficient Hamming weight. Consequently, the transition as  $n \rightarrow \infty$  creates complications that are not present for finite  $n$ , and one needs to depart from intuition based on these finite-blocklength considerations.

**Proof.** See Appendix E.  $\square$

**Proposition 7** (Lebesgue Measure & Hausdorff Dimension).  $\mathcal{H}(\rho)$  is Lebesgue measurable and has Lebesgue measure

$$\lambda(\mathcal{H}(\rho)) = \begin{cases} 1, & \text{if } \rho < 1/2 \\ 0, & \text{if } \rho \geq 1/2. \end{cases} \tag{27}$$

The Hausdorff dimension satisfies

$$d(\mathcal{H}(\rho)) \begin{cases} = 1, & \text{if } \rho \leq 1/2 \\ \geq h_2(\rho), & \text{if } \rho > 1/2 \end{cases} \tag{28}$$

where  $h_2(x) := -x \log_2 x - (1 - x) \log_2(1 - x)$ .

**Proof.** See Appendix F.  $\square$

Loosely speaking, the Lebesgue measure of  $\mathcal{H}(\rho)$  is the asymptotic equivalent of the rate of the fractional order- $\rho$  Reed–Muller code. As we showed in Proposition 4, the Lebesgue measure of  $\mathcal{G}$  is equal to the symmetric capacity of  $W$ . In contrast, the set  $\mathcal{H}(\rho)$  does not depend on  $W$ . Rather, Proposition 7 suggests that the order parameter  $\rho$  induces a phase transition for the rate of Reed–Muller codes: If  $\rho < 1/2$ , the “infinite-blocklength” Reed–Muller code consists of almost all (in the sense of Lebesgue measure) possible binary sequences. In contrast, if  $\rho \geq 1/2$ , the “infinite-blocklength” Reed–Muller code consists of almost no codewords (again, in the sense of Lebesgue measure).

Let us briefly consider the case  $\rho = 1/2$ . For this case, Proposition 7 states that  $\mathcal{H}(\rho)$  is a Lebesgue null set that has a Hausdorff dimension equal to 1. Thus, the set  $\mathcal{H}(1/2)$  is a superfractal. Unfortunately, we were not able to give an exact expression for the Hausdorff dimension of  $\mathcal{H}(\rho)$  for  $\rho > 1/2$ . While the set of all non-normal numbers is superfractal, we are not sure if this holds also for the specific proper subset  $\mathcal{H}(\rho)$ .

The sets  $\mathcal{G}$  and  $\mathcal{B}$  exhibit self-similarity, i.e., detailed structure on every scale (cf. Figure 1). We next show that also  $\mathcal{H}(\rho)$  is self-similar. At least for  $\mathcal{H}(0)$  and  $\mathcal{H}(1)$  (cf. Example 5) this is as trivial as the

self-similarity of a point or a line. For  $\rho \in (0, 1)$  this self-similarity is more interesting, and related to the fact that Reed–Muller codes are decreasing monomial codes [6] (Proposition 2).

**Proposition 8** (Self-Similarity). *Let  $\mathcal{H}_n(\rho, k) := \mathcal{H}(\rho) \cap [(k-1)2^{-n}, k2^{-n}]$  for  $k = 1, \dots, 2^n$ .  $\mathcal{H}(\rho) = \mathcal{H}_0(\rho, 1)$  is quasi self-similar in the sense that, for all  $n$  and all  $k$ ,  $\mathcal{H}_n(\rho, k) = \mathcal{H}_{n+1}(\rho, 2k-1) \cup \mathcal{H}_{n+1}(\rho, 2k)$  is quasi self-similar:*

$$2\mathcal{H}_{n+1}(\rho, 2k-1) - (k-1)2^{-n} \subseteq \mathcal{H}_n(\rho, k) \subseteq 2\mathcal{H}_{n+1}(\rho, 2k) - k2^{-n}. \quad (29)$$

**Proof.** See Appendix G.  $\square$

## 6. Discussion and Outlook

That Kronecker product-based codes possess fractal properties has long been suspected. The present manuscript contains several results that back this suspicion with solid mathematical analyses. Specifically, we assumed that the blocklength tends to infinity and investigated the properties of the set  $\mathcal{G}$  of virtual channels that are perfect and the set  $\mathcal{H}(\rho)$  of codewords that have a fractional Hamming weight no less than  $\rho$ . Since both polar codes and Reed–Muller codes are obtained by a simple, recursive procedure, it remains to investigate whether the sets  $\mathcal{G}$  and  $\mathcal{H}(\rho)$  satisfy any of the following properties [1] (p. xxviii):

- The set has a fine structure, i.e., there is detail on arbitrarily small scales;
- It does not admit a description in traditional geometrical language, neither locally nor globally; it is irregular in some sense;
- It has some form of self-similarity, at least approximate or statistical;
- The fractal dimension of the set exceeds its topological dimension.

Indeed, the sets  $\mathcal{G}$  and  $\mathcal{H}(\rho)$  possess a fine structure in the sense that they are dense in the unit interval, but that also their complements are dense in the unit interval (cf. Propositions 2 and 6). Therefore, at an arbitrarily small scale, the sets  $\mathcal{G}$  and  $\mathcal{H}(\rho)$  admit no simple description in geometrical language. Both of these sets are self-similar in a specific sense, as we outlined in Propositions 5 and 8. Finally, while  $\mathcal{G}$  has a fractal dimension of one (cf. Proposition 4), the set  $\mathcal{H}(\rho)$  has, for a certain range of  $\rho$ , a positive (fractional?) Hausdorff dimension despite being a Lebesgue null set. This result, which we proved in Proposition 7, is one of the defining properties of a fractal set.

