

## FRACTAL DIMENSIONS OF $k$ -AUTOMATIC SETS

ALEXI BLOCK GORMAN AND CHRIS SCHULZ

**Abstract.** This paper seeks to build on the extensive connections that have arisen between automata theory, combinatorics on words, fractal geometry, and model theory. Results in this paper establish a characterization for the behavior of the fractal geometry of “ $k$ -automatic” sets, subsets of  $[0, 1]^d$  that are recognized by Büchi automata. The primary tools for building this characterization include the entropy of a regular language and the digraph structure of an automaton. Via an analysis of the strongly connected components of such a structure, we give an algorithmic description of the box-counting dimension, Hausdorff dimension, and Hausdorff measure of the corresponding subset of the unit box. Applications to definability in model-theoretic expansions of the real additive group are laid out as well.

### §1. Introduction.

**1.1. Main results.** In this paper, we consider the  $k$ -automatic subsets of  $\mathbb{R}$ , and analyze both the  $k$ -representations of such sets and the Büchi automata that recognize their base- $k$  representations. The methods used in this paper integrate multiple perspectives previously taken regarding Büchi automata and  $k$ -automatic subsets of finite-dimensional Euclidean spaces. These include the perspective given by viewing automata as directed graphs, as well as characterizations of  $k$ -regular  $\omega$ -languages coming from combinatorics on words.

Our primary result describes how to obtain the Hausdorff and box-counting dimensions of  $k$ -automatic subsets of  $[0, 1]^d \subseteq \mathbb{R}^d$  (with  $d \in \mathbb{N}$ ) not quite in terms of some of the induced subautomata, but by considering slight variants thereof. Further results include a similar mechanism for computing Hausdorff measure (in the appropriate dimension) in terms of the same variant of an induced subautomaton, as well as a characterization of which expansions of the first order structure  $(\mathbb{R}, <, +)$  by  $k$ -automatic subsets of  $[0, 1]^d$  have definable unary sets whose Hausdorff dimension differs from its box-counting dimension.

Structures of this form are of some interest to tame geometers since they are “well-behaved” from the perspectives of topology and computability theory. Though they often fail to have stability and its generalizations (we will elaborate below), they retain the decidability found in the real additive group (see [3, Theorem 6]). It is moreover rare to find examples of structures with some geometric notion of tameness in which there are definable sets whose Hausdorff and box-counting dimensions differ.

---

Received May 13, 2022.

2020 *Mathematics Subject Classification.* Primary 03D05, Secondary 28A80.

*Key words and phrases.* Büchi automata, finite automata, fractal geometry, Hausdorff dimension, Hausdorff measure, box-counting dimension, entropy, model theory, tame geometry.

© The Author(s), 2023. Published by Cambridge University Press on behalf of The Association for Symbolic Logic.  
0022-4812/24/8903-0007  
DOI:10.1017/jsl.2023.55



Recall that an automaton is “trim” if each state is accessible from some start state, and each state is also coaccessible from some accept state. Below, we use  $d_H$  to denote Hausdorff dimension, we use  $d_B$  to denote box-counting dimension, and  $h(X)$  denotes the entropy of  $X$ ; for formal definitions of each, see Section 2. We show that as a corollary of recent work of Evans [9], if  $X$  is a closed  $k$ -automatic subset of  $[0, 1]^n$  recognized by closed automaton  $\mathcal{A}$ , then  $d_H(X) = d_B(X) = \frac{1}{\log(k)}h(L(\mathcal{A}))$ , where  $L(\mathcal{A})$  is the set of strings  $\mathcal{A}$  recognizes.

To state our main theorem, we will briefly describe the “cycle language” associated with a state  $q$  in an automaton  $\mathcal{A}$ , the definition of which is stated with more detail in Section 5. Suppose that  $\mathcal{A}$  is an automaton (finite or Büchi) with  $Q$  as its set of states. For state  $q \in Q$ , the cycle language  $C_q(\mathcal{A})$  is the language consisting of all  $w \in \Sigma^*$  such that there is a run of  $\mathcal{A}$  from state  $q$  to itself via  $w$ .

**THEOREM A.** *Let  $\mathcal{A}$  be a trim Büchi automaton with set of states  $Q$ , and let  $X$  be the set of elements in  $[0, 1]^d \subseteq \mathbb{R}^d$  that have a base- $k$  representation that  $\mathcal{A}$  accepts. Let  $F$  be the set of accept states of  $\mathcal{A}$ . Then:*

- (i)  $d_H(X) = \frac{1}{\log k} \max_{q \in F} h(C_q(\mathcal{A}))$ .
- (ii)  $d_B(X) = \frac{1}{\log k} \max_{q \in Q} h(C_q(\mathcal{A}))$ .

From this theorem, and using the crucial notion of “unambiguous” Büchi automata, we establish a similar result that describes the Hausdorff measure of a  $k$ -automatic set in terms of the structure of the automaton that recognizes it. For a definition of unambiguous automata and details about the partition  $\{M_q : q \in Q'\}$  of the language  $L$  in the theorem below, see Section 6.

**THEOREM B.** *Let  $\mathcal{A}$  be an unambiguous Büchi automaton with set of states  $Q$  and recognizing an  $\omega$ -language  $L$ . Let  $Q' \subseteq Q$  be the set of states whose strongly connected component contains an accept state. For each  $q \in Q'$ , let  $\mathcal{A}_q$  be the automaton created by moving the start state of  $\mathcal{A}$  to  $q$  and removing all transitions out of its strongly connected component, and let  $L_q$  be the  $\omega$ -language it accepts. Then we can effectively partition  $L$  into sublanguages  $\{M_q : q \in Q'\}$  such that:*

- (i)  $d_H(v_k(L)) = \max_{q \in Q'} d_H(v_k(L_q))$ ,
- (ii) with  $\alpha = d_H(v_k(L))$ ,  $\mu_H^\alpha(L) = \sum_{q \in Q'} \mu_H^\alpha(M_q)$ .

Finally, we give a dividing line for the fractal dimensions of definable sets in certain first-order structures related to Büchi automata. This dividing line has implications for the model-theoretic tameness of structures of the form  $(\mathbb{R}, <, +, X)$  where  $X \subseteq [0, 1]^n$  is  $k$ -automatic, since Hieronymi and Walsberg have shown in [12] that if  $\mathcal{C}$  is a Cantor set (a subset of  $\mathbb{R}$  that is compact, has no interior, and has no isolated points), then  $(\mathbb{R}, <, +, \mathcal{C})$  is not tame with respect to any notion coming from Shelah-style generalizations of stability, including NIP and  $NTP_2$ .

**THEOREM C.** *Suppose  $X \subseteq [0, 1]^n$  is  $k$ -automatic. There exists a set  $A \subseteq [0, 1]$  definable in  $(\mathbb{R}, <, +, X)$  such that  $d_B(A) \neq d_H(A)$  if and only if either a Cantor set is definable in  $(\mathbb{R}, <, +, X)$  or a set that is both dense and codense on an interval is definable in  $(\mathbb{R}, <, +, X)$ .*

**1.2. Background.** In his seminal work [4], Büchi introduced the notion of what we now call a Büchi automaton, and he identified a connection between these

automata and the monadic second-order theory of the natural numbers with the successor function. Notably, Büchi automata take countably infinite-length inputs, unlike standard automata (which we will also call “finite automata”), which only accept or reject finite-length input strings. In addition to the work of Büchi to extend the notion of automatic sets to infinite words, McNaughton broadened the realm of generating infinite sequences by a finite automaton in [15], and many more notions of automatic or regular sets of infinite words arose.

There is natural topological structure on the space of infinite words on a finite alphabet; hence, the topological features of subsets of such a space that is recognized by an appropriate Büchi automaton have been investigated since the 1980s. Languages recognized by Büchi automata are commonly called regular  $\omega$ -languages. One topological property that was first introduced in the context of information theory by Shannon in [19] is that of entropy, also called “topological entropy” in some settings.

Staiger established in [20] that extending the definition of entropy to  $\omega$ -languages yields compelling topological characterizations of closed regular  $\omega$ -languages. For example, he shows in [20] that a closed regular  $\omega$ -language is countable if and only if the entropy is 0 and that the entropy of regular  $\omega$ -languages is countably additive. From another perspective, in [6], the authors Charlier, Leroy, and Rigo show that there is a close connection between regular  $\omega$ -languages and Graph Directed Iterated Function Systems, or GDIFSs for brevity. Due to the work in [14], there have long existed means of computing geometric properties like Hausdorff measure and Hausdorff dimension for GDIFSs.

In light of the connection between Büchi automata and GDIFSs, the connections between fractal dimensions and entropy for automatic sets of real numbers can now yield a total characterization, as our paper illuminates. The work of Staiger and Charlier, Leroy, and Rigo focus on computing fractal dimensions and equality of fractal dimensions in the case of either closed or deterministic  $k$ -automatic sets (with the Cantor space as the setting for Staiger and the reals as the setting for Charlier, Leroy, and Rigo), and in this paper we extend such results to the non-deterministic and non-closed cases, while also illustrating precisely when the equality of such dimensions fails. The new methods we introduce also are applied to computing Hausdorff measure, which has been relatively explored for  $k$ -automatic sets of real numbers.

## §2. Preliminaries.

**2.1. Definition of Büchi automata.** Below, for a set  $X$  we use  $X^*$  to denote the Kleene star of  $X$ , i.e.,  $X^* = \{x_1x_2 \dots x_n : n \in \mathbb{N}, x_1, \dots, x_n \in X\}$ , and we similarly use  $X^\omega$  to denote the set  $\{x_1x_2 \dots : x_1, x_2, \dots \in X\}$ . For a language  $L$  of finite strings, we will use  $\bar{L}$  to denote the limit language of  $L$ , i.e., the set of infinite strings with infinitely many prefixes in  $L$ .

**DEFINITION 2.1.** A *finite automaton* is a 5-tuple  $\mathcal{A} = (Q, \Sigma, \delta, S, F)$  where:

- $Q$ , the set of *states*, is a finite set;
- $\Sigma$ , the *alphabet*, is a finite set;
- $\delta$ , the *transition function*, is a function  $Q \times \Sigma \rightarrow \mathcal{P}(Q)$ ;

- $S$ , the set of *start states* or *initial states*, is a nonempty subset of  $Q$ ;
- $F$ , the set of *accept states* or *final states*, is a subset of  $Q$ .

A finite automaton is said to *run from  $q_0$  to  $q_n$  on a string  $w = w_1 \dots w_n \in \Sigma^n$* , for  $q_0, q_n \in Q$ , if there exist states  $q_1, \dots, q_{n-1}$  such that for  $i = 1, \dots, n$  we have  $q_i \in \delta(q_{i-1}, w_i)$ . If  $q_0 \in S$ , such a sequence of states may be called a *run* of  $w$  in  $\mathcal{A}$ , which is *accepting* if  $q_n \in F$ . The automaton *accepts*  $w$  if there is an accepting run of  $w$ . The language *recognized* (or *accepted*) by  $\mathcal{A}$  is the set of all strings in  $\Sigma^*$  it accepts. Two finite automata are *equivalent* if they recognize the same language.

DEFINITION 2.2. A *Büchi automaton* is a 5-tuple  $\mathcal{A} = (Q, \Sigma, \delta, S, F)$  where:

- $Q$ , the set of *states*, is a finite set;
- $\Sigma$ , the *alphabet*, is a finite set;
- $\delta$ , the *transition function*, is a function  $Q \times \Sigma \rightarrow \mathcal{P}(Q)$ ;
- $S$ , the set of *start states* or *initial states*, is a nonempty subset of  $Q$ ;
- $F$ , the set of *accept states* or *final states*, is a subset of  $Q$ .

For an infinite string  $w = w_1 w_2 \dots \in \Sigma^\omega$ , a *run* of  $w$  in  $\mathcal{A}$  is a sequence of states  $q_0, q_1, \dots \in Q^\omega$  such that  $q_0 \in S$  and for  $i \in \mathbb{Z}^+$  we have  $q_i \in \delta(q_{i-1}, w_i)$ . A run is *accepting* if  $q_i \in F$  for infinitely many  $i$ . The automaton *accepts*  $w$  if there is an accepting run of  $w$ . The  $\omega$ -language *recognized* (or *accepted*) by  $\mathcal{A}$  is the set of all strings in  $\Sigma^\omega$  it accepts. Two Büchi automata are *equivalent* if they recognize the same language.

Note that the only difference between these definitions is in the accept condition; thus, the same tuple  $(Q, \Sigma, \delta, S, F)$  may be alternately treated as either a finite or Büchi automaton, which will be useful several times in this paper. A finite or Büchi automaton also has a canonical digraph structure whose vertex set is  $Q$  and whose edge set contains precisely those  $(q, q') \in Q^2$  for which there exists  $\sigma \in \Sigma$  such that  $q' \in \delta(q, \sigma)$ . We will often implicitly refer to this digraph structure, speaking of such concepts as paths between states and strongly connected components containing states. If we refer to the graphical structure on an automaton as simply a graph, we implicitly mean the structure of the automaton as a directed graph.

We will also use several properties that such an automaton may have:

DEFINITION 2.3. Let  $\mathcal{A} = (Q, \Sigma, \delta, S, F)$  be a finite or Büchi automaton.

- (i) We say  $\mathcal{A}$  is *deterministic* if  $|S| = 1$  and  $|\delta(q, c)| \leq 1$  for all  $q \in Q, c \in \Sigma$ . (Note that this definition guarantees that there is at most one run of a given  $w$  in  $\mathcal{A}$ .) Every finite automaton has an equivalent deterministic automaton; this is not true in general for Büchi automata.
- (ii) We say  $\mathcal{A}$  is *finite-trim* if for every  $q \in Q$ , there is a path from a start state to  $q$  (possibly of zero length) and a path from  $q$  to an accept state (also possibly of zero length). On the additional condition that the path from  $q$  to an accept state must be of nonzero length, we say that  $\mathcal{A}$  is *trim*. Every Büchi automaton has an equivalent trim automaton; every finite automaton has an equivalent finite-trim automaton. In fact, given an automaton  $\mathcal{A}$ , we may always produce a finite-trim automaton that is equivalent to  $\mathcal{A}$  as both a finite and Büchi automaton.
- (iii) We say  $\mathcal{A}$  is *closed* if it is trim and every state is an accept state (i.e.,  $Q = F$ ).

- (iv) Given a trim automaton  $\mathcal{A} = (Q, \Sigma, \delta, S, F)$ , call  $\overline{\mathcal{A}} = (Q, \Sigma, \delta, S, Q)$  (this is the resulting automaton when all the states of  $\mathcal{A}$  are added to the set of accept states) the *closure* of  $\mathcal{A}$ . Note that if an automaton  $\mathcal{B} = (Q', \Sigma, \delta', S', F')$  is equivalent to  $\mathcal{A}$  but not trim, then  $(Q', \Sigma, \delta', S', Q')$  need not recognize the same language as  $\overline{\mathcal{A}}$ .
- (v) We say an automaton  $\mathcal{A} = (Q, \Sigma, \delta, S, F)$  is *weak* if for every  $q, q' \in Q$  such that  $q$  and  $q'$  are in the same strongly connected component of  $\mathcal{A}$  (as a digraph), either  $q$  and  $q'$  are both accept states, or both are not accept states.

**2.2. Regularity and  $k$ -representations.** Let  $k \in \mathbb{N}_{>1}$ , and set  $[k] = \{0, 1, \dots, k - 1\}$  for the remainder of this paper. We will use the terms “base- $k$  representation” and “ $k$ -representation” interchangeably to mean the expression of an element  $x \in \mathbb{R}$  as a countable sum of integer powers of  $k$ , each multiplied by a coefficient in  $[k]$ . Note that we will sometimes conflate elements of  $[0, 1]$  and their  $k$ -representations, and we may occasionally say that an automaton  $\mathcal{A}$  accepts *the*  $k$ -representation of  $x \in [0, 1]$ . For the countable subset of  $[0, 1]^d$  whose elements have multiple (in particular, at most  $2^d$ )  $k$ -representations, we mean that  $\mathcal{A}$  accepts *at least one* of the  $k$ -representations of  $x$ . For ease of switching between  $x$  and its  $k$ -representation, we will define a valuation for elements of  $[k]^\omega$ .

DEFINITION 2.4. Define  $v_k : [k]^\omega \rightarrow [0, 1]$  by

$$v_k(w) = \sum_{i=0}^{\infty} \frac{w_i}{k^{i+1}},$$

where  $w = w_0w_1w_2 \dots$  with  $w_i \in [k]$  for each  $i \in \mathbb{N}$ .

Note that the equivalence relation  $v \equiv w \iff v_k(v) = v_k(w)$  is not only a finite equivalence relation, but moreover each equivalence class has size at most two. As noted above, only countably many elements in  $[k]^\omega$  are not the unique element of their  $v_k$ -equivalence class. For  $L \subseteq ([k]^d)^\omega$ , set

$$v_k(L) = \{(v_k(w_1), \dots, v_k(w_d)) : w_1, \dots, w_d \in [k]^\omega, ((w_{1,i}, \dots, w_{d,i}))_{i < \omega} \in L\}.$$

We can now formally define what it means for a subset of  $[0, 1] \subseteq \mathbb{R}$  to be  $k$ -automatic. Let  $k \in \mathbb{N}$  be greater than one, and let  $d \in \mathbb{N}$  be greater than zero.

DEFINITION 2.5. Say that  $L \subseteq ([k]^d)^\omega$  is  $k$ -regular if there is some Büchi automaton  $\mathcal{A}$  with alphabet  $[k]^d$  that recognizes  $L$ . Say that  $A \subseteq [0, 1]^d$  is  $k$ -automatic if there is a Büchi automaton  $\mathcal{A}$  with alphabet  $[k]^d$  that recognizes the maximal language  $L \subseteq ([k]^d)^\omega$  such that  $A = v_k(L)$ . Moreover, if this holds, say that  $\mathcal{A}$  recognizes  $A$ .

We also use the notation that for a Büchi automaton  $\mathcal{A}$  with alphabet  $[k]$ ,  $V_k(\mathcal{A})$  will denote the set of elements  $x \in [0, 1]$  for which some  $k$ -representation of  $x$  is accepted by  $\mathcal{A}$ .

The fact below follows immediately from the existence of a Büchi automaton with alphabet  $\Sigma^2$  that accepts a pair of elements  $x, y \in \Sigma^\omega$  precisely if both are  $k$ -representations of the same element of  $[0, 1]$ .

**FACT 2.6.** For  $A \subseteq [0, 1]^d$ , if there is some  $k$ -regular language  $L \subseteq ([k]^d)^\omega$  such that  $A = v_k(L)$ , then the set of all  $k$ -representations of elements of  $A$  is  $k$ -regular as well.

Call an element  $x \in [0, 1]^d$  a  $k$ -rational if there exists  $w \in ([k]^d)^*$  such that  $x = v_k(w\vec{0}^\omega)$ , where  $\vec{0}$  is the  $d$ -tuple  $(0, \dots, 0)$ . Clearly, these are the elements of  $[0, 1]^d$  whose coordinates can all be written as fractions with powers of  $k$  in the denominators.

Throughout this paper  $d$  denotes the (finite, but arbitrary) arity of the Euclidean space we are working in. We use  $\mathcal{A}$  to denote both finite automata and Büchi automata, and we use  $L$  to denote the subset of  $([k]^d)^*$  that  $\mathcal{A}$  recognizes if it is a finite automaton, or to denote the subset of  $([k]^d)^\omega$  that  $\mathcal{A}$  recognizes if it is a Büchi automaton. If  $\mathcal{A}$  is a Büchi automaton, we will often use  $A$  to denote  $v_k(L)$ , unless specified otherwise. We will say that a Büchi automaton  $\mathcal{A}$  accepts  $x \in [0, 1]^d$  if  $\mathcal{A}$  accepts some  $w \in ([k]^d)^\omega$  such that  $v_k(w) = x$ .

Given  $\mathcal{A}$  a trim Büchi automaton we let  $\overline{\mathcal{A}}$  denote the image under  $v_k$  of the language that  $\overline{\mathcal{A}}$ , the closure of  $\mathcal{A}$ , recognizes. In [6], the authors show that every closed trim Büchi automaton recognizes a (topologically) closed set, hence the conflation of the set recognized by  $\overline{\mathcal{A}}$  with  $\overline{\mathcal{A}}$ . This conflation will be further justified in Section 4. In addition, we define closed  $k$ -automatic  $\omega$ -languages and the closure of a  $k$ -automatic  $\omega$ -language analogously. If  $L$  is a language (either a subset of  $\Sigma^*$  or a subset of  $\Sigma^\omega$ ) let  $L^{pre} \subseteq \Sigma^*$  denote the set of all finite prefixes of elements of  $L$ . Similarly, let  $L_n^{pre}$  denote the set of all length- $n$  prefixes of elements of  $L$ , and let  $L_{<n}^{pre}$  denote the set of all prefixes of  $L$  with length at most  $n$ .

**2.3. Definition of entropy.** A key concept that turns out to be very helpful in the study of dimension of  $k$ -automatic sets is the notion of *entropy*. The entropy of a formal language was perhaps first used for regular languages by Chomsky and Miller in [7] and was called entropy as an analogue for topological entropy by Hansen, Perrin, and Simon in [11]. In [1], the authors note a seeming correspondence between the Hausdorff dimension of a  $k$ -automatic fractal and the entropy of the language of substrings of the base- $k$  expansions of its points. Proving this conjecture is one of the main results of this paper. In order to do so, we find it most convenient to extend the definition of entropy to sequences of real numbers as follows:

**DEFINITION 2.7.** Let  $(a_n)_{n \in \mathbb{N}}$  be a sequence of nonnegative real numbers such that infinitely many terms are nonzero and such that  $a_n \in O(k^n)$  for some  $k$ . The *entropy* of  $a_n$  is defined as the limit superior:

$$h((a_n)_n) = \limsup_{n \rightarrow \infty} \frac{\log a_n}{n}.$$

The entropy  $h(L)$  of an infinite language  $L$  is the entropy of  $(|L|_n)_n$ .

We choose to leave the entropy undefined for  $a_n$  eventually zero and  $a_n$  growing faster than exponentially, as this way the entropy is always a real number (and is nonnegative if  $a_n$  is an integer sequence), which simplifies some results regarding entropy.

**2.4. Definition of box-counting dimension.** There are two different notions of dimension that will play large roles in this paper. The first is the concept of

*box-counting dimension*, also known as *Minkowski dimension*. Intuitively, this is defined by quantifying how the number of boxes required to cover a given set increases as the size of the boxes decreases. This matches our intuition regarding “nice” sets that have a well-defined length, area, etc. For instance, it is natural that to cover a polygonal area of  $\mathbb{R}^2$  with boxes, when the boxes are half the size, this will require four times as many boxes. The box-counting dimension of such a polygon is  $\frac{\log 4}{\log 2} = 2$ .

In order to fully formalize this notion, many decisions must be made about the details. Is the “size” of a box its diameter or its side length? Must we use boxes, or could we use another shape, like a closed ball, instead? What if we allow the covering sets to be *any* set of a given diameter? Should we place restrictions on the positioning of each box, such as requiring them to come from a grid? It turns out that most of these decisions have no effect on the resulting notion of dimension, i.e., they are equivalent. Therefore, we use one of the several versions of the definition given in [10]:

DEFINITION 2.8 [10, Section 3.1]. Let  $X \subseteq \mathbb{R}^d$  be nonempty and bounded.

- (i) We define  $N(X, \varepsilon, \vec{r})$  to be the number of sets of the form  $I_{\vec{z}} = [z_1\varepsilon + r_1, (z_1 + 1)\varepsilon + r_1] \times \cdots \times [z_d\varepsilon + r_d, (z_d + 1)\varepsilon + r_d]$ , where  $\vec{z} = (z_1, \dots, z_d)$  are integers, required to cover  $X$ .
- (ii) The *upper box-counting dimension* of  $X$  is

$$d_{\overline{B}}(X) = \limsup_{\varepsilon \rightarrow 0} \sup_{\vec{r}} \frac{\log N(X, \varepsilon, \vec{r})}{\log \frac{1}{\varepsilon}}.$$

- (iii) The *lower box-counting dimension* of  $X$  is

$$d_{\underline{B}}(X) = \liminf_{\varepsilon \rightarrow 0} \inf_{\vec{r}} \frac{\log N(X, \varepsilon, \vec{r})}{\log \frac{1}{\varepsilon}}.$$

- (iv) If the upper and lower box-counting dimensions of  $X$  are equal, we refer to their value as simply *the box-counting dimension*  $d_B(X)$ .

There are several properties of the box-counting dimension that justify the use of the word “dimension” in the above definition. These include the following:

FACT 2.9 [10, Section 3.2]. Let  $X \subseteq \mathbb{R}^d$  be nonempty and bounded.

- (i) If  $X$  is a smooth  $n$ -manifold (embedded in  $\mathbb{R}^d$ ), then  $d_B(X) = n$ .
- (ii) If  $X \subseteq Y$ , then  $d_{\overline{B}}(X) \leq d_{\overline{B}}(Y)$  and  $d_{\underline{B}}(X) \leq d_{\underline{B}}(Y)$ .
- (iii) If  $X = Y_1 \cup Y_2$ , then  $d_{\overline{B}}(X) = \max(d_{\overline{B}}(Y_1), d_{\overline{B}}(Y_2))$ .
- (iv) Invertible affine transformations of  $\mathbb{R}^d$  preserve  $d_{\overline{B}}$  and  $d_{\underline{B}}$ .

In addition to these, box-counting dimension has one more property that turns out to be quite useful (and that other notions of dimension do not possess):

FACT 2.10 [10, Proposition 3.4]. Let  $X \subseteq \mathbb{R}^d$  be nonempty and bounded. Then  $d_{\overline{B}}(X) = d_{\overline{B}}(\overline{X})$ , and  $d_{\underline{B}}(X) = d_{\underline{B}}(\overline{X})$ .

**2.5. Definition of Hausdorff dimension.** Hausdorff dimension is the other notion of dimension that we will use in this paper. It is considerably more popular within



fractal geometry, probably due to its compatibility with measure-theoretic notions. To define Hausdorff dimension, we must first define the notion of Hausdorff *measure*, a family of outer measures on subsets of  $\mathbb{R}^d$ :

DEFINITION 2.11 [10, Section 2.1]. Let  $X$  be a nonempty Borel subset of  $\mathbb{R}^d$ . For  $s \geq 0, \varepsilon > 0$ , we define

$$\mu_H^s(X, \varepsilon) = \inf \left\{ \sum_{i=1}^{\infty} (\text{Diam } U_i)^s : \{U_i\}_i \text{ is a collection sets of diameter at most } \varepsilon \text{ covering } X \right\}.$$

The  $s$ -dimensional Hausdorff measure of  $X$ ,  $\mu_H^s(X)$ , is the limit of  $\mu_H^s(X, \varepsilon)$  as  $\varepsilon \rightarrow 0$ .

One precaution: recall that when we defined box-counting dimension above, we mentioned that it does not matter if the covering sets are boxes, balls, or any set with a given diameter. This is not the case with Hausdorff measure. Although the Hausdorff *dimension* of  $X$  would ultimately be the same if we changed these details in the above definition, the measure itself could be different.

Note that for subsets of  $\mathbb{R}$  with Hausdorff measure one, the Hausdorff measure agrees with the Lebesgue measure. A given set  $X$  will only have a “meaningful” (i.e., nonzero and finite) Hausdorff measure for at most one value of  $s$ . Consider once more the example of a polygon in  $\mathbb{R}^2$ . The two-dimensional Hausdorff measure of a polygon is, up to a constant factor of  $\frac{\pi}{4}$ , its area. But the  $s$ -dimensional Hausdorff measure will be infinite for any  $s < 2$  and zero for any  $s > 2$ . This suggests the following definition of Hausdorff dimension:

DEFINITION 2.12 [10, Section 2.2]. For  $X \subseteq \mathbb{R}^d$  nonempty, the Hausdorff dimension  $d_H(X)$  is the unique real number  $s$  such that  $\mu_H^{s'}(X) = \infty$  for  $s' < s$  and  $\mu_H^{s'}(X) = 0$  for  $s' > s$ .

Note that when  $s' = s$ , the Hausdorff measure may or may not be finite and may or may not be zero. What matters for determining dimension is the limiting behavior on either side of the critical value.

Hausdorff dimension has the properties that we expect any notion of dimension to have. These include the following:

- FACT 2.13 [10, Section 2.2]. (i) If  $X$  is a smooth  $n$ -manifold (embedded in  $\mathbb{R}^d$ ), then  $d_H(X) = n$ .
- (ii) If  $X \subseteq Y$ , then  $d_H(X) \leq d_H(Y)$ .
- (iii) If  $X = Y_1 \cup Y_2$ , then  $d_H(X) = \max(d_H(Y_1), d_H(Y_2))$ .
- (iv) Invertible affine transformations of  $\mathbb{R}^d$  preserve  $d_H$ .

In fact, Hausdorff dimension satisfies a stronger version of (iii) above:

FACT 2.14 [10, Section 2.2]. If  $X = \bigcup_{i \in \mathbb{N}} Y_i$ , then  $d_H(X) = \max_{i \in \mathbb{N}} d_H(Y_i)$ .