One reviewer pointed out that our definition of  $\mathcal{H}(\rho)$  can be complemented by a different one. Specifically, while  $\mathcal{H}(\rho)$  indexes the codewords with a fractional Hamming weight not smaller than  $\rho$ , one could define a set  $\mathcal{H}'(R)$  indexing the codewords of a Reed–Muller code with rate  $R$ . In other words, while  $\mathcal{H}(\rho)$  is parameterized via the *fractional order* of the code,  $\mathcal{H}'(R)$  is parameterized via its rate. We expect that the Lebesgue measure of the (adequately defined) set  $\mathcal{H}'(R)$  should be  $R$  and that, thus, its Hausdorff dimension equals one. An appropriate definition of  $\mathcal{H}'(R)$  is tied to the set  $\mathcal{F}$  of a rate- $R$ , length- $2^n$  Reed–Muller code (such as is our Definition 4). Since finding such a definition has so far eluded us, we postpone this investigation to future work.

Another obvious extension of our work are non-binary polar and Reed–Muller codes. For example, consider an  $\ell \times \ell$  matrix with entries from  $\{0, \dots, q-1\}$ , where  $q$  is prime. One can show that this matrix is polarizing as long as it is not upper-triangular [15] (Theorem 5.2). We believe that our analysis can be replicated by considering the  $\ell$ -ary expansion of real numbers in  $[0, 1]$ . Along the same lines, it would be interesting to examine the properties of  $q$ -ary Reed–Muller codes, e.g., [23,24].

**Supplementary Materials:** Supplementary material are available online at [www.mdpi.com/1099-4300/20/1/70/s1](http://www.mdpi.com/1099-4300/20/1/70/s1).

**Acknowledgments:** The author thanks Emmanuel Abbe, Princeton University, and Hamed Hassani, University of Pennsylvania, for fruitful discussions and suggesting material. The author is particularly indebted to Jean-Pierre Tillich, French Institute for Research in Computer Science and Automation (INRIA), for helpful suggestions and generalizing Proposition 5. The author is also indebted anonymous reviewers pointing to a missing step in the proof of Proposition 4 and for extending Proposition 2. This work was supported by TU Graz

Open Access Publishing Fund, by the Erwin Schrödinger Fellowship J 3765 of the Austrian Science Fund, and by the German Ministry of Education and Research in the framework of an Alexander von Humboldt Professorship.

**Conflicts of Interest:** The author declares no conflict of interest.

### Appendix A. Proof of Proposition 2

That  $\mathcal{G} \cap \mathcal{B} \subseteq \mathbb{D}$  follows from the fact that only dyadic rationals have a non-unique binary expansion. In particular, the preimage of  $x \in \mathbb{D}$  consists of two elements, namely

$$b' := (b^n 0000000 \dots) \tag{A1}$$

and

$$b'' := (b^{n-1} \bar{b}_n 1111111 \dots). \tag{A2}$$

By the properties of polarization and with the assumption that  $0 < I(W) < 1$ , we have that  $0 < I(W_{2^n}^{b^{n-1} \bar{b}_n}), I(W_{2^n}^{b^n}) < 1$ , and hence also  $0 < Z(W_{2^n}^{b^{n-1} \bar{b}_n}), Z(W_{2^n}^{b^n}) < 1$ . Moreover, it follows from Lemma 2 and (9) that

$$Z(W_\infty^{b''}) \leq \lim_{\ell \rightarrow \infty} \left( Z(W_{2^n}^{b^{n-1} \bar{b}_n}) \right)^{2^\ell} = 0 \tag{A3}$$

from which we obtain  $I(W_\infty^{b''}) = 1$  and  $x \in \mathcal{G}$ . Similarly, with Lemma 2 and (10) we obtain

$$Z(W_\infty^{b'}) \geq \lim_{\ell \rightarrow \infty} \sqrt{1 - (1 - Z^2(W_{2^n}^{b^n}))^{2^\ell}} = 1. \tag{A4}$$

Hence,  $I(W_\infty^{b'}) = 0$  and  $x \in \mathcal{B}$ . Since this holds for every  $x \in \mathbb{D}$ , we have that  $\mathcal{G} \cap \mathcal{B} = \mathbb{D}$ .

The proof that  $\mathcal{G} \setminus \mathbb{D}$  and  $\mathcal{B} \setminus \mathbb{D}$  are dense in  $[0, 1]$  follows along similar lines. Specifically, we show that between every dyadic rational we can find rational numbers  $x, x' \in \mathbb{Q} \setminus \mathbb{D}$  such that  $x \in \mathcal{G}$  and  $x' \in \mathcal{B}$ . To this end, fix  $x_1 = p/2^n$  and  $x_2 = (p + 1)/2^n$ . Let further  $b^n$  be the terminating binary expansion of  $x_1$ , i.e.,  $f(b^n 000 \dots) = x_1$ .

Let  $a^k$  be such that  $a_1 = \dots = a_{k-1} = 1$  and  $a_k = 0$ . We can bound the polynomial  $p_{a^k}$  from above:

$$p_{a^k}(z) = 2z^{2^{k-1}} - z^{2^k} \leq 2z^{2^{k-1}}$$

The bound crosses  $z$  at  $z = 0$  and at  $z^* = 2^{-1/(2^{k-1}-1)}$ , where  $z^*$  can be made arbitrarily close to one for  $k$  sufficiently large. Now let  $z_{i+1} = p_{a^k}(z_i)$ , where  $z_0 = Z(W_{2^n}^{b^n})$ . It follows that  $Z(W_\infty^{b^n a^k a^k \dots}) \leq \lim_{i \rightarrow \infty} z_i$ . However,  $z_i \rightarrow 0$  if  $z_0 < z^*$ , hence if  $k$  is sufficiently large such that this holds, then  $Z(W_\infty^{b^n a^k a^k \dots}) \leq 0$ . Thus,  $I(W_\infty^{b^n a^k a^k \dots}) = 1$  and  $x = f(b^n a^k a^k \dots) \in \mathcal{G}$ .

Recall that  $\bar{a}^k$  is such that  $\bar{a}_1 = \dots = \bar{a}_{k-1} = 0$  and  $\bar{a}_k = 1$ . We next bound the polynomial  $q_{\bar{a}^k}$  from below:

$$q_{\bar{a}^k}(z) = 1 - (1 - z^2)^{2^{k-1}} \geq 1 - e^{-2^{k-1}z^2} \tag{A5}$$

which intersects  $z$  at  $z = 0$ , at some root  $z^*$  that can be made arbitrarily close to zero for  $k$  sufficiently large, and at some root  $z^\dagger < 1$  that tends to one if  $k$  becomes large. Note further that the slope of  $q_{\bar{a}^k}$  equals  $2^k z (1 - z^2)^{2^{k-1}-1}$ . By setting  $k$  sufficiently large, one can guarantee that this slope is smaller than one on the interval  $[z^\dagger, 1]$ . Now let  $z_{i+1} = q_{\bar{a}^k}(z_i)$ , where  $z_0 = Z(W_{2^n}^{b^n})$ . Suppose further that we have chosen  $k$  sufficiently large such that  $z_0 > z^*$  and that the slope of  $q_{\bar{a}^k}$  is smaller than one on the interval  $[z^\dagger, 1]$ . We know that  $q_{\bar{a}^k}(z) > z$  on the interval  $(z^*, z^\dagger)$ , since this is the case for the lower bound in (A5). However, since the slope of  $q_{\bar{a}^k}$  is smaller than one on the interval  $[z^\dagger, 1]$  and since  $q_{\bar{a}^k}$  intersects  $z$  at one, there can be no further intersection between  $q_{\bar{a}^k}$  and  $z$  on the interval  $[z^\dagger, 1]$ . Hence,  $q_{\bar{a}^k}(z) > z$  on the interval  $(z^*, 1]$ , and  $z_i \rightarrow 1$  since  $z_0 > z^*$ . Since furthermore  $Z(W_\infty^{b^n \bar{a}^k \bar{a}^k \dots}) \geq \lim_{i \rightarrow \infty} z_i$ , we obtain  $Z(W_\infty^{b^n \bar{a}^k \bar{a}^k \dots}) \geq 1$ . Thus,  $I(W_\infty^{b^n \bar{a}^k \bar{a}^k \dots}) = 0$  and  $x' = f(b^n \bar{a}^k \bar{a}^k \dots) \in \mathcal{B}$ .

Since both  $f(b^n a^k a^k a^k \dots)$  and  $f(b^n \bar{a}^k \bar{a}^k \bar{a}^k \dots)$  are in the interval  $(x_1, x_2)$ , this shows that between every two dyadic rationals, there are numbers  $x$  and  $x'$  that are good and bad. This proves that  $\mathcal{G} \setminus \mathbb{D}$  and  $\mathcal{B} \setminus \mathbb{D}$  are dense in  $[0, 1]$ .

**Appendix B. Proof of Proposition 3**

**Lemma A1** ([25] (Lemma 11)). *For  $\mathbb{P}$ -almost every realization  $b \in \Omega$ , there exists a point  $\theta(b) \in [0, 1]$  such that*

$$\lim_{n \rightarrow \infty} p_{b^n}(z) = \begin{cases} 0, & z \in [0, \theta(b)) \\ 1, & z \in (\theta(b), 1]. \end{cases} \tag{A6}$$

Furthermore, the thus constructed RV  $\theta$  is uniformly distributed on  $[0, 1]$ .

It can be easily verified that  $g_i(1 - z) = 1 - g_{1-i}(z)$  for  $i = 0, 1$ . Hence,

$$\begin{aligned} p_{b^n}(z) &= g_{b_n}(g_{b_{n-1}}(\dots g_{b_2}(g_{b_1}(z)) \dots)) \\ &= g_{b_n}(g_{b_{n-1}}(\dots g_{b_2}(1 - g_{\bar{b}_1}(1 - z)) \dots)) \\ &= g_{b_n}(g_{b_{n-1}}(\dots 1 - g_{\bar{b}_2}(g_{\bar{b}_1}(1 - z)) \dots)) \\ &= 1 - g_{\bar{b}_n}(g_{\bar{b}_{n-1}}(\dots g_{\bar{b}_2}(g_{\bar{b}_1}(1 - z)) \dots)) \\ &= 1 - p_{\bar{b}^n}(1 - z). \end{aligned}$$

Now suppose that  $b \notin f^{-1}(\mathbb{D})$  and that  $\theta(b)$  is defined. If  $0 \leq z < \theta(b)$ , then  $1 - \theta(b) < 1 - z \leq 1$ . Since, by Lemma A1,  $0 \leq z < \theta(b)$  implies  $p_{b^n}(z) \rightarrow 0$  and  $p_{\bar{b}^n}(1 - z) \rightarrow 1$ , we get  $\theta(\bar{b}) = 1 - \theta(b)$ . Now set  $\vartheta(f(b)) := \theta(b)$ . Since  $b + \bar{b} = 11111\dots$ , it follows from the linearity of  $f$  that  $f(b) + f(\bar{b}) = f(b + \bar{b}) = 1$ , i.e., if  $x \notin \mathbb{D}$  has binary expansion  $b$ , then  $1 - x$  has binary expansion  $\bar{b}$ . Therefore, for all  $x \notin \mathbb{D}$  for which  $\theta(f^{-1}(x))$  is defined, we have  $\vartheta(1 - x) = 1 - \vartheta(x)$ .