As a corollary, the Hausdorff dimension of any countable set is zero (as that of a point is zero), so Hausdorff dimension is invariant under the addition or removal of a countable set of points. This is very unlike box-counting dimension: note that box-counting dimension is preserved under closures. Since  $\mathbb{R}^d$  is separable, stability under closures and stability under the addition of countably many points are properties directly at odds with each other.



In particular, consider the set  $X = \mathbb{Q} \cap [0, 1]$ . The closure of  $X$  is the interval  $[0, 1]$ , which has box-counting dimension 1; hence,  $X$  has box-counting dimension 1. Yet  $X$  is countable and thus has Hausdorff dimension 0. This gives an explicit example of when Hausdorff and box-counting dimension may differ. Note, however, that when they do differ, it is always the box-counting dimension that is higher:

FACT 2.15 [10, Equation 3.17]. *Let  $X$  be a nonempty and bounded subset of  $\mathbb{R}^d$ . Then  $d_H(X) \leq d_B(X) \leq d_{\overline{B}}(X)$ .*

**§3. Entropy and its relationship to dimension.**

**3.1. Properties of entropy.** It will be helpful to establish several properties of entropy before connecting it to fractal dimension. First, we need the following results from [20] concerning the monotonicity of entropy under the subset relation and union operation, and  $\omega$ -language prefixes.

FACT 3.1 [20, Proposition 1]. *Let  $L_1$  and  $L_2$  be infinite languages.*

- (i) *If  $L_1 \subseteq L_2$ , then  $h(L_1) \leq h(L_2)$ .*
- (ii)  *$h(L_1 \cup L_2) = \max(h(L_1), h(L_2))$ .*

FACT 3.2 [20]. *Let  $L$  be an infinite language. Then*

$$h(L^{pre}) = \limsup_{n \rightarrow \infty} \frac{\log |L|_{\leq n}}{n} = h(L).$$

In the case where  $L$  is closed under prefixes, we can define its entropy to be a limit, rather than a limit superior. This fact is used in Corollary 9 of [20], but not proven explicitly. Hence, we include the following result for completeness.

LEMMA 3.3. *Let  $L$  be a regular language closed under prefixes. Then  $\lim_{n \rightarrow \infty} \frac{\log |L|_n}{n}$  exists.*

PROOF. Let  $\Sigma$  be the alphabet of  $L$ .

By Theorem 13 of [17], there exists a constant  $c$  and an increasing sequence  $(n_i)_{i \in \mathbb{N}}$  such that  $0 < n_{i+1} - n_i \leq c$  and such that  $\left(\frac{\log |L|_{n_i}}{n_i}\right)_i$  converges.

For any  $n \in \mathbb{N}$  with  $n \geq n_1$ , let  $i(n)$  be the index such that  $n_{i(n)} \leq n < n_{i(n)+1}$ . Note that  $n - n_{i(n)} \leq c$ , and  $n_{i(n)+1} - n \leq c$ . Set  $d_{n+} := n_{i(n)+1} - n$  and  $d_{n-} := n - n_{i(n)}$ .

Let  $k_1 < k_2 \in \mathbb{N}$ . Since  $L$  is prefix-closed, any string of length  $k_2$  in  $L$  is a string of length  $k_1$  in  $L$ , prepended to one of the  $|\Sigma|^{k_2-k_1}$  strings of length  $k_2 - k_1$  over the alphabet  $\Sigma$ . Thus  $|L|_{k_2} \leq |\Sigma|^{k_2-k_1} |L|_{k_1}$ .

Therefore, given any  $n \geq n_1$ ,

$$\begin{aligned} & \frac{1}{|\Sigma|^{d_{n+}}} |L|_{n_{i(n)+1}} \leq |L|_n \leq |\Sigma|^{d_{n-}} |L|_{n_{i(n)}} \\ \implies & \frac{1}{|\Sigma|^c} |L|_{n_{i(n)+1}} \leq |L|_n \leq |\Sigma|^c |L|_{n_{i(n)}} \\ \implies & \log |L|_{n_{i(n)+1}} - c \log |\Sigma| \leq \log |L|_n \leq \log |L|_{n_{i(n)}} + c \log |\Sigma| \\ \implies & \frac{\log |L|_{n_{i(n)+1}} - c \log |\Sigma|}{n} \leq \frac{\log |L|_n}{n} \leq \frac{\log |L|_{n_{i(n)}} + c \log |\Sigma|}{n} \\ \implies & \frac{\log |L|_{n_{i(n)+1}} - \frac{c \log |\Sigma|}{n}}{n} \leq \frac{\log |L|_n}{n} \leq \frac{\log |L|_{n_{i(n)}} + \frac{c \log |\Sigma|}{n}}{n} \\ \implies & \frac{\log |L|_{n_{i(n)+1}} - \frac{c \log |\Sigma|}{n}}{n_{i(n)+1}} \leq \frac{\log |L|_n}{n} \leq \frac{\log |L|_{n_{i(n)}} + \frac{c \log |\Sigma|}{n}}{n_{i(n)}}. \end{aligned}$$

Note that  $\left(\frac{c \log |\Sigma|}{n}\right)_n \rightarrow 0$ . Therefore,  $\left(\frac{\log |L|_n}{n}\right)_n$  is bounded between two sequences with the same limit; hence, it converges.

As an additional consequence, note that if  $L$  is a regular language closed under prefixes, then  $\lim_{n \rightarrow \infty} \frac{\log |L|_{\leq n}}{n}$  exists as well. -1

**3.2. Entropy and box-counting dimension.** Note that box-counting dimension appears to be quite hard to compute by the definition we have given, because there are fairly general quantifications over multiple variables. But with a bit of work, the process of proving the box-counting dimension of a set can be simplified. In particular, we have the following result showing that we do not need to check many values of  $\varepsilon$  and  $h$ :

LEMMA 3.4. *Let  $X \subseteq [0, 1]^d$ . Let  $r > 1$ , and assume that the limit*

$$L = \lim_{n \rightarrow \infty} \frac{\log N(X, r^{-n}, \vec{0})}{\log r^n}$$

*exists. Then  $d_B(X) = L$ .*

PROOF. Let  $(\varepsilon_i)_i$  be any sequence of positive values converging to zero, and let  $(\vec{h}_i)_i$  be any sequence of values in  $[0, 1]^d$ . We will show that

$$\lim_{i \rightarrow \infty} \frac{\log N(X, \varepsilon_i, \vec{h}_i)}{\log \frac{1}{\varepsilon_i}} = L.$$

Choose some arbitrary  $i$ , and find  $n_i$  such that  $r^{-n_i} > \varepsilon_i > r^{-n_i-1}$ . Because  $\varepsilon_i \rightarrow 0$ , we have  $n_i \rightarrow \infty$ .

Consider the boxes of the form  $I_{\vec{z}} = [z_1 r^{-n_i}, (z_1 + 1)r^{-n_i}] \times \dots \times [z_d r^{-n_i}, (z_d + 1)r^{-n_i}]$  where  $\vec{z} = (z_1, \dots, z_d)$  are integers. Similarly, choose an arbitrary vector of integers  $\vec{z}' = (z'_1, \dots, z'_d)$  and let  $I' = [(z'_1 - h_{i,1})\varepsilon_i, (z'_1 - h_{i,1} + 1)\varepsilon_i] \times \dots \times [(z'_d - h_{i,d})\varepsilon_i, (z'_d - h_{i,d} + 1)\varepsilon_i]$ . Note that because the  $I_{\vec{z}}$  are adjacent (i.e., they partition  $\mathbb{R}^d$ ) and longer than  $I'$ , then there exist at most  $2^d$  such  $I_{\vec{z}}$  that cover  $I'$ . We chose  $I'$  arbitrarily, so  $N(S, r^{-n_i}, 0) \leq 2^d N(S, \varepsilon_i, \vec{h}_i)$ .

Similarly, choose an arbitrary  $\vec{z}$  and let  $I = [z_1 r^{-n_i-1}, (z_1 + 1)r^{-n_i-1}] \times \dots \times [z_d r^{-n_i-1}, (z_d + 1)r^{-n_i-1}]$ , and consider the intervals of the form  $I_{\vec{z}'} = [(z'_1 - h_{i,1})\varepsilon_i, (z'_1 - h_{i,1} + 1)\varepsilon_i] \times \dots \times [(z'_d - h_{i,d})\varepsilon_i, (z'_d - h_{i,d} + 1)\varepsilon_i]$  for integer vectors  $\vec{z}'$ . Note that because the  $I_{\vec{z}'}$  are adjacent (i.e., they partition  $\mathbb{R}$ ) and longer than  $I$ , then there exist at most  $2^d$  such  $I_{\vec{z}'}$  that cover  $I$ . We chose  $I$  arbitrarily, so  $N(X, \varepsilon_i, \vec{h}_i) \leq 2^d N(X, r^{-n_i-1}, 0)$ .

These inequalities give us

$$\frac{\log \frac{1}{2^d} N(X, r^{-n_i}, 0)}{\log r^{n_i+1}} \leq \frac{\log N(X, \varepsilon_i, \vec{h}_i)}{\log \frac{1}{\varepsilon_i}} \leq \frac{\log 2^d N(X, r^{-n_i-1}, 0)}{\log r^{n_i}}$$

Apply laws of logarithms:

$$\frac{-\log 2^d + \log N(X, r^{-n_i}, 0)}{\log r + \log r^{n_i}} \leq \frac{\log N(X, \varepsilon_i, \vec{h}_i)}{\log \frac{1}{\varepsilon_i}} \leq \frac{\log 2^d + \log N(X, r^{-n_i-1}, 0)}{-\log r + \log r^{n_i+1}}$$

Now note that as long as  $a_i, b_i \rightarrow \infty$  with  $C_1, C_2$  constant, we have  $\lim_{i \rightarrow \infty} \frac{C_1 + a_i}{C_2 + b_i} = \lim_{i \rightarrow \infty} \frac{a_i}{b_i}$ . So,

$$\lim_{i \rightarrow \infty} \frac{-\log 2^d + \log N(X, r^{-n_i}, 0)}{\log r + \log r^{n_i}} = \lim_{i \rightarrow \infty} \frac{\log N(X, r^{-n_i}, 0)}{\log r^{n_i}} = L.$$

Similarly,

$$\lim_{i \rightarrow \infty} \frac{\log 2^d + \log N(X, r^{-n_i-1}, 0)}{-\log r + \log r^{n_i+1}} = \lim_{i \rightarrow \infty} \frac{\log N(X, r^{-n_i-1}, 0)}{\log r^{n_i+1}} = L.$$

Since  $\frac{\log N(X, \varepsilon_i, \vec{h}_i)}{\log \frac{1}{\varepsilon_i}}$  is bounded between two sequences converging to  $L$ , we conclude that its limit is  $L$  as well. ⊖

**LEMMA 3.5.** *Let  $L$  be an  $\omega$ -language of base- $k$  representations of points in  $[0, 1]^d$ . Let  $X = v_k(L)$ . Then  $d_B(X) = \frac{1}{\log k} h(L^{pre})$ .*

**PROOF.** Lemma 3.3 tells us that the entropy of  $L^{pre}$  is defined as a limit and not just as a limit superior. By Lemma 3.4, it then suffices to show

$$\lim_{n \rightarrow \infty} \frac{\log N(X, k^{-n}, \vec{0})}{\log k^n} = \frac{1}{\log k} \lim_{n \rightarrow \infty} \frac{\log |L^{pre}|_n}{n}.$$

Now consider each string  $w$  in  $L^{pre}$  of length  $n$ . The infinite strings with this prefix represent precisely the numbers in  $I_{\vec{z}} = [z_1 k^{-n}, (z_1 + 1)k^{-n}] \times \dots \times [z_d k^{-n}, (z_d + 1)k^{-n}]$  where  $\vec{z} = (z_1, \dots, z_d)$  is the integer vector with base- $k$  representation given by  $w$ . So all of these boxes for all such strings  $w$  will cover  $X$ ; thus,  $N(X, k^{-n}, \vec{0}) \leq |L^{pre}|_n$ .

This covering may not be optimal, because the boxes  $I_{\vec{z}}$  are not disjoint; they overlap at points with at least one coordinate that is  $k$ -rational. So a single  $I_{\vec{z}}$  may contain points that the above covering method would place in any adjacent box. However, this is the only overlap, so a box in the optimal covering corresponds to at most  $3^d$  boxes in the above covering. Hence  $N(X, k^{-n}, \vec{0}) \geq \frac{1}{3^d} |L^{pre}|_n$ . Therefore,

$$\begin{aligned} \frac{1}{\log k} \lim_{n \rightarrow \infty} \frac{\log |L^{pre}|_n}{n} &= \frac{1}{\log k} \lim_{n \rightarrow \infty} \frac{\log \frac{1}{3^d} + \log |L^{pre}|_n}{n} \\ &= \frac{1}{\log k} \lim_{n \rightarrow \infty} \frac{\log \left( \frac{1}{3^d} |L^{pre}|_n \right)}{n} \\ &\leq \frac{1}{\log k} \lim_{n \rightarrow \infty} \frac{\log N(X, k^{-n}, \vec{0})}{n} \\ &\leq \frac{1}{\log k} \lim_{n \rightarrow \infty} \frac{\log |L^{pre}|_n}{n}. \end{aligned}$$

It follows that the two limits are equal, as required. ⊖

We conclude this section by noting that the choice of regular language to associate with a  $k$ -automatic set is not always obvious. In particular, in their conjecture regarding the connection between entropy and dimension, Adamczewski and Bell [1]

use the regular language of *substrings* of the base- $k$  expansions, not the language of prefixes as we are using here. We will show that this does not make a difference.

This result follows from [20], in which the author shows that the entropy of a  $k$ -regular  $\omega$ -language  $L$  is equal to that of the (normal) regular language  $L^{pre}$ . Theorem 14 of [20] further shows that if  $L$  is an  $\omega$ -language, then there is a strongly connected  $\omega$ -language  $L'$  and a word  $w \in \Sigma$  such that  $\{v : v = w\ell, \ell \in L'\} \subseteq L$ , and  $h(L) = h(L')$ . It follows that the entropy of  $L$  is equal to the maximum over the entropies of the subsets corresponding to the strongly connected components of any Büchi automaton  $\mathcal{A}$  recognizing  $L$ , which together implies the entropy of  $L$  is the same as that of the set of substrings of  $L$ .