Recall that, by Lemma 2, we have

$$Z(W_{2^n}^{b^n}) \leq p_{b^n}(Z(W)) = g_{b_n}(g_{b_{n-1}}(\dots g_{b_1}(Z(W)) \dots)). \tag{A7}$$

If  $Z(W) < \vartheta(x)$ , then by Lemma A1,  $Z(W_\infty^b) \leq \lim_{n \rightarrow \infty} p_{b^n}(Z(W)) = 0$  and  $x \in \mathcal{G}$ . If  $W$  is a BEC, then (A7) holds with equality. Thus, by Lemma A1, if  $Z(W) > \vartheta(x)$ , then  $Z(W_\infty^b) = \lim_{n \rightarrow \infty} p_{b^n}(Z(W)) = 1$  and  $x \in \mathcal{B}$ .

**Appendix C. Proof of Proposition 4**

For the proof we utilize properties of  $f$  derived in the proof of [26] (Theorem 2.1, p. 7). Specifically, let  $E \subset \Omega$  be the set of all binary sequences with infinitely many zeros, i.e.,  $E$  contains the binary expansion of all numbers  $x \in [0, 1] \setminus \mathbb{D}$  and the terminating binary expansions of all numbers  $x \in \mathbb{D}$ . Since  $\Omega \setminus E$  is countable,  $E$  is a Borel set. The function  $f$  restricted to  $E$ ,  $f: E \rightarrow [0, 1]$ , is bijective and measurable (since  $f$  is measurable by Lemma 1). Finally, the inverse function  $f^{-1}: [0, 1] \rightarrow E$  is measurable, i.e., for all  $A \in \mathfrak{A}$ ,  $A \subseteq E$ , we have  $f(A) \in \mathfrak{B}_{[0,1]}$ .

The set  $E$  contains all binary sequences except for non-terminating expansions of dyadic rationals, which lead to good channels by the proof of Proposition 2. Thus, we have

$$\{b \in \Omega: I(W_\infty^b) = 0\} \cap E = \{b \in \Omega: I(W_\infty^b) = 0\}. \tag{A8}$$

Since this set has probability  $1 - I(W)$  by Proposition 1, it is a Borel set (otherwise, it would not be measurable). However, since  $f(\{b \in E: I(W_\infty^b) = 0\}) = \mathcal{B}$ , it follows that  $\mathcal{B}$  is a Borel set of  $[0, 1]$ . That  $\mathcal{G}$  is a Borel set can be shown along similar lines, with  $E$  containing all binary sequences with infinitely many ones.

Every Borel set is Lebesgue measurable. To evaluate the Lebesgue measure of  $\mathcal{B}$ , note that

$$f^{-1}(\mathcal{B}) = \{b \in \Omega: I(W_\infty^b) = 0\} \cup f^{-1}(\mathbb{D}). \tag{A9}$$

Since  $\mathcal{B} \in \mathfrak{B}_{[0,1]}$ , we get from Lemma 1 that  $\lambda(\mathcal{B}) = \mathbb{P}(f^{-1}(\mathcal{B}))$ . By the monotonicity and countable subadditivity of measures, we have

$$\mathbb{P}(\{b \in \Omega: I(W_\infty^b) = 0\}) \leq \mathbb{P}(f^{-1}(\mathcal{B})) \leq \mathbb{P}(\{b \in \Omega: I(W_\infty^b) = 0\}) + \underbrace{\mathbb{P}(f^{-1}(\mathbb{D}))}_{=\lambda(\mathbb{D})=0}. \tag{A10}$$

Hence, by Proposition 1,  $\lambda(\mathcal{B}) = \mathbb{P}(\{b \in \Omega: I(W_\infty^b) = 0\}) = 1 - I(W)$ . The proof for the set of good channels follows along the same lines.

Since the one-dimensional Hausdorff measure of a Borel set equals its Lebesgue measure [1] (Equation (3.4), p. 45), it immediately follows that  $\mathcal{G}$  and  $\mathcal{B}$  have a Hausdorff dimension equal to one.