**FACT 3.6 [20].** *Let  $L$  be a regular language closed under prefixes, and let  $L^{sub}$  be the regular language of substrings of strings in  $L$ . Then  $h(L) = h(L^{sub})$ .*

**§4. The closed case.**

**4.1. Spectral radius and box-counting dimension.** We know a lot about the dimension of  $V_k(\mathcal{A})$  when it is a closed set, and we know that if  $\mathcal{A}$  is deterministic (and trim), then we can take the closure of  $V_k(\mathcal{A})$  by making all states accepting. It is not immediately obvious that this will hold when  $\mathcal{A}$  is nondeterministic, but in the following lemma we show this is case. The following result is essentially a corollary of Lemma 58 in [6]:

**LEMMA 4.1.** *Let  $\mathcal{A}$  be a trim Büchi automaton. Then  $V_k(\overline{\mathcal{A}}) = \overline{V_k(\mathcal{A})}$ .*

**PROOF.** First, we prove that  $V_k(\overline{\mathcal{A}})$  is closed. Let  $x \in \overline{V_k(\overline{\mathcal{A}})}$ . Then there exist  $(x_m)_{m \in \mathbb{N}} \in V_k(\overline{\mathcal{A}})$  with  $x_m \rightarrow x$ . Without loss of generality assume that either  $x_m < x$  for all  $m$ , or else  $x_m > x$  for all  $m$ . We will assume the latter; the proof in the former case is analogous.

Let  $w$  be the infinite base- $k$  representation of  $x$  such that if  $x$  is  $k$ -rational, then  $w$  ends in  $0^\omega$ ; this uniquely identifies  $w$ . (We choose this representation because we assumed  $x_m > x$ .) Then for all  $n$  there exists  $m_n$  such that the first  $n$  characters of some base- $k$  representation of  $x_{m_n}$  are the first  $n$  characters of  $w$ . Let  $w_n$  be this base- $k$  representation for a given  $n$ .

Now, we will define a graph  $G$  inductively. The vertex set will be a subset of  $\Sigma^* \times Q^*$ . Our initial condition is that  $G$  contains the vertices  $(\varepsilon, q)$ , where  $q$  ranges over  $S$  (the set of start states of  $\overline{\mathcal{A}}$ ). Then for every vertex  $(u, v)$ , we require that  $G$  contain the vertex  $(uc, vq)$  where  $uc$  is a prefix of  $w$  and  $\overline{\mathcal{A}}$  transitions from the last state in  $v$  to state  $q$  on  $c$ ; and we require that  $G$  contain the edge from  $(u, v)$  to  $(uc, vq)$ .

Observe that:

- $G$  has finitely many connected components: as we have defined  $G$ , every vertex is connected to a vertex whose string is shorter. So if the length of the string for a vertex is  $n$ , that vertex has a path of length  $n$  to an initial vertex.
- $G$  is locally finite: every vertex has at most  $k|Q|$  incident edges.
- $G$  is infinite: if  $w_n$  is accepted, there must be an accepting path for the first  $n$  characters of  $w_n$ , which are also the first  $n$  characters of  $w$ . This gives a vertex for every natural number  $n$ .

Therefore, we can apply König’s lemma to one of the connected components of  $G$  and conclude that  $G$  has an infinite path. Moreover, note that because strings have finite length, infinitely many of the edges in the path must lead to a longer string; and because there is only one edge from any vertex to a shorter string, it in fact must be the case that eventually the path only contains edges to longer strings. Choose a vertex sufficiently far into the path that this is the case; there is a path from an initial vertex to this vertex and then infinitely farther through longer and longer strings; and this path must correspond to an accepting path for  $w$ . So  $\mathcal{A}$  accepts  $w$ , and thus,  $V_k(\overline{\mathcal{A}})$  is closed.

Now it remains to show that  $V_k(\overline{\mathcal{A}})$  is the closure of  $V_k(\mathcal{A})$ . Note that any accepting run in  $\mathcal{A}$  is also accepting in  $\overline{\mathcal{A}}$ , so  $V_k(\mathcal{A}) \subseteq V_k(\overline{\mathcal{A}})$ . Moreover, let  $w$  be accepted by  $\overline{\mathcal{A}}$ . Then for any  $n$ , the first  $n$  characters of  $w$  have a run to some state in  $\overline{\mathcal{A}}$ . As the only difference between  $\mathcal{A}$  and  $\overline{\mathcal{A}}$  is the set of accept states, there is also a run for these characters in  $\mathcal{A}$ , and because  $\mathcal{A}$  is trim, this run can be extended to an accepting run. So every infinite string accepted by  $\overline{\mathcal{A}}$  has, for all  $n$ , its first  $n$  characters in common with some infinite string accepted by  $\mathcal{A}$ . It follows that  $V_k(\overline{\mathcal{A}}) \subseteq \overline{V_k(\mathcal{A})}$ . But the only closed subset of  $\overline{V_k(\mathcal{A})}$  that is a superset of  $V_k(\mathcal{A})$  is  $\overline{V_k(\mathcal{A})}$  itself. –

**4.2. Spectral radius and Hausdorff dimension.** In [14] Mauldin and Williams work with what they call a “geometric graph directed construction” (or GGDC) on  $\mathbb{R}^m$ , and in [9] Evans works with the same notion. These constructions would be described in the current terminology of metric geometry as GDIFSs that satisfy the open set condition and have compact attractors. The objects of interest in [9] are also called “ $k$ -self-similar” sets, and these include the closed  $k$ -automatic subsets of  $[0, 1]^d$ , as we will verify shortly. When a  $k$ -automatic set is not closed, it is not  $k$ -self-similar in the sense of [9] by definition, and it should be noted that in general  $k$ -self-similar sets need not be  $k$ -automatic.

Recall the open set condition for GDIFSs as defined in [8].

**DEFINITION 4.2.** If  $\mathcal{G} = (V, E, s, t, X, S)$  is a GDIFS, then it satisfies the *open set condition* precisely if for all  $v \in V$  there are nonempty open sets  $U_v \subseteq \mathbb{R}^d$  such that

$$U_u \supseteq f_e[U_v]$$

for all  $u, v \in V$  such that  $e \in E_{u,v}$ , and additionally,

$$f_e[U_v] \cap f_{e'}[U_{v'}] = \emptyset$$

for all  $u, v, v' \in V$  such that  $e \in E_{u,v}$  and  $e' \in E_{u,v'}$ .

In [9], Evans shows that Theorem 1.1 holds for the sets he calls  $k$ -self-similar by first showing that all  $k$ -self-similar sets are GGDCs. He then proves the conclusion of Theorem 1.1 for GGDCs. Since it is already demonstrated in [6] that we can associate with any compact  $k$ -automatic set a GDIFS, and because knowing that every compact  $k$ -automatic is a GGDC will give us additional tools and information, we will show that any closed  $k$ -automatic set  $X \subseteq [0, 1]^d$  is a GGDC, and the consequence of Theorem 1.1 of [9] will follow.

**THEOREM 4.3.** *Suppose that  $X \subseteq [0, 1]^d$  is a  $k$ -automatic set recognized by a deterministic Büchi automaton  $\mathcal{A}$  with corresponding GDIFS  $\mathcal{G}_{\mathcal{A}} = (V, E, s, t, X, S)$ . Then  $\mathcal{G}_{\mathcal{A}}$  satisfies the open set condition as defined in [8].*

**PROOF.** Let  $X, \mathcal{A}$ , and  $\mathcal{G}_{\mathcal{A}}$  be as in the hypotheses. To see that the open set condition holds, for each  $v \in V$  let  $U_v = (0, 1)^d$ . Observe that for any  $e \in E$ , it is the case that  $f_e[U_v] = \{\frac{x+\sigma_e}{r} : x \in U_v\}$ , where  $\sigma_e \in \Sigma$  is the label of the transition arrow in  $\mathcal{A}$  that corresponds to the edge  $e$ . Hence it is clear that  $U_u \supseteq f_e(U_v)$  for any vertices  $u, v \in V$  such that  $e$  is an edge from  $u$  to  $v$ , since

$$(0, 1)^d \supseteq \prod_{i=1}^d \left( \frac{\sigma_{e,i}}{r}, \frac{\sigma_{e,i} + 1}{r} \right).$$

Suppose now that  $u, v, v' \in V$  and that  $e \in E_{u,v}$  and  $e' \in E_{u,v'}$ . Due to  $\mathcal{A}$  being deterministic, it cannot be the case that the transition arrows corresponding to  $e$  and  $e'$  respectively in  $\mathcal{A}$  have the same label. Thus we know  $\sigma_e \neq \sigma_{e'}$ , and as a consequence if an element  $\bar{a}$  is in  $\{\frac{x+\sigma_e}{r} : x \in (0, 1)^d\}$ , then it cannot be the case that  $\bar{a} \in \{\frac{x+\sigma_{e'}}{r} : x \in U_v\}$ , since such sets carve out the interior of disjoint subcubes of  $(0, 1)^d$  of size  $(\frac{1}{r})^d$ . Hence the open set condition is satisfied.  $\dashv$

The following corollary is immediate from the fact that a closed Büchi automaton is weak, by applying the above theorem. Recall that weak automata (which include all closed automata) have an equivalent deterministic Büchi automaton [18].

**COROLLARY 4.4.** *All closed Büchi automata satisfy the open set condition.*

The following lemma concerns strongly connected automata, rather than closed ones, and will be broadly useful in our methods.

**LEMMA 4.5.** *Let  $\mathcal{A}$  and  $\mathcal{A}'$  be two strongly connected Büchi automata with the same set of states, set of accept states, and transition relation. Then  $d_H(V_k(\mathcal{A})) = d_H(V_k(\mathcal{A}'))$ .*

**PROOF.** Note that  $V_k(\mathcal{A}) = \bigcup_{q_i \in S} V_k(\mathcal{A}_i)$ , where  $S$  is the set of start states of  $\mathcal{A}$  and  $\mathcal{A}_i$  is a modification of  $\mathcal{A}$  where  $q_i$  is the only start state. Therefore  $d_H(V_k(\mathcal{A})) = \max_{q_i \in S} d_H(V_k(\mathcal{A}_i))$ . Let  $r$  be the value of  $i$  giving maximal dimension, i.e.,  $d_H(V_k(\mathcal{A})) = d_H(V_k(\mathcal{A}_r))$ . Define  $\mathcal{A}'_s$  and choose  $s$  analogously, such that  $d_H(V_k(\mathcal{A}')) = d_H(V_k(\mathcal{A}'_s))$  and such that  $q_s$  is the only start state in  $\mathcal{A}'_s$ .

Because  $\mathcal{A}_r$  is strongly connected, it must contain a path  $P$  from  $q_s$  to  $q_r$ ; let  $v \in ([k]^d)^*$  be a string witnessing path  $P$ . Let  $w \in ([k]^d)^\omega$  be accepted by  $\mathcal{A}_r$ . Then  $P$  concatenated with an accepting run of  $w$  in  $\mathcal{A}_r$  forms an accepting run of  $vw$  in  $\mathcal{A}'_s$ . Moreover, note that

$$\begin{aligned} v_k(vw) &= \sum_{i=0}^{\infty} \frac{(vw)_i}{k^{i+1}} \\ &= \sum_{i=0}^{|v|-1} \frac{(vw)_i}{k^{i+1}} + \sum_{i=|v|}^{\infty} \frac{(vw)_i}{k^{i+1}} \end{aligned}$$

$$\begin{aligned}
 &= \sum_{i=0}^{|v|-1} \frac{v_i}{k^{i+1}} + \sum_{i=0}^{\infty} \frac{w_i}{k^{|v|+i+1}} \\
 &= v_k(v\vec{0}^\omega) + k^{-|v|}v_k(w).
 \end{aligned}$$

Let  $f : [0, 1]^d \rightarrow [0, 1]^d$  map  $x$  to  $v_k(v\vec{0}^\omega) + k^{-|v|x}$ ; then the above chain of equalities gives us that  $f(V_k(\mathcal{A}_r)) \subseteq V_k(\mathcal{A}'_s)$ . Because  $f$  is an invertible affine transformation, it follows that  $d_H(V_k(\mathcal{A}_r)) \leq d_H(V_k(\mathcal{A}'_s))$ . The analogous argument in the opposite direction gives us  $d_H(V_k(\mathcal{A}'_s)) \leq d_H(V_k(\mathcal{A}_r))$ . Thus,

$$d_H(V_k(\mathcal{A})) = d_H(V_k(\mathcal{A}_r)) = d_H(V_k(\mathcal{A}'_s)) = d_H(V_k(\mathcal{A}')). \quad \dashv$$

In order to establish the final results of this section, we will need to talk about the adjacency matrix of a Büchi automaton.

**DEFINITION 4.6.** Let  $\mathcal{A}$  be a Büchi automaton. Suppose that  $Q$ , the set of states of  $\mathcal{A}$ , is size  $n$  and let  $\iota : \{1, \dots, n\} \rightarrow Q$  be a fixed bijection. Then we associate with  $\mathcal{A}$  a weighted adjacency matrix  $\mathbf{M}(\mathcal{A}, s) = (m_{i,j})$  given by

$$m_{i,j} = \left( \frac{|\{\sigma \in \Sigma : (\iota(i), \iota(j), \sigma) \in \Delta\}|}{k} \right)^s.$$

Recall that  $\Sigma$  is the alphabet and  $\Delta$  is the transition function of  $\mathcal{A}$ .

We will need the following fact regarding sequences defined by repeated matrix multiplication. The following result appears in [20]:

**FACT 4.7.** *Let  $L$  be a regular language closed under prefixes. Let  $\mathcal{A}$  be a finite automaton recognizing  $L$ , and assume that  $\mathcal{A}$  is deterministic, trim, and has every state accepting. Label the states of  $\mathcal{A}$  by numbers 1 through  $m$ . Then the entropy of  $L$  is equal to the spectral radius  $\text{sprad } k\mathbf{M}(\mathcal{A}, 1)$  of weighted adjacency matrix  $\mathbf{M}(\mathcal{A}, 1)$  scaled by the constant  $k$ .*

We observe that the GDIFS associated with a closed Büchi automaton is nearly of the form that gives rise to a GGDC in the sense of [9, 14]. However in [14], the authors consider only iterated function systems directed by a graph, whereas the GDIFSs associated with closed automata are in general multigraphs. To account for this, though, we observe that we can “translate” between multigraph representations of Büchi automata and automaton diagrams which are true digraphs.

**DEFINITION 4.8.** For a deterministic Büchi automaton  $\mathcal{A} = (Q, \Sigma, \delta, S, F)$ , consider the automaton  $(Q \times \Sigma, \Sigma, \delta', S', F \times \Sigma)$ , with  $\delta'$  and  $S'$  defined as follows:

- For all  $\sigma, \tau \in \Sigma$  and  $q, s \in Q$ , if  $\delta(q, \sigma) = \emptyset$ , then  $\delta'((q, \tau), \sigma) = \emptyset$ . If  $\delta(q, \sigma) = \{s\}$ , then  $\delta'((q, \tau), \sigma) = \{(s, \sigma)\}$ . (These are the only possibilities, because  $\mathcal{A}$  is deterministic.)
- Fix some  $\sigma_0 \in \Sigma$  arbitrarily; then with  $S = \{q\}$ , we have  $S' = \{(q, \sigma_0)\}$ .

Since this automaton is not necessarily trim, let  $\mathcal{A}_{dg}$  denote the automaton that results from removing any states that are not accessible from the start state or not co-accessible from some accept state.

**LEMMA 4.9.** *If the automaton  $\mathcal{A}$  is a closed Büchi automaton, then  $\mathcal{A}_{dg}$  is an equivalent Büchi automaton that is also closed.*



PROOF. To see that  $\mathcal{A}$  and  $\mathcal{A}_{dg}$  are equivalent as automata, observe that if  $(q_i)_{i \in \mathbb{N}}$  is an accepting run of  $(\tau_i)_{i \in \mathbb{N}}$  in  $\mathcal{A}$ , then set  $\tau_{-1} = \sigma_0$  and observe that the run  $(q_i, \tau_{i-1})_{i \in \mathbb{N}}$  of  $(\tau_i)_{i \in \mathbb{N}}$  in  $\mathcal{A}_{dg}$  is an acceptance run for  $\mathcal{A}_{dg}$ . Similarly if  $(q_i, \sigma_i)_{i \in \mathbb{N}}$  is an acceptance run of  $(\tau_i)_{i \in \mathbb{N}}$  in  $\mathcal{A}_{dg}$ , then  $(q_i)_{i \in \mathbb{N}}$  is an acceptance run in  $\mathcal{A}$  for the same string.

Since we define  $\mathcal{A}_{dg}$  to be trim and since  $Q = F$  for  $\mathcal{A}$ , we observe  $Q \times \Sigma = F \times \Sigma$  makes  $\mathcal{A}_{dg}$  closed precisely if  $\mathcal{A}$  is closed. ⊖

Now that we have established that closed  $k$ -automatic subsets of  $[0, 1]^d$  are GGDCs in the sense of [14], the following is immediate from Theorem 1.1 in [9]:

COROLLARY 4.10. *If  $X$  is a closed  $k$ -automatic set in  $\mathbb{R}$  recognized by closed automaton  $\mathcal{A}$ , then  $d_H(X) = d_B(X) = \frac{1}{\log(k)}h(L(\mathcal{A}))$ , where  $L(\mathcal{A})$  is the set of strings  $\mathcal{A}$  recognizes.*

The next corollary follows from the remarks made in the proof of Lemma 4.9 and essentially says that connected components are preserved by the construction in Definition 4.8.

COROLLARY 4.11. *If  $C$  is a strongly connected component of  $\mathcal{A}$ , then there is a corresponding strongly connected component of  $\mathcal{A}_{dg}$ , call it  $C_{dg}$ , such that the induced sub-automata generated by  $C$  and  $C_{dg}$  (respectively) recognize the same strings up to prefixes. In other words, let  $L$  be recognized by the induced sub-automaton generated by  $C$ , and likewise for  $L_{dg}$  and  $C_{dg}$ ; then there exist words  $u, v$  such that  $L = uL_{dg}$  and  $L_{dg} = vL$ .*

PROOF. First, we note that the choice of start state in each induced sub-automaton does not matter. In a strongly connected automaton, if the start state is moved, the resulting language and original language are equal up to prefixes (as defined in the statement of the corollary). So we will allow the start states to be chosen freely.

Let  $\mathcal{A}_C$  denote the sub-automaton of  $\mathcal{A}$  consisting of the states and transitions contained within  $C$ , with arbitrarily chosen  $q \in C$  as its only start state. Then by strong connectedness, there exists a word  $w$  and a corresponding run  $(q_i)_{0 \leq i \leq |w|}$  where  $q_0 = q_n = q$  and where, for each transition within  $C$ , there exists an  $i$  witnessing the transition, i.e., if there is an arrow from state  $r$  to state  $s$  on the character  $\sigma$ , then there exists  $i$  such that  $q_{i-1} = r, q_i = s, w_i = \sigma$ .

Let  $C_{dg}$  be the strongly connected component of  $(q, w_n)$  in  $\mathcal{A}_{dg}$ , and let  $\mathcal{A}_{C_{dg}}$  be its induced sub-automaton with  $(q, w_n)$  as the start state. Then there is a corresponding path  $((q_i, w_i))_{0 \leq i \leq |w|}$  (letting  $w_0 = w_n$ ) in  $\mathcal{A}_{C_{dg}}$ ; this path is a cycle, so each state  $(q_i, w_i)$  is in  $C_{dg}$ .

Let  $u$  be any infinite word accepted by  $\mathcal{A}_C$ . Following the proof of the previous lemma, note that because each transition in the acceptance run for  $u$  is also somewhere in the run for  $w$ , and therefore it has a corresponding transition in  $\mathcal{A}_{C_{dg}}$ . So  $u$  is accepted by  $\mathcal{A}_{C_{dg}}$ . The reverse inclusion follows immediately from the same argument as in Lemma 4.9. ⊖

The following is implied by the results in [21] in which the author shows that for closed languages the Hausdorff dimension and entropy agree, in conjunction with the countable additivity of entropy, as shown in [20].

LEMMA 4.12. *Let  $\mathcal{A}$  be a closed Büchi automaton. Let  $SC(\mathcal{A})$  denote the set of strongly connected components of  $\mathcal{A}$ . For each strongly connected component  $C$  of  $\mathcal{A}$ , let  $\mathcal{A}_C$  denote the sub-automaton of  $\mathcal{A}$  consisting of the states and transitions contained within  $C$ , with arbitrarily chosen  $q \in C$  as its only start state.*

*Then*

$$d_H(V_k(\mathcal{A})) = \max_{C \in SC(\mathcal{A})} (d_H(V_k(\mathcal{A}_C))).$$

PROOF. By Lemma 4.5, the Hausdorff dimension of the automaton  $\mathcal{A}_C$  does not depend on the choice of  $q$ , its start state. Let  $K$  be the attractor of  $G_{\mathcal{A}}$ , i.e., the set that  $\mathcal{A}$  recognizes. We recall the statement of Theorem 5 of [14]. It states that if  $K = \bigcup_{v \in V} K_v$  is the attractor of GDIFS  $G$ , then  $d_H(K_v) = \alpha_v = \max\{\alpha_H \mid H \in SC_v(G)\}$ . Here,  $SC_v(G)$  denotes the set of strongly connected components of  $G$  that are accessible from  $v$ .

By Corollary 4.4 we know that  $\mathcal{A}_{dg}$  satisfies the definition in [14] of a geometric graph directed construction. By Lemma 4.9 we know that the automaton  $\mathcal{A}_{dg}$  recognizes the same set as  $\mathcal{A}$ ; hence,  $d_H(V_k(\mathcal{A})) = d_H(\mathcal{A}_{dg})$ . For each strongly connected component  $C$  of  $\mathcal{A}$ , let  $\mathcal{A}_C$  be the automaton whose automaton diagram is given by taking  $C$  and assigning an arbitrary start state from those in the vertex set of  $C$ . By Corollary 4.11, for each  $C$  a strongly connected component of  $G$  there is a corresponding strongly connected component  $C_{dg}$  of  $\mathcal{A}_{dg}$  such that  $d_H(V_k(\mathcal{A}_C)) = d_H(V_k(\mathcal{A}_{C,dg}))$ , where  $\mathcal{A}_{C,dg}$  is an automaton whose automaton diagram is given by taking  $C_{dg}$  and assigning an arbitrary start state from those in the vertex set of  $C_{dg}$ .

Since  $\mathcal{A}_{dg}$  has a true digraph structure (rather than that of a multigraph), we can apply Theorem 3 of [14] to obtain that the Hausdorff dimension of  $V_k(\mathcal{A}_{C,dg})$  is the unique  $\alpha$  such that  $\text{spradM}(\mathcal{A}_C, \alpha) = 1$ . Note that the value of  $\max_{C \in SC(\mathcal{A})} d_H(V_k(\mathcal{A}_C))$  is well-defined by Lemma 4.5. By Theorem 4 of [14], we conclude that the dimension of  $V_k(\mathcal{A}_{dg})$  is the maximum over all such  $\alpha$  for  $\mathcal{A}_{C,dg} \in SC(\mathcal{A}_{dg})$ . ◻

**§5. Cycle languages and when dimensions disagree.** We would like to know in the general case when the Hausdorff dimension of a  $k$ -automatic set is not equal to its box-counting dimension. A fundamental example of this is the set of dyadic rationals in  $[0, 1]$ :  $\{\frac{a}{2^n} : a, n \in \mathbb{N}\} \cap [0, 1]$ . This set is 2-automatic because it can be equivalently phrased as the set of numbers in  $[0, 1]$  whose binary expansions only have a finite number of one of the two bits (0 or 1). It has Hausdorff dimension 0, as it is countable, and box-counting dimension 1, as it is dense in the interval. Examining a Büchi automaton for this set, we note that it seems the reason for this disparity in dimension is that there are many ways to get from the start state to an accept state but few ways (one way, in fact) to loop from an accept state back to itself. We can in fact formalize this notion and obtain a sufficient and necessary condition for the equivalence of the two notions of dimension by defining the notion of a *cycle language*.

DEFINITION 5.1. In a finite or Büchi automaton  $\mathcal{A}$  with  $q$  as one of its states, the *cycle language*  $C_q(\mathcal{A}) \subseteq \Sigma^*$  contains all strings  $w$  for which there is a run of  $\mathcal{A}$  from

state  $q$  to itself on the string  $w$ . Let  $C_q(\mathcal{A})$  denote the automaton constructed by taking  $\mathcal{A}$ , making state  $q$  the only start and accept state, and trimming the resulting automaton. Call this the *cycle automaton*.

Note that cycle languages are regular, and this is witnessed by the cycle automaton. Note also that because looping to an accept state multiple times (or zero times) is still a loop,  $C_q(\mathcal{A})^* = C_q(\mathcal{A})$ .

LEMMA 5.2. *Let  $\mathcal{A} = (Q, \Sigma, \delta, S, F)$  be a trim Büchi automaton, and let  $X = V_k(\mathcal{A})$ . Let  $X_q = V_k(C_q(\mathcal{A}))$  for each  $q \in Q$ . Then:*

- (i)  $d_H(X) = \max_{q \in F} d_H(X_q)$ .
- (ii)  $d_B(X) = \max_{q \in Q} d_H(X_q)$ .

PROOF. (i) Let  $L_q$  be the language of words that have a path from a start state of  $\mathcal{A}$  to state  $q$ . An infinite string is accepted by  $\mathcal{A}$  precisely when it has a path that runs from a start state to an accept state and then cycles back to that accept state infinitely often. Thus,

$$L(\mathcal{A}) = \bigcup_{q \in F} \bigcup_{w \in L_q} wC_i(\mathcal{A})^\omega.$$

Let  $T_w$  be the (linear) transformation on a set corresponding to prefixing the string  $w$ . Applying  $d_H \circ V_k$  to both sides, we get

$$d_H(X) = d_H \left( \bigcup_{q \in F} \bigcup_{w \in L_q} T_w(X_q) \right).$$

Applying the formula for Hausdorff dimension of a countable union,

$$d_H(X) = \sup_{q \in F} \sup_{w \in L_q} d_H(T_w(X_q)).$$

Because (invertible) linear transformations do not affect dimension,

$$d_H(X) = \sup_{q \in F} \sup_{w \in L_q} d_H(X_q).$$

Eliminating the now-useless quantification over  $w$ ,

$$d_H(X) = \sup_{q \in F} d_H(X_q).$$

Because  $\mathcal{A}$  is finite,

$$d_H(X) = \max_{q \in F} d_H(X_q).$$

- (ii) Because  $\mathcal{A}$  is trim, we have  $\overline{X} = V_k(\overline{\mathcal{A}})$  according to Lemma 4.1. So applying (i), we get that  $d_H(\overline{X}) = \max_{q \in Q} d_H(X_q)$  (note that the cycle language  $C_q(\mathcal{A})$  does not depend on which states are accepting). Yet we know from Theorem 4.10 that  $d_H(\overline{X}) = d_B(\overline{X}) = d_B(X)$ . □

We do immediately get as a corollary that  $d_H(X) < d_B(X)$  when  $d_H(X_q)$  is larger for some  $q \notin F$  than for all  $q \in F$ . However, this is not a very useful version of the characterization as it stands, because the Hausdorff dimension of  $v_k(C_q(\mathcal{A})^\omega)$  is not an easy value to compute *a priori*. The rest of this section will be focused on reducing the above to an easier problem.

The main step in the process of simplifying the result of Lemma 5.2 is the following result:

**LEMMA 5.3.** *Let  $\mathcal{A}$  be a finite automaton, not necessarily deterministic. Let  $L'$  be the  $\omega$ -language recognized by  $\mathcal{A}$  as a Büchi automaton, and let  $L$  be the language recognized by  $\mathcal{A}$  as a finite automaton. Let  $X = v_k(L')$  and  $Y = v_k(L^\omega)$ . Then  $d_H(\bar{X}) \leq d_H(Y)$ .*

**PROOF.** Without loss of generality, assume  $\mathcal{A}$  is finite-trim. Let  $\mathcal{B}$  be the induced sub-automaton containing states in  $\mathcal{A}$  from which there are arbitrarily long paths to accept states. We note that  $\mathcal{B}$  is trim and is equivalent as a Büchi automaton to  $\mathcal{A}$ . Assume that the desired lemma holds for  $\mathcal{B}$ ; let  $L_{\mathcal{B}}$  be the language recognized by  $\mathcal{B}$  as a finite automaton, and let  $Z = v_k(L_{\mathcal{B}}^\omega)$ . Then  $L_{\mathcal{B}} \subseteq L$ ; hence,  $d_H(Z) \leq d_H(Y)$ . So  $d_H(\bar{X}) \leq d_H(Y)$  as well; because  $\mathcal{A}$  and  $\mathcal{B}$  are equivalent as Büchi automata, the lemma then holds for  $\mathcal{A}$ . Thus we have reduced the lemma to the case where the automaton in question is trim.

Starting over, assume  $\mathcal{A}$  is trim. Then  $\bar{\mathcal{A}}$ , as a finite automaton, recognizes the language  $M$  of prefixes of  $L$ . Let  $M'$  be the language  $\bar{\mathcal{A}}$  recognizes as a Büchi automaton.