#### Appendix D. Proof of Proposition 5

Since the dyadic rationals are self-similar and since  $\mathbb{D} \subset \mathcal{G}$  from Proposition 2, one has for all  $n$  and  $k$ ,

$$\mathcal{G}_n(k) \cap \mathbb{D} = 2(\mathcal{G}_{n+1}(2k) \cap \mathbb{D}) - k2^{-n}. \tag{A11}$$

We now treat those values in  $[0, 1]$  that are not dyadic rationals. If  $b_k^n = (b_1 b_2 \dots b_n)$  is the terminating binary expansion of  $(k - 1)2^{-n}$ , every value in  $[(k - 1)2^{-n}, k2^{-n}]$  has a binary expansion  $b_k^n a$  for some  $a \in \Omega$ , where  $b_n = 1$  if and only if  $(k - 1)$  is odd. Similarly, and since  $(2k - 1)$  is always odd, every value in  $[(2k - 1)2^{-n-1}, k2^{-n}]$  has a binary expansion  $b_k^n 1a'$  for some  $a' \in \Omega$ . Assume that  $a' = a$ . Then, by Lemmas 3 and 4,  $W_\infty^{b_k^n a} \preceq W_\infty^{b_k^n 1a}$  for all  $a$ . Hence, if  $f(b_k^n a) \in \mathcal{G}_n(k)$ , then  $f(b_k^n 1a) \in \mathcal{G}_{n+1}(2k)$ . It remains to show that  $2f(b_k^n 1a) - f(b_{k+1}^n) = f(b_k^n a)$ :

$$\begin{aligned} f(b_k^n a) + f(b_{k+1}^n) &= f(b_k^n) + 2^{-n}f(a) + f(b_{k+1}^n) \\ &= (k - 1)2^{-n} + 2^{-n}f(a) + k2^{-n} \\ &= (2k - 1)2^{-n} + 2^{-n}f(a) \\ &= 2(2k - 1)2^{-n-1} + 2 \times 2^{-n-1}f(a) \\ &= 2f(b_k^n 1) + 2 \times 2^{-n-1}f(a) \\ &= 2f(b_k^n 1a) \end{aligned}$$

**Proof for Symmetric Channels.** Since  $(2k - 2)$  is always even, every value in  $[(2k - 2)2^{-n-1}, (2k - 1)2^{-n-1}]$  has a binary expansion  $b_k^n 0a$  for some  $a \in \Omega$ . Then, by Lemmas 3 and 4,  $W_\infty^{b_k^n 0a} \preceq W_\infty^{b_k^n a}$  for all  $a$ . Hence, if  $f(b_k^n 0a) \in \mathcal{G}_{n+1}(2k)$ , then  $f(b_k^n a) \in \mathcal{G}_n(k)$ . It remains to show that  $2f(b_k^n 0a) - f(b_k^n) = f(b_k^n a)$ :

$$\begin{aligned} f(b_k^n a) + f(b_k^n) &= f(b_k^n) + 2^{-n}f(a) + f(b_k^n) \\ &= (k - 1)2^{-n} + 2^{-n}f(a) + (k - 1)2^{-n} \\ &= (2k - 2)2^{-n} + 2^{-n}f(a) \\ &= 2(2k - 2)2^{-n-1} + 2 \times 2^{-n-1}f(a) \\ &= 2f(b_k^n 0) + 2 \times 2^{-n-1}f(a) \\ &= 2f(b_k^n 0a) \end{aligned}$$

□

**Appendix E. Proof of Proposition 6**

Note that in Definition 4 we can take the binary logarithm on both sides of the inequality to get the condition

$$x \in \mathcal{H}(\rho) \Leftrightarrow \exists b \in f^{-1}(x): \liminf_{n \rightarrow \infty} w(b^n) - n\rho \geq 0. \tag{A12}$$

We first show that  $\mathbb{D} \subset \mathcal{H}(\rho)$  for every  $\rho \in [0, 1)$ . To this end, consider the non-terminating expansion of  $x \in \mathbb{D}$ , i.e., there is a  $b^k \in \{0, 1\}^k$  such that  $f(b^k 1111 \dots) = x$ . Hence,  $w(b^n) \geq n - k$  for  $n \geq k$  and, for  $\rho < 1$ ,

$$\liminf_{n \rightarrow \infty} w(b^n) - n\rho \geq \lim_{n \rightarrow \infty} n(1 - \rho) - k = \infty. \tag{A13}$$

We next show that also  $\mathcal{H}(\rho) \setminus \mathbb{D}$  is dense in  $[0, 1]$  for  $\rho < 1$ . We do so by showing that between any two dyadic rationals  $x_1 = p/2^n$  and  $x_2 = (p + 1)/2^n$ , there exists a rational number  $x \in \mathbb{Q} \setminus \mathbb{D}$  that is in  $\mathcal{H}(\rho)$ . Let  $b^\ell$  be the terminating binary expansion of  $x_1$ , i.e.,  $f(b^\ell 000 \dots) = x_1$ . Furthermore, let  $a^k$  be such that  $a_1 = \dots = a_{k-1} = 1$  and  $a_k = 0$ . We set  $b = b^\ell a^k a^k a^k \dots$  and get  $f(b) \in (x_1, x_2)$ . One can show that  $w(b^n) \geq n - \ell - 1 - \frac{n-\ell}{k}$ . Hence, choosing  $k > 1/(1 - \rho)$ ,  $\rho < 1$ , leads to

$$\liminf_{n \rightarrow \infty} w(b^n) - n\rho \geq \lim_{n \rightarrow \infty} n \left( 1 - \frac{1}{k} - \rho \right) - \ell - \frac{\ell}{k} - 1 = \infty. \tag{A14}$$

This proves that  $x \in \mathcal{H}(\rho)$  if  $\rho < 1$ , from which follows that  $\mathcal{H}(\rho) \setminus \mathbb{D}$  is dense in  $[0, 1]$ .