By the proof of Lemma 4.1, the language  $M'$  is closed, in the sense that if infinitely many of an infinite string  $w$ 's prefixes are prefixes of strings in  $M'$ , then  $w \in M'$ . Therefore  $M' = \bar{M}$ , as  $M$  is the language of prefixes of strings in  $M'$ . Note furthermore that  $L' \subseteq M'$ . Moreover, any prefix of a string in  $M'$  is a string in  $M$  and thus a prefix of a string in  $L$ , and thus, from trimness, a prefix of a string in  $L'$ ; so  $v_k(\bar{M}) = \bar{X}$ .

Let  $\mathcal{Q}$  be the state set of  $\mathcal{A}$  (and  $\bar{\mathcal{A}}$ ). Choose for each  $q \in \mathcal{Q}$  a string  $u_q$  on which  $\mathcal{A}$  runs from a start state to state  $q$ , and choose a string  $v_q$  on which  $\mathcal{A}$  runs from state  $q$  to an accept state. Note that  $u_q v_q \in L$ . Let  $\ell$  be the maximum length of  $v_q u_q$  for  $q \in \mathcal{Q}$ .

Observe that because  $\bar{X} = v_k(\bar{M})$  is closed, and  $M$  is closed under prefixes, we have that  $d_H(\bar{X}) = \frac{1}{\log k} h(M)$  by Theorem 4.10. Define the language  $M_n$  for each positive integer  $n$  as follows: for each string in  $M$ , and for every accepting run of this string in  $\mathcal{A}$ , after every  $n$  characters, we find the index  $q$  of the state  $\mathcal{A}$  is in and insert  $v_q u_q$ ; the resulting strings are the elements of  $M_n$ . We note that  $M_n$  and  $X_n = v_k(\bar{M}_n)$  have the following properties:

- $\bar{M}_n$  is a subset of  $L^\omega$ . A string in  $\bar{M}_n$  must have the form  $w_1 v_{q_1} u_{q_1} w_2 v_{q_2} u_{q_2} w_3 \dots$  with each  $w_j$  a string of length  $n$ . Note that  $w_1 v_{q_1} \in L$ , and  $u_{q_j} w_{j+1} v_{q_{j+1}}$  is in  $L$  as well. So this string is in  $L^\omega$ , and hence  $X_n \subseteq Y$ .
- The set  $X_n$  is still closed; observe that by considering every possible path through  $L$  of length  $n$  and connecting them with strings of states recognizing  $v_q u_q$ , it is possible to create a closed Büchi automaton for  $\bar{M}_n$  and apply Lemma 4.1 again. Therefore,  $d_H(X_n) = \frac{1}{\log k} h(M_n^{pre})$  by Theorem 4.10.

- Let  $r$  be a positive integer; every string of length at most  $rn$  in  $M$  has a corresponding string of length at most  $r(n + \ell)$  in  $M_n^{pre}$ . Thus  $|M_n^{pre}|_{\leq r(n+\ell)} \geq |M|_{\leq rn}$ , and so

$$\begin{aligned} h(M_n^{pre}) &= \limsup_{r \rightarrow \infty} \frac{\log |M_n^{pre}|_{\leq r(n+\ell)}}{r(n + \ell)} \\ &\geq \limsup_{r \rightarrow \infty} \frac{\log |M|_{\leq rn}}{r(n + \ell)} \\ &= \frac{n}{n + \ell} \limsup_{r \rightarrow \infty} \frac{\log |M|_{\leq rn}}{rn}. \end{aligned}$$

Now, without loss of generality, assume our alphabet does not contain the character  $\$,$  and let  $N = M\$\ast$ . Observe that  $N$  is still prefix-closed and that  $|M|_{\leq n} = |N|_n$  for all  $n$ . Then the above expression is equal to  $\frac{n}{n+\ell}h(N)$ , because from Lemma 3.3, we know that the limit superior defining  $h(N)$  is a limit (thus it does not matter if we take a subsequence of the indices).

We would then like to replace  $N$  with  $M$  in this equation; hence, we show  $h(N) = h(M)$ . Note that because  $\bar{A}$  is trim, every string in  $M$  can be extended to a longer string in  $M$ ; and because  $M$  is prefix-closed, said longer string can have any length. Therefore  $|M|_n$  is monotone in  $n$ , and so  $|N|_n = |M|_{\leq n} \leq (n + 1)|M|_n$ . Conversely we trivially have  $|M|_n \leq |N|_n$ . Since multiplication by a linear factor does not change the entropy, we have  $h(N) = h(M)$  as required.

- Let  $r$  be a positive integer. Each string in  $M$  corresponds to at most  $|Q|(\ell + 1)$  strings in  $M_n^{pre}$  because, at worst, the length of the original string is a multiple of  $n$ , and thus  $M_n^{pre}$  has  $|Q|(\ell + 1)$  corresponding strings depending on which state the string ends in (which could, due to nondeterminism, be any of them) and how much of the last  $v_q u_q$  is added. These corresponding strings are never shorter than the original string; therefore,  $|M_n^{pre}|_{\leq r} \leq |Q|(\ell + 1)|M|_{\leq r}$ . Thus,

$$\begin{aligned} h(M_n^{pre}) &= \limsup_{r \rightarrow \infty} \frac{\log |M_n^{pre}|_{\leq r}}{r} \\ &\leq \limsup_{r \rightarrow \infty} \frac{\log(|Q|(\ell + 1)|M|_{\leq r})}{r} \\ &= \limsup_{r \rightarrow \infty} \frac{\log |Q|(\ell + 1) + \log |M|_{\leq r}}{r} \\ &= \limsup_{r \rightarrow \infty} \frac{\log |Q|(\ell + 1)}{r} + \limsup_{r \rightarrow \infty} \frac{\log |M|_{\leq r}}{r} \\ &= \limsup_{r \rightarrow \infty} \frac{\log |M|_{\leq r}}{r} = h(M). \end{aligned}$$

We thus have a collection of subsets  $X_n$  of  $Y$  such that

$$\frac{n}{n + \ell} d_H(\bar{X}) = \frac{1}{\log k} \frac{n}{n + \ell} h(M) \leq \frac{1}{\log k} h(M_n^{pre}) = d_H(X_n) \leq \frac{1}{\log k} h(M) = d_H(\bar{X}).$$

Because  $\frac{n}{n+\ell} \rightarrow 1$  as  $n \rightarrow \infty$ , we conclude that  $d_H(X_n) \rightarrow d_H(\bar{X})$ . Therefore,  $d_H(Y) \geq \sup d_H(X_n) \geq d_H(\bar{X})$ , as required. -1

**COROLLARY 5.4.** *Let  $\mathcal{A}$  be a Büchi automaton recognizing an  $\omega$ -language  $L'$ . Assume that  $\mathcal{A}$  is trim and has one accept state and one start state, which are the same state. Let  $L$  be the language recognized by  $\mathcal{A}$  as a finite automaton. Then  $L' = L^\omega$ , and  $d_H(v_k(L')) = d_B(v_k(L'))$ .*

**PROOF.** Let  $X = v_k(L')$ . From Theorem 4.10, we know  $d_H(\overline{X}) = d_B(\overline{X}) = d_B(X)$ , so it suffices to show  $d_H(X) = d_H(\overline{X})$ ; in fact, we need only show  $d_H(\overline{X}) \leq d_H(X)$ .

Note that if an infinite string  $w$  belongs to  $L'$ , it must have a run that starts at the single start/accept state and revisits it infinitely many times; hence,  $w$  can be broken into infinitely many substrings, each of which starts and ends at this state. So  $L' = L^\omega$ . The second result then follows from Lemma 5.3.  $\dashv$

**LEMMA 5.5.** *Let  $L = C_q(\mathcal{A})$  be a cycle language for a Büchi automaton. Then  $d_H(v_k(L^\omega)) = \frac{1}{\log k} h(L)$ .*

**PROOF.** Let  $L'$  be the language  $C_q(\mathcal{A})$  accepts as a Büchi automaton. Then by Corollary 5.4, it suffices to show  $d_B(v_k(L')) = \frac{1}{\log k} h(L)$ . Moreover, by Lemma 3.5, it suffices to show  $h((L')^{pre}) = h(L)$ , where  $(L')^{pre}$  is the language of prefixes of strings in  $L'$ .

We claim that  $(L')^{pre} = L^{pre}$ . We have  $L^{pre} \subseteq (L')^{pre}$  because  $C_q(\mathcal{A})$  is trim, so any string that can be extended to a string in  $L^{pre}$  can be further extended to an infinite string in  $(L')^{pre}$ . We have  $(L')^{pre} \subseteq L^{pre}$  because any string in  $L'$  must have infinitely many prefixes in  $L$ , so we may extend a string in  $(L')^{pre}$  to an infinite string and then cut it off at a sufficiently late occurrence of an accept state.

So we have reduced the lemma to demonstrating that  $h(L) = h(L^{pre})$ , which is Fact 3.2.  $\dashv$

Combining the above with Lemma 5.2 gives us the theorem:

**THEOREM 5.6.** *Let  $\mathcal{A} = (Q, \Sigma, \delta, S, F)$  be a trim Büchi automaton, and let  $X = V_k(\mathcal{A})$ . Let  $F$  be the set of indices of accept states in  $\mathcal{A}$ . Then:*

- (i)  $d_H(X) = \frac{1}{\log k} \max_{q \in F} h(C_q(\mathcal{A}))$ .
- (ii)  $d_B(X) = \frac{1}{\log k} \max_{q \in Q} h(C_q(\mathcal{A}))$ .

**COROLLARY 5.7.** *Let  $\mathcal{A}$  be a trim Büchi automaton such that  $V_k(\mathcal{A}) = X$ . Then  $d_H(X) < d_B(X) = h(X)$  if and only if there exists a non-final state  $q \in Q \setminus F$  such that for cofinitely many  $m \in \mathbb{N}$  there exists  $n_m \in \mathbb{N}$  such that for each  $f \in F$ , the ratio of  $|C_f(\mathcal{A})|_{n_m}$ , which denotes the number of paths from  $f$  to itself of length  $n_m$ , to  $k^{n_m}$  is strictly less than the ratio of  $|C_q(\mathcal{A})|_{n_m}$  to  $k^{n_m}$ .*

**PROOF.** For the forward implication, we will illustrate the contrapositive; suppose that for every non-final state  $q \in Q \setminus F$  there is a final state  $f \in F$  such that  $\limsup_{n \rightarrow \infty} |\{\sigma : q \rightarrow q \mid \sigma \text{ is a path of length } n\}| \leq \limsup_{n \rightarrow \infty} |\{\sigma : f \rightarrow f \mid \sigma \text{ is a path of length } n\}|$ . If this is the case, we conclude that for all  $q \in Q$  we have  $h(C_q(\mathcal{A})) \leq h(C_f(\mathcal{A}))$  for some  $f \in F$ , where  $C_q(\mathcal{A})$  is the cycle language of  $q$  as defined in Definition 5.1. This follows because taking logarithms commutes with  $\limsup$ . By Theorem 5.6, we conclude that  $d_B(X) = d_H(X)$ , proving the contrapositive.

For the backwards implication, we suppose that there does exist some  $q \in Q \setminus F$  such that for cofinitely many  $m \in \mathbb{N}$  there exists  $n_m \in \mathbb{N}$  such that  $\frac{|C_f(\mathcal{A})|_m}{k^m}$  (the ratio of paths from  $f$  to itself of length  $m$  to  $k^m$ ) is strictly less than  $\frac{|C_q(\mathcal{A})|_{n_m}}{k^{n_m}}$  (the ratio of paths from  $q$  to itself of length  $n_m$  to  $k^{n_m}$ ) for each  $f \in F$ . Then it is evident that the sequence  $(n_m)_{m \in \mathbb{N}}$  witnesses that for all  $f \in F$  we have  $\limsup_{n \rightarrow \infty} \frac{|C_f|_n}{k^n} < \limsup_{n \rightarrow \infty} \frac{|C_q(\mathcal{A})|_n}{k^n}$ . Taking the logarithm of each side, we conclude  $h(C_f(\mathcal{A})) < h(C_q(\mathcal{A}))$  for all  $f \in F$ . Hence by Theorem 5.6, we get that  $d_H(X) < d_B(X)$ .  $\dashv$

**§6. Hausdorff measure.** As mentioned in Section 2.5, the Hausdorff dimension of a fractal  $X$  is defined by considering the  $s$ -dimensional Hausdorff measure  $\mu_H^s(X)$  for different values of  $s$ , and in particular,  $\mu_H^s(X)$  is  $\infty$  for  $s < d_H(X)$  and  $0$  for  $s > d_H(X)$ . In this section, we will give methods for determining the  $d_H(X)$ -dimensional Hausdorff measure of various types of  $k$ -automatic fractal  $X$  and make some observations that result from these methods. Note that this measure may be zero, a positive real number, or infinite.

We begin by leveraging former work of Merzenich and Staiger. The following comprises Lemma 15 and Procedure 1 of [16]:

**FACT 6.1.** *Let  $\mathcal{A}$  be a strongly connected deterministic Büchi automaton such that  $X = V_k(\mathcal{A})$  is closed. Then there is exactly one vector  $\vec{u}$  such that  $\vec{u}$  is an eigenvector corresponding to an eigenvalue of maximum magnitude of  $\mathbf{M}(\mathcal{A}, 0)$  and such that  $\vec{u}$  contains nonnegative entries, the maximum of which is 1. The Hausdorff measure  $\mu_H^{d_H(X)}(X)$  is the entry in  $\vec{u}$  corresponding to the start state of  $\mathcal{A}$ .*

Note that in particular, this implies  $0 \leq \mu_H^{d_H(X)}(X) \leq 1$  in this case.

The next step in this analysis is to extend this work to the case of a general strongly connected automaton. In particular, we might rightfully suspect that the Hausdorff dimension and measure of a set recognized by a strongly connected Büchi automaton are the same as those of the set's closure. It is most convenient to show this first for the dimensions and then for the measures.

**LEMMA 6.2.** *If  $\mathcal{A}$  is a strongly connected Büchi automaton, then  $d_H(V_k(\mathcal{A})) = d_H(\overline{V_k(\mathcal{A})})$ .*

**PROOF.** Note that removing accept states from  $\mathcal{A}$  can only make the resulting Hausdorff dimension lower, so it suffices to consider the case where  $\mathcal{A}$  has only one accept state. Similarly, by Lemma 4.5, it suffices to assume  $\mathcal{A}$  has only one start state, which is also its accept state.

In this case,  $\mathcal{A}$  satisfies the conditions of Corollary 5.4; so  $d_H(V_k(\mathcal{A})) = d_B(V_k(\mathcal{A}))$ . This is then equal to  $d_B(\overline{V_k(\mathcal{A})})$ , which is in turn equal to  $d_H(\overline{V_k(\mathcal{A})})$  by Theorem 4.10, because  $\overline{V_k(\mathcal{A})}$  is a closed  $k$ -automatic set.  $\dashv$

We show the analogous result for Hausdorff measures by first noting that when such an  $X$  is embedded in its closure, the set-difference has lower dimension and hence is null in the higher-dimensional measure.