We finally show that there exists a non-dyadic rational number  $x \in (x_1, x_2)$  such that  $x \notin \mathcal{H}(\rho)$ . From this follows that also the complement of  $\mathcal{H}(\rho) \setminus \mathbb{D}$  is dense in  $[0, 1]$ . To this end, let  $\bar{a}^k$  be such that  $\bar{a}_1 = \dots = \bar{a}_{k-1} = 0$  and  $\bar{a}_k = 1$ . Set  $b = b^\ell \bar{a}^k \bar{a}^k \bar{a}^k \dots$ . One can show that  $w(b^n) \leq \frac{n}{k} + \ell - \frac{\ell}{k} + 1$ . Hence, choosing  $k > 1/\rho$ ,  $\rho > 0$ , leads to

$$\liminf_{n \rightarrow \infty} w(b^n) - n\rho \leq \lim_{n \rightarrow \infty} n \left( \frac{1}{k} - \rho \right) + \ell - \frac{\ell}{k} + 1 = -\infty. \tag{A15}$$

Therefore,  $x \notin \mathcal{H}(\rho)$  if  $\rho > 0$ .

**Appendix F. Proof of Proposition 7**

By Example 4, dyadic rationals are not simply normal, hence let  $\mathcal{N} \subset [0, 1] \setminus \mathbb{D}$  be the set of simply normal numbers in  $[0, 1]$ . Note that  $f$  is bijective on  $\mathcal{N}$  by Lemma 1. We furthermore have  $\lambda(\mathcal{N}) = 1$ , hence  $[0, 1] \setminus \mathcal{N}$  is a Lebesgue null set (both  $\mathcal{N}$  and its complement are measurable). Every subset of a Lebesgue null set is a Lebesgue null set and, a fortiori, Lebesgue measurable.

By Lemma 5 we have

$$\forall b \in f^{-1}(\mathcal{N}): w(b^n) = \frac{1}{2}n + o(n). \tag{A16}$$

Fix  $\rho$ . Then,

$$\liminf_{n \rightarrow \infty} w(b^n) - n\rho = \lim_{n \rightarrow \infty} n \left( \frac{1}{2} - \rho \right) + o(n). \tag{A17}$$

If  $\rho < 1/2$ , then this limit diverges to infinity, and hence  $\mathcal{N} \subset \mathcal{H}(\rho)$ . Thus,  $[0, 1] \setminus \mathcal{H}(\rho)$  is a subset of  $[0, 1] \setminus \mathcal{N}$ , hence measurable, from which measurability of  $\mathcal{H}(\rho)$  follows. Since  $\lambda(\mathcal{N}) = 1$ , we have  $\lambda(\mathcal{H}(\rho)) = 1$ . If  $\rho > 1/2$ , the limit diverges to minus infinity, and hence  $\mathcal{N} \not\subset \mathcal{H}(\rho)$ . Thus,  $\mathcal{H}(\rho) \subset [0, 1] \setminus \mathcal{N}$ , from which  $\lambda(\mathcal{H}(\rho)) = 0$  follows.

Now let  $\rho = 1/2$ . We define a random variable  $B$  on our probability space, such that for all  $b \in \Omega$ ,  $B(b) = b$ .  $B$  is a sequence of independent, identically distributed Bernoulli-1/2 random variables, i.e., for all  $i$  we have  $\mathbb{P}(B_i = 1) = \mathbb{P}(B_i = 0) = 1/2$ . We now evaluate

$$\mathbb{P} \left( \liminf_{n \rightarrow \infty} w(B^n) - \frac{n}{2} \geq 0 \right). \tag{A18}$$

Consider the simple random walk  $S_n := w(B^n) - \frac{n}{2}$ . Let  $N_0(n)$  be the number of zero crossings of the sequence  $S_1, \dots, S_n$  and let  $N_0(n, b)$  be the number of zero crossings corresponding to the realization  $b \in \Omega$ . The event  $\liminf_{n \rightarrow \infty} w(B^n) - \frac{n}{2} \geq 0$  can only happen if the realization of  $S_n$  corresponding to  $b$  has only finitely many zero crossings, i.e.,

$$\begin{aligned} \{b \in \Omega: \liminf_{n \rightarrow \infty} w(B^n) - \frac{n}{2} \geq 0\} &\subseteq \{b \in \Omega: \exists R \in \mathbb{N}_0: \lim_{n \rightarrow \infty} N_0(n, b) \leq R\} \\ &= \bigcup_{R=0}^{\infty} \{b \in \Omega: \lim_{n \rightarrow \infty} N_0(n, b) \leq R\} \\ &= \bigcup_{R=0}^{\infty} \liminf_{n \rightarrow \infty} \{b \in \Omega: N_0(n, b) \leq R\} \end{aligned}$$

and hence

$$\begin{aligned} &\mathbb{P}(\{b \in \Omega: \liminf_{n \rightarrow \infty} w(B^n) - \frac{n}{2} \geq 0\}) \\ &\leq \sum_{R=0}^{\infty} \mathbb{P}(\liminf_{n \rightarrow \infty} \{b \in \Omega: N_0(n, b) \leq R\}) \\ &\leq \sum_{R=0}^{\infty} \lim_{n \rightarrow \infty} \mathbb{P}(N_0(n) \leq R) \end{aligned} \tag{A19}$$

where the second inequality is due to Fatou’s lemma [27] (Lemma 1.28, p. 23).

With [28] (Chapter III.5, p. 84)

$$\mathbb{P}(N_0(n) = R) = 2\mathbb{P}(S_{2n+1} = 2R + 1) \tag{A20}$$

we get

$$\begin{aligned} \mathbb{P}(N_0(n) \leq R) &= 2 \sum_{r=0}^R \mathbb{P}(S_{2n+1} = 2r + 1) \\ &\stackrel{(a)}{=} 2 \sum_{r=0}^R \binom{2n+1}{n-r} 2^{-2n-1} \\ &\leq 2^{-2n} \sum_{r=0}^R \binom{2n+2}{n+1} \\ &= 2^{-2n} (R+1) \binom{2n+2}{n+1} \\ &\stackrel{(b)}{\leq} 2^{-2n} (R+1) e 2^{2n+2} \frac{1}{\sqrt{(n+1)\pi}} \\ &= \frac{4e(R+1)}{\sqrt{(n+1)\pi}} \end{aligned}$$

where (a) is [28] (Equation (2.2), p. 75) and (b) is due to Stirling’s approximation [29] (Equation (6.1.38), p. 257). Since this probability tends to zero as  $n \rightarrow \infty$ , we have by (A19)

$$\mathbb{P}\left(\liminf_{n \rightarrow \infty} w(B^n) - \frac{n}{2} \geq 0\right) = 0. \tag{A21}$$

Since the set inside the probability measure is thus measurable, we can apply the same reasoning as in the proof of Proposition 4 to argue that  $\mathcal{H}(1/2)$  is Lebesgue measurable. We then obtain  $\lambda(\mathcal{H}(1/2)) = 0$  which completes the proof for the Lebesgue measure.

We now turn to the proof of the Hausdorff dimension. For  $\rho < 1/2$ ,  $\mathcal{H}(\rho)$  has full Lebesgue measure. Since every Lebesgue measurable set has a Borel subset with the same Lebesgue measure, and since Hausdorff dimension is monotonic, we have  $d(\mathcal{H}(\rho)) = 1$  for  $\rho < 1/2$ .

For  $\rho \geq 1/2$ , we define

$$\tilde{\mathcal{N}}_{\xi} := \left\{ x \in [0, 1]: \exists b \in f^{-1}(x): \lim_{n \rightarrow \infty} \frac{w(b^n)}{n} = \xi \right\} \tag{A22}$$

for some  $\xi \in (0, 1)$ . Note that  $\tilde{\mathcal{N}}_{1/2} = \mathcal{N}$ . By [30] (cf. [21] (Chapter 8) for further notes), the Hausdorff dimension of this set is given by (Interestingly, in Eggleston’s paper, the dimension was not connected to entropy; it was submitted earlier in the same year as Shannon’s Mathematical Theory of Communication was published).

$$d(\tilde{\mathcal{N}}_{\rho}) = h_2(\xi). \tag{A23}$$

Reasoning as in the proof for the Lebesgue measure,  $\tilde{\mathcal{N}}_{\xi} \subset \mathcal{H}(\rho)$  if  $\xi > \rho$  and  $\tilde{\mathcal{N}}_{\rho} \not\subset \mathcal{H}(\rho)$  if  $\xi < \rho$ . As a consequence,

$$\bigcup_{n=1}^{\infty} \tilde{\mathcal{N}}_{\rho+1/n} \subset \mathcal{H}(\rho). \tag{A24}$$

For a countable sequence of sets  $A_n$ , Hausdorff dimension satisfies [1] (p. 49)

$$d\left(\bigcup_{n=1}^{\infty} A_n\right) = \sup_{n \geq 1} d(A_n) \tag{A25}$$

and hence, by the monotonicity of Hausdorff dimension [1] (p. 48),

$$d(\mathcal{H}(\rho)) \geq \sup_{n \geq 1} h_2(\rho + 1/n) = h_2(\rho) \tag{A26}$$

where the last equality follows from the fact that the binary entropy function decreases with increasing  $\rho$  for  $\rho \geq 1/2$ . In particular, for  $\rho = 1/2$ ,  $d(\mathcal{H}(\rho)) = 1$ . This completes the proof.

### Appendix G. Proof of Proposition 8

The proof follows along the lines of the proof of Proposition 5. Let again  $b_k^n$  be the terminating expansion of  $(k - 1)2^{-n}$  and let  $a \in \Omega$ . The connections between the sequences  $b := b_k^n a$ ,  $b_- := b_k^n 0a$ , and  $b_+ := b_k^n 1a$  have been established above. To prove the theorem, we have to show that

$$\liminf_{m \rightarrow \infty} w(b_-^m) - \rho m \geq 0 \tag{A27}$$

$$\Rightarrow \liminf_{m \rightarrow \infty} w(b^m) - \rho m \geq 0 \tag{A28}$$

$$\Rightarrow \liminf_{m \rightarrow \infty} w(b_+^m) - \rho m \geq 0. \tag{A29}$$

This is obtained by

$$\begin{aligned} \liminf_{m \rightarrow \infty} w(b_-^m) - \rho m &= w(b_k^n 0) - \rho(n + 1) + \liminf_{m \rightarrow \infty} w(a^m) - \rho m \\ &= w(b_k^n) - \rho n - \rho + \liminf_{m \rightarrow \infty} w(a^m) - \rho m \\ &\leq w(b_k^n) - \rho n + \liminf_{m \rightarrow \infty} w(a^m) - \rho m \end{aligned} \tag{A30}$$

$$\begin{aligned} &\leq w(b_k^n) - \rho n + (1 - \rho) + \liminf_{m \rightarrow \infty} w(a^m) - \rho m \\ &= w(b_k^n 1) - \rho(n + 1) + \liminf_{m \rightarrow \infty} w(a^m) - \rho m \end{aligned} \tag{A31}$$

where (A30) equals (A28) and where (A31) equals (A29). The inequalities yield the desired result.