**LEMMA 6.3.** *Let  $\mathcal{A}$  be a strongly connected Büchi automaton, and let  $X = V_k(\mathcal{A})$ . If  $d_H(X)$  is positive, then  $d_H(\overline{X} \setminus X) < d_H(X)$ .*



PROOF. First, we let  $\bar{L}$  be the  $\omega$ -language accepted by  $\bar{\mathcal{A}}$ . Note that  $v_k(\bar{L}) = \bar{X}$ , so an element of  $\bar{X} \setminus X$  must correspond to an infinite string on which  $\mathcal{A}$  passes through finitely many accept states before infinitely passing through exclusively non-accepting states. Let  $L'$  be the language of strings for which every infinite run is of this form (and for which there is at least one such infinite run). This correspondence is not exact; it is possible that if an element of  $\bar{X} \setminus X$  is  $k$ -rational, one of its representations is in  $L'$ , while the other is not. This will not affect an analysis of Hausdorff dimension, because (nonzero) Hausdorff dimension does not change based on the membership of a countable set. Our goal is then to show that  $d_H(v_k(L')) < d_H(X)$ .

Let  $F^c = Q \setminus F$  be the set of non-accepting states of  $\mathcal{A}$ , and let  $L_q$  for  $q \in F^c$  be the set of infinite strings for which every infinite run starting at state  $q$  only passes through non-accepting states in  $\mathcal{A}$  (and for which there is at least one such infinite run). Since every string in  $L'$  has a tail in  $L_q$  for some  $q$ , we know  $v_k(L')$  is a countable union of scaled copies of  $v_k(L_q)$ . So it suffices to show that  $d_H(v_k(L_q)) < d_H(\bar{X})$  for all  $q \in F^c$ . We will show a stronger statement, that  $d_H(\overline{v_k(L_q)}) < d_H(\bar{X})$ . By Theorem 4.10, it suffices to show the corresponding entropy statement, that  $h(L_q^{pre}) < h(\bar{L}^{pre})$ .

Let  $\log \alpha = h(L_q^{pre})$ . By Lemma 3.3,  $\log \alpha = \lim_{n \rightarrow \infty} \frac{\log |L_q^{pre}|_n}{n}$ . For every non-accepting state  $q'$  of  $\mathcal{A}$ , there exists a string on which  $\mathcal{A}$  cycles from state  $q'$  to itself while passing through an accept state. Choose such a cycle for each  $q' \in F^c$ , and let  $m$  be the least common multiple of their lengths; by repeating the cycles if necessary, we may assume they all have the same length  $m$ .

Let  $M$  be the  $\omega$ -language defined as follows: every string in  $M$  is made up of blocks  $m$  characters long, with each block a “normal block” or a “cycle block.” The normal blocks, when taken together with the cycle blocks removed, must form a string in  $L_q$ . Each cycle block corresponds to one of the cycles of length  $m$  mentioned above, in particular the cycle corresponding to a state in which a run on the prefix terminates. Note that when every string in  $M$  is prefixed by a constant string that corresponds to a path from the start state of  $\mathcal{A}$  to state  $q$  (which does not affect entropy), the result is a subset of  $\bar{L}$ . So  $h(M^{pre}) \leq h(\bar{L}^{pre})$ , and it suffices to show  $h(L_q^{pre}) < h(M^{pre})$ .

Now for every string in  $L_q^{pre}$ , choose a run that witnesses this. Let  $k$  be a positive integer, and let  $0 \leq r \leq k$ . For any string in  $L_q^{pre}$  of length  $rm$ , we may apply the chosen run and insert  $(k - r)$  cycle blocks at any choice of block boundaries in order to produce a string in  $M^{pre}$  of length  $km$ . Therefore,

$$|M^{pre}|_{km} \geq \sum_{r=0}^k \binom{k}{r} |L_q^{pre}|_{rm}.$$

Let  $1 < \beta < \alpha$ . One can check that eventually  $|L_q^{pre}|_n > \beta^n$ . So for sufficiently large  $k$ ,

$$|M^{pre}|_{km} > \sum_{r=0}^k \binom{k}{r} \beta^{rm} = (\beta^m + 1)^k.$$

So,

$$\begin{aligned}
 h(M^{pre}) &= \limsup_{n \rightarrow \infty} \frac{\log |M^{pre}|_n}{n} \\
 &\geq \limsup_{k \rightarrow \infty} \frac{\log |M^{pre}|_{km}}{km} \\
 &\geq \limsup_{k \rightarrow \infty} \frac{\log ((\beta^m + 1)^k)}{km} \\
 &= \limsup_{k \rightarrow \infty} \frac{k \log (\beta^m + 1)}{km} \\
 &= \frac{\log (\beta^m + 1)}{m}.
 \end{aligned}$$

So for all  $1 < \beta < \alpha$ , we have  $h(M^{pre}) \geq \frac{\log(\beta^m + 1)}{m}$ . Thus,

$$h(M^{pre}) \geq \lim_{\beta \nearrow \alpha} \frac{\log (\beta^m + 1)}{m} = \frac{\log (\alpha^m + 1)}{m} > \frac{\log (\alpha^m)}{m} = \log \alpha = h(L_q^{pre}).$$

This concludes the proof. ⊖

**COROLLARY 6.4.** *Let  $\mathcal{A}$  be a strongly connected Büchi automaton, and let  $X = V_k(\mathcal{A})$  have Hausdorff dimension  $d_H(X) = \alpha$ . Then  $\mu_H^\alpha(X) = \mu_H^\alpha(\overline{X})$ .*

**PROOF.** We examine two cases. If  $\alpha$  is positive, then the previous lemma gives us  $d_H(\overline{X} \setminus X) < \alpha$ ; hence,  $\mu_H^\alpha(\overline{X} \setminus X) = 0$ . The result then follows from additivity of Hausdorff measure.

Assume on the other hand that  $\alpha = 0$ ; then  $\mu_H^\alpha$  is the counting measure. If  $X$  is infinite, then so is  $\overline{X}$ ; hence, the counting measure agrees for both. If  $X$  is finite, then  $\overline{X} = X$ , as any finite set in a metric space is closed. So in either case  $\mu_H^0(X) = \mu_H^0(\overline{X})$ . ⊖

Our goal is now to extend this result and compute the Hausdorff measure of any  $k$ -automatic fractal. We will do this by making use of a class of Büchi automata called *unambiguous* Büchi automata:

**DEFINITION 6.5.** Let  $\mathcal{A}$  be a Büchi automaton. We say that  $\mathcal{A}$  is *unambiguous* if for every infinite string  $w$  accepted by  $\mathcal{A}$ , there is exactly one accepting run of  $\mathcal{A}$  on  $w$ .

Note that any deterministic Büchi automaton is unambiguous. But in the case of Büchi automata, we know that nondeterminism is sometimes necessary in order to achieve full computational capability. Fortunately, a result of [5] tells us that this capability can still be fully realized in the unambiguous case:

**FACT 6.6.** *Let  $L$  be a regular  $\omega$ -language. Then there is an unambiguous Büchi automaton  $\mathcal{A}$  with  $L(\mathcal{A}) = L$ . Moreover, given any Büchi automaton for  $L$ , we can effectively convert it to an unambiguous automaton.*

We will now extend Corollary 6.4, giving us a method to compute the Hausdorff measure of any  $k$ -automatic fractal. Using the previous fact, we may assume that the Büchi automaton defining the fractal is unambiguous. Next, it is useful for us

to define functions associating each infinite string in an  $\omega$ -language  $L$  with its *key state* and *key prefix*:

DEFINITION 6.7. Let  $\mathcal{A}$  be an unambiguous Büchi automaton with set of states  $Q$  and recognizing an  $\omega$ -language  $L$ . The function  $k_{\mathcal{A}} : L \rightarrow Q$  maps a string  $w$  to the unique *key state*  $q$  such that when  $\mathcal{A}$  performs an accepting run on  $w$ :

- the run passes through state  $q$ ;
- after first passing through state  $q$ , the run never leaves its strongly connected component;
- before first passing through state  $q$ , the run never enters its strongly connected component.

The function  $p_{\mathcal{A}} : L \rightarrow \Sigma^*$  maps  $w$  to the *key prefix* of  $w$  running from the start state of  $\mathcal{A}$  to the first occurrence of  $k_{\mathcal{A}}(w)$ .

Note that when a run leaves a strongly connected component, it can never return to it, or else the states in between would also be a part of the same component, a contradiction. So  $k_{\mathcal{A}}$  and  $p_{\mathcal{A}}$  are well-defined functions.

LEMMA 6.8. Let  $\mathcal{A}$  be an unambiguous Büchi automaton with set of states  $Q$  and recognizing an  $\omega$ -language  $L$ . Let  $Q' \subseteq Q$  be the set of states whose strongly connected component contains an accept state. For each  $q \in Q'$ , let  $\mathcal{A}_q$  be the automaton created by moving the start state of  $\mathcal{A}$  to  $q$  and removing all transitions out of its strongly connected component, and let  $L_q$  be the  $\omega$ -language it accepts. Let  $M_q$  be the  $\omega$ -language containing all strings  $w \in L$  with  $k_{\mathcal{A}}(w) = q$ . Then  $d_H(v_k(M_q)) = d_H(v_k(L_q))$ , assuming  $M_q \neq \emptyset$ . Moreover, let  $P_q = \{p_{\mathcal{A}}(w) : w \in M_q\}$ ; then for any  $\alpha$ ,

$$\mu_H^\alpha(M_q) = \sum_{u \in P_q} \frac{\mu_H^\alpha(L_q)}{k^{\alpha|u|}}$$

(here  $|u|$  is the length of  $u$ ).

PROOF. Let  $M_{q,u}$  be the set of  $w \in L$  with  $k_{\mathcal{A}}(w) = q$  and  $p_{\mathcal{A}}(w) = u$ . Such strings decompose as  $w = uv$ , where  $v$  is a string that, starting from state  $q$ , passes through accept states in its strongly connected component infinitely often without leaving the component. In other words,  $M_{q,u} = uL_q$ . It follows that  $v_k(M_{q,u})$  is an affine copy of  $v_k(L_q)$ , which is scaled down with a factor of  $f = \frac{1}{k^{|u|}}$ . By properties of Hausdorff measure,  $\mu_H^\alpha(M_{q,u}) = \frac{\mu_H^\alpha(L_q)}{k^{\alpha|u|}}$ .

Note then that  $M_q = \bigcup_{u \in P_q} M_{q,u}$ , because  $M_{q,u}$  is empty for  $u \notin P_q$ . Since  $P_q \subseteq \Sigma^*$  is a countable set, we may apply countable additivity to get the desired formula. ⊖

This process may then be applied to every key state to produce the following result:

THEOREM 6.9. Let  $\mathcal{A}$ ,  $Q$ ,  $Q'$ ,  $L$ ,  $L_q$ , and  $M_q$  be as in the previous lemma. Then:

- (i)  $d_H(v_k(L)) = \max_{q \in Q'} d_H(v_k(L_q))$ ,
- (ii) with  $\alpha = d_H(v_k(L))$ ,  $\mu_H^\alpha(L) = \sum_{q \in Q'} \mu_H^\alpha(M_q)$ .

PROOF. For (i), we need only note that from the previous lemma  $d_H(v_k(L_q)) = d_H(v_k(M_q))$  and that  $L = \bigcup_{q \in Q'} M_q$ . Then (ii) also follows from  $L = \bigcup_{q \in Q} M_q$ .  $\dashv$

**§7. Model-theoretic consequences.** For the basics of first-order logic and model theory, see [13], from which we also take our notation and conventions. We recall the following correspondence between subsets of  $\mathbb{R}$  recognized by Büchi automata and definable subsets of the first-order structure  $\mathcal{R}_k = (\mathbb{R}, <, +, \mathbb{Z}, X_k(x, u, d))$  where the ternary predicate  $X_k(x, u, d)$  holds precisely if  $u$  is an integer power of  $k$ , and in the  $k$ -ary expansion of  $x$  the digit specified by  $u$  is  $d$ , i.e.,  $d$  is the coefficient of  $u$  in the expansion. In particular, note that  $u \in k^{\mathbb{Z}}$  and  $d \in \{0, \dots, k - 1\}$  in any tuple  $(x, u, d)$  such that  $X_k(x, u, d)$  holds.

**THEOREM 7.1 [3].** *For each  $n \in \mathbb{N}$ , any  $A \subseteq \mathbb{R}^n$  is  $k$ -automatic if and only if  $A$  is definable in  $\mathcal{R}_k$ .*

In conjunction with results from [2], we obtain the following consequences for definable subsets of the structure  $\mathcal{R}_k$ . Below, by ‘‘Cantor set’’ we mean a nonempty compact subset of  $\mathbb{R}$  that has no isolated points and has no interior.

**LEMMA 7.2.** *Suppose that  $\mathcal{A}$  is a trim Büchi automaton with  $m$  states on alphabet  $[k]$  that has one accept state, which is also the start state. Let  $L$  be the language it accepts as a Büchi automaton. Suppose that  $w \in [k]^n$  is not the prefix of any word in the language  $L$ . Then every subset of  $[0, 1]$  of the form  $(v_k(w0^\omega), v_k(w(k-1)^\omega))$ , with  $w \in [k]^\ell$ ,  $\ell \in \mathbb{N}$ , has a subinterval of size at least  $k^{-(\ell+m+n)}$  disjoint from  $v_k(L)$ . Moreover, the box-counting dimension of  $v_k(L)$  is at most  $\log_{k^{n+m}}(k^{n+m} - 1) < 1$ .*

PROOF. For each  $q \in Q$  we let  $s_q$  be a minimal path from  $q$  back to the start state. We know a minimal path visits each state at most once. For a prefix  $w \in [k]^*$  such that a run on  $\mathcal{A}$  stops at state  $q$ , the interval  $\tilde{I} = (v_k(ws_qv0^\omega), v_k(ws_qv(k-1)^\omega))$  cannot be in  $v_k(L)$ . Observe that the length of  $\tilde{I}$  is at least  $k^{-(\ell+m+n)}$ .