## References

1. Falconer, K. *Fractal Geometry: Mathematical Foundations and Applications*, 3rd ed.; John Wiley & Sons: Chichester, UK, 2014.
2. Kahraman, S.; Viterbo, E.; Çelebi, M.E. Folded Tree Maximum-Likelihood Decoder for Kronecker Product-based Codes. In Proceedings of the Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, 2–4 October 2013, pp. 629–636.
3. Abbe, E. (Princeton University, Princeton, NJ, USA). Personal communication, 2011.
4. Arikan, E. A Performance Comparison of Polar Codes and Reed–Muller Codes. *IEEE Commun. Lett.* **2008**, *12*, 447–449.
5. Renes, J.M.; Sutter, D.; Hassani, S.H. Alignment of Polarized Sets. In Proceedings of the 2015 IEEE International Symposium on Information Theory (ISIT), Hong Kong, China, 14–19 June 2015; pp. 2446–2450.
6. Bardet, M.; Dragoi, V.; Otmani, A.; Tillich, J.P. Algebraic Properties of Polar Codes From a New Polynomial Formalism. *arXiv* **2016**, arXiv:1601.06215v2.
7. Bardet, M.; Dragoi, V.; Otmani, A.; Tillich, J.P. Algebraic properties of polar codes from a new polynomial formalism. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Barcelona, Spain, 10–15 July 2016; pp. 230–234.
8. Haghhighatshoar, S.; Abbe, E. Polarization of the Rényi information dimension for single and multi terminal analog compression. In Proceedings of the 2013 IEEE International Symposium on Information Theory Proceedings (ISIT), Istanbul, Turkey, 7–12 July 2013, pp. 779–783.
9. Abbe, E.; Wigderson, Y. High-Girth matrices and polarization. In Proceedings of the IEEE International Symposium on Information Theory Proceedings (ISIT), Hong Kong, China, 14–19 June 2015; pp. 2461–2465.
10. Nasser, R. An Ergodic Theory of Binary Operations—Part I: Key Properties. *IEEE Trans. Inf. Theory* **2016**, *62*, 6931–6952.
11. Nasser, R. An Ergodic Theory of Binary Operations—Part II: Applications to Polarization. *IEEE Trans. Inf. Theory* **2017**, *63*, 1063–1083.
12. Taylor, M. *Measure Theory and Integration*; Graduate Studies In Mathematics Series 76; American Mathematical Society: Providence, RI, USA, 2006.
13. Arikan, E. Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Trans. Inf. Theory* **2009**, *55*, 3051–3073.
14. Korada, S.B.; Urbanke, R.L. Polar Codes are Optimal for Lossy Source Coding. *IEEE Trans. Inf. Theory* **2010**, *56*, 1751–1768.
15. Şaşıoğlu, E. Polarization and Polar Codes; *Found. Trends Commun. Inf. Theory* **2011**, *8*, 259–381.
16. Korada, S.B. Polar Codes for Channel and Source Coding. Ph.D. Thesis, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2009.
17. Tal, I.; Vardy, A. How to Construct Polar Codes. *IEEE Trans. Inf. Theory* **2013**, *59*, 6562–6582.
18. El-Khamy, M.; Mahdavifar, H.; Feygin, G.; Lee, J.; Kang, I. Relaxed channel polarization for reduced complexity polar coding. In Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC), New Orleans, LA, USA, 9–12 March 2015; pp. 207–212.
19. El-Khamy, M.; Mahdavifar, H.; Feygin, G.; Lee, J.; Kang, I. Relaxed Polar Codes. *IEEE Trans. Inf. Theory* **2017**, *63*, 1986–2000.
20. Hassani, S.H.; Korada, S.; Urbanke, R. The compound capacity of polar codes. In Proceedings of the Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, 30 September–2 October 2009; pp. 16–21.
21. Kuipers, I.; Niederreiter, H. *Uniform Distribution of Sequences*; John Wiley & Sons: New York, NY, USA, 1974.
22. Albeverio, S.; Pratsiuvityi, M.; Torbin, G. Topological and fractal properties of real numbers which are not normal. *Bull. Sci. Math.* **2005**, *129*, 615–630.
23. Kasami, T.; Lin, S.; Peterson, W. New generalizations of the Reed–Muller codes—I: Primitive codes. *IEEE Trans. Inf. Theory* **1968**, *14*, 189–199.
24. Delsarte, P.; Goethals, J.; Williams, F.M. On generalized Reed–Muller codes and their relatives. *Inf. Control* **1970**, *16*, 403–442.
25. Hassani, S.H.; Alishahi, K.; Urbanke, R.L. Finite-Length Scaling for Polar Codes. *IEEE Trans. Inf. Theory* **2014**, *60*, 5875–5898.

26. Parthasarathy, K.R. *Probability Measures on Metric Spaces*; Academic Press: New York, NY, USA, 1967.
27. Rudin, W. *Real and Complex Analysis*, 3rd ed.; McGraw-Hill: New York, NY, USA, 1987.
28. Feller, W. *An Introduction to Probability Theory and Its Applications*, 3rd ed.; John Wiley & Sons: New York, NY, USA, 1968; Volume 1.
29. Abramowitz, M.; Stegun, I.A., Eds. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed.; Dover Publications: New York, NY, USA, 1972.
30. Eggleston, H.G. The Fractional Dimension of a Set defined by decimal properties. *Q. J. Math.* **1949**, *os-20*, 31–36.



© 2018 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).