For the ‘‘moreover,’’ observe that by Lemmas 3.5 and 4.1, we have the following:

$$d_B(v_k(L)) = d_B(\overline{v_k(L)}) = \frac{1}{\log(k)} h(L^{pre}) = \lim_{d \rightarrow \infty} \frac{\log(|L^{pre}|_d)}{d \log(k)} = \lim_{d \rightarrow \infty} \frac{\log(|L^{pre}|_{d(n+m)})}{d(n+m) \log(k)}.$$

We proceed by induction on  $d$ . For the base case, note  $|L^{pre}|_{n+m} \leq k^m(k^n - 1) < k^{n+m} - 1$ . For  $d > 1$ , we know  $|L^{pre}|_{d(n+m)} \leq |L^{pre}|_{(d-1)(n+m)}(k^{n+m} - 1)$  because for each  $w \in |L^{pre}|_{(d-1)(n+m)}$  there exists  $q' \in Q$  such that  $ws_{q'}v$  is not in  $|L^{pre}|_{d(n+m)}$ . By induction, we get the following:

$$\lim_{d \rightarrow \infty} \frac{\log(|L^{pre}|_{d(n+m)})}{d(n+m) \log(k)} \leq \lim_{d \rightarrow \infty} \frac{\log((k^{n+m} - 1)^d)}{d(n+m) \log(k)},$$

and hence  $d_B(v_k(L)) \leq \log_{k^{n+m}}(k^{n+m} - 1)$ .  $\dashv$

Recall that  $\mu_H^1$  is the Hausdorff 1-measure, which equals Lebesgue measure on subsets of  $\mathbb{R}$ .

**LEMMA 7.3.** *Suppose that  $X \subseteq [0, 1]$  is a  $k$ -automatic set with Hausdorff dimension 1, and  $X$  is recognized by Büchi automaton  $\mathcal{A}$ . If  $\mu_H^1(X \cap I) > 0$  for every interval  $I \subseteq [0, 1]$  with  $k$ -rational endpoints, then there is  $\varepsilon > 0$  such that for each such  $I$  we have  $\mu_H^1(X \cap I) \geq \varepsilon \cdot \text{diam}(I)$ .*

PROOF. Let  $m$  be the number of states in  $\mathcal{A}$ . By hypothesis the Hausdorff dimension of  $X \cap I$  is 1 for every interval  $I \subseteq [0, 1]$  with  $k$ -rational endpoints, and  $\mu_H^1(X \cap I)$  is positive. This means that for each prefix  $w$  of  $L(\mathcal{A})$ , there exists a strongly connected component  $\mathcal{C}_q(\mathcal{A})$  that contains an accept state  $q$  and for which the image under  $V_k$  of the set of words with prefix  $w$  that are accepted via a run contained in  $\mathcal{C}_q(\mathcal{A})$  has positive Hausdorff 1-measure. In particular, there is a word  $v$  with length at most  $m + |w|$  such that  $v$  extends  $w$  and  $\mu_H^1(v_k(v\mathcal{C}_q(\mathcal{A})^\omega)) > 0$ , where  $\mathcal{C}_q(\mathcal{A})$  is the cycle language for  $\mathcal{C}_q(\mathcal{A})$ , assuming without loss of generality that a run of  $v$  terminates at state  $q$ . We set  $a$  to be the minimum of  $\mu_H^1(V_k(\mathcal{C}_q(\mathcal{A})))$  where  $q$  ranges over all states in a strongly connected component of  $\mathcal{A}$  that contains an accept state and has positive Hausdorff 1-measure. We let  $\varepsilon = a \cdot k^{-m}$ , and conclude that for each  $I \subseteq [0, 1]$ , we have  $\mu_H^1(X \cap I) \geq \varepsilon \cdot \text{diam}(I)$ , as desired.  $\dashv$

THEOREM 7.4. *Suppose  $X \subseteq [0, 1]^n$  is  $k$ -automatic. There exists a set  $A \subseteq [0, 1]$  definable in  $(\mathbb{R}, <, +, X)$  such that  $d_B(A) \neq d_H(A)$  if and only if either a Cantor set is definable in  $(\mathbb{R}, <, +, X)$  or a set that is both dense and codense on an interval is definable in  $(\mathbb{R}, <, +, X)$ .*

PROOF. ( $\Leftarrow$ ) Suppose that  $C$ , a Cantor set, is definable in the structure  $(\mathbb{R}, <, +, X)$ . Define the set  $A = \{x \in C : \exists y \in C : [x < y \wedge (x, y) \cap C = \emptyset] \vee [y < x \wedge (y, x) \cap C = \emptyset]\}$ ;  $A$  is first-order definable in  $(\mathbb{R}, <, +, X)$ . Then  $A$  is the set of endpoints of the complementary intervals (e.g., the endpoints of the middle thirds that are removed in constructing the ternary Cantor set), which form a countable set whose closure is the entire Cantor set. This follows since each element of a Cantor set can be approximated by endpoints of the complement within  $[0, 1]$ . We conclude that  $d_H(A) = 0$  and  $d_B(A) = d_B(\overline{A}) \neq 0$  by Lemma 6.7 in [2].

Now suppose that  $D$  is a set that is dense and codense in an interval  $J \subseteq [0, 1]$ , and is definable in  $\mathcal{R}_k$ . Note that if there exists any interval  $I \subseteq J$  such that  $d_H(D \cap I) < 1$ , then because  $D$  is dense in  $J$  and hence also in  $I$ , we conclude that  $d_B(D \cap I) = d_H(\overline{D \cap I}) = 1 > d_H(D \cap I)$ , and we are done. Similarly if  $d_H(\overline{D} \setminus D \cap I) < 1$  for some interval  $I \subseteq J$ , since  $D$  is codense in  $J$ .

The only case remaining is the one in which for every interval  $I \subseteq J$ , we know  $d_H(D \cap I) = 1$  and  $d_H(\overline{D} \setminus D \cap I) = 1$ . If this is the case then both  $\mathcal{A}$ , the automaton that recognizes  $D$ , and  $\mathcal{A}_c$ , the automaton that recognizes  $\overline{D} \setminus D$ , each have a strongly connected component containing an accept state whose cycle language has the same Hausdorff dimension as the whole automaton, namely Hausdorff dimension 1. Suppose that  $S_q(\mathcal{A})$  is the strongly connected automaton that recognizes the cycle language of state  $q$  in  $\mathcal{A}$  that has the same Hausdorff dimension as  $D$ . Similarly suppose that  $S_{q'}(\mathcal{A}_c)$  is the strongly connected automaton that recognizes the cycle language of state  $q'$  that has the same Hausdorff dimension as  $\overline{D} \setminus D$ .

If the image of  $S_q(\mathcal{A})$  under  $V_k$  is nowhere dense in  $[0, 1]$ , then  $S_q(\mathcal{A})$  satisfies the hypotheses of Lemma 7.2, since being nowhere dense necessitates that at least one finite word is not a prefix of the language. Similar logic holds for  $S_{q'}(\mathcal{A}_c)$ . So by Lemma 7.2, they would have box-counting dimension less than one, and hence their closures would as well. This would contradict our assumption that they each respectively witness the Hausdorff dimensions of  $\mathcal{A}$  and  $\mathcal{A}_c$  being one, since

Hausdorff dimension is bounded above by box-counting dimension, so  $S_q(\mathcal{A})$  and  $S_{q'}(\mathcal{A}_c)$  cannot have box-counting dimension less than one.

Therefore  $S_q(\mathcal{A})$  and  $S_{q'}(\mathcal{A}_c)$  both must recognize somewhere dense sets. Hence the closure of each must contain an interval and thus has positive Lebesgue measure. By Corollary 6.4 we know that both  $S_q(\mathcal{A})$  and  $S_{q'}(\mathcal{A}_c)$  have the same Hausdorff 1-measure as their closures (as Büchi automata). So by Lemma 6.8, both  $D$  and  $\overline{D} \setminus D$  have positive Lebesgue measure, which equals Hausdorff 1-measure, on every subinterval  $I \subseteq J$ . In particular, both  $D \cap I$  and  $(\overline{D} \setminus D) \cap I$  have positive Lebesgue measure for every  $I \subseteq J$ .

Fixing one such  $I$ , we may assume without loss of generality that  $\overline{D}$  is an interval. Applying Lemma 7.3, we know that both  $D \cap I$  and  $\overline{D} \setminus D \cap I$  have positive Lebesgue measure bounded uniformly below by  $\varepsilon \cdot \text{diam}(I)$  and  $\varepsilon' \cdot \text{diam}(I)$  respectively, for fixed  $\varepsilon, \varepsilon' > 0$  given to us by Lemma 7.3. Since  $\overline{D} \setminus D \cap I$  is a subset of the complement of  $D \cap I$ , we conclude  $\varepsilon' \leq 1 - \varepsilon$ , and also get the following:

$$\varepsilon' \cdot \text{diam}(I) \leq \mu_H^1(D \cap I) \leq \varepsilon \text{diam}(I).$$

By the Lebesgue density theorem, since  $D \cap I$  is dense in  $I$  for each  $I$  with  $k$ -rational endpoints, we know that

$$\lim_{\text{diam}(I) \rightarrow 0} \frac{\mu_H^1(D \cap I)}{\text{diam}(I)}$$

exists and equals 1. Yet we observe that  $\varepsilon' \leq \frac{\mu_H^1(D \cap I)}{\text{diam}(I)} \leq \varepsilon$ , so we must have  $\varepsilon = 1$ . But we conclude the same holds for  $\frac{\mu_H^1(\overline{D} \setminus D \cap I)}{\text{diam}(I)}$ , so we also need  $\varepsilon' = 1$ . This contradicts the fact that  $\varepsilon' \leq 1 - \varepsilon$ , and we are done.

( $\implies$ ) Suppose that there is some set  $A \subseteq [0, 1]$  definable in  $(\mathbb{R}, <, +, X)$  such that  $d_B(A) \neq d_H(A)$ . Then it must be the case that  $A \neq \overline{A}$ . Let  $\mathcal{A}$  be a trim Büchi automaton recognizing  $A$ . Note that we know  $A$  has no interior. Suppose it is not the case that  $A$  is somewhere dense, i.e.,  $A$  is nowhere dense. Then, we may pass from  $A$  to  $\overline{A}$ , and since  $A$  is nowhere dense we know that  $\overline{A}$  is nowhere dense as well.

We know that  $d_B(\overline{A}) = d_B(A) > 0$ ; otherwise, since  $d_H(A) \leq d_B(A)$ , they would have to agree on  $A$ . If  $d_B(\overline{A}) = 1$ , then  $d_H(\overline{A}) = 1$  by Theorem 4.10. This implies that there is a state  $q$  in  $\overline{A}$  such that  $d_H(V_k(S_q(\overline{A}))) = 1$  by Lemma 6.2. In order for  $\overline{A}$  to be nowhere dense, it must be the case that each strongly connected component of  $\overline{A}$  omits at least one string of finite length as a prefix. Otherwise, every finite-length string is the prefix of some string in  $C_q(\overline{A})$ . Since there exists a string  $w \in [k]^*$  such that  $v_k(wC_q(\overline{A})^\omega) \subseteq \overline{A}$ , this would imply that  $\overline{A}$  is dense in (and in fact, since  $\overline{A}$  is closed, contains) the interval  $v_k(w[k]^*)$ , contradicting that  $\overline{A}$  is nowhere dense. So for each state  $q$  in  $\overline{A}$ , let  $v_q \in [k]^*$  be a word omitted from the prefixes of  $C_q(\overline{A})$  whose length is minimal among such finite words. By Lemma 7.2, we conclude that  $V_k(S_q(\overline{A}))$  has Hausdorff dimension less than 1 for all states  $q$  in  $\overline{A}$ . So by Lemma 4.12, we know that  $d_H(\overline{A})$  is the maximum of  $d_H(V_k(S_q(\overline{A})))$  for all states  $q$  in  $\overline{A}$ , and we conclude that  $1 > d_H(\overline{A}) > 0$ . By [2], there is a definable Cantor set in  $(\mathbb{R}, <, +, X)$ .

In the case that  $A$  is somewhere dense, say on interval  $I \subseteq [0, 1]$ , suppose it is not also codense in  $I$ . Then some subinterval of  $I$  is contained entirely in  $A$ , so  $d_H(A) = 1 = d_B(A)$ , a contradiction.  $\dashv$

**Acknowledgments.** Many thanks to Philipp Hieronymi for the interesting ideas, questions, and generous guidance provided for this paper. Thanks to Elliot Kaplan, Jason Bell, and Rahim Moosa for their helpful comments, suggestions, and questions regarding the contents of this paper.

**Funding.** We gratefully acknowledge that this research was supported by the Fields Institute for Research in Mathematical Sciences. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Institute. The first author was partially supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE—1746047. The second author was partially supported by the National Science Foundation Grant No. DMS—1654725.

#### REFERENCES

- [1] B. ADAMCZEWSKI and J. BELL, *An analogue of Cobham's theorem for fractals*. *Transactions of the American Mathematical Society*, vol. 363 (2011), pp. 4421–4442.
- [2] A. BLOCK GORMAN, *Pairs and predicates in expansions of o-minimal structures*, Ph.D. thesis, University of Illinois at Urbana–Champaign, 2021.
- [3] B. BOIGELOT, S. RASSART, and P. WOLPER, *On the expressiveness of real and integer arithmetic automata (extended abstract)*, *Proceedings of the 25th International Colloquium on Automata, Languages and Programming*, Springer, Berlin, Heidelberg, 1998, pp. 152–163.
- [4] J. R. BÜCHI, *On a decision method in restricted second order arithmetic*, *Proceedings of the International Congress on Logic, Method, and Philosophy of Science*, Stanford University Press, Stanford, 1962, pp. 1–12.
- [5] O. CARTON and M. MICHEL, *Unambiguous Büchi automata*. *Theoretical Computer Science*, vol. 297 (2003), pp. 37–81.
- [6] E. CHARLIER, J. LEROY, and M. RIGO, *An analogue of Cobham's theorem for graph directed iterated function systems*. *Advances in Mathematics*, vol. 280 (2015), pp. 86–120.
- [7] N. CHOMSKY and G. MILLER, *Finite state languages*. *Information and Control*, vol. 1 (1958), pp. 91–112.
- [8] G. EDGAR, *Measure, Topology, and Fractal Geometry*, second ed., Undergraduate Texts in Mathematics, Springer, New York, 2008.
- [9] J. EVANS, *The entropy and Hausdorff dimension of self-similar sets*. *Proceedings of the American Mathematical Society*, vol. 149 (2021), pp. 4387–4396.
- [10] K. FALCONER, *Fractal geometry: Mathematical foundations and applications*, Wiley, New York, 2013.
- [11] G. HANSEL, D. PERRIN, and I. SIMON, *Compression and entropy*, *Stacs 92* (A. Finkel and M. Jantzen, editors), Springer, Berlin–Heidelberg, 1992, pp. 513–528.
- [12] P. HIERONYMI and E. WALSBERG, *Interpreting the monadic second order theory of one successor in expansions of the real line*. *Israel Journal of Mathematics*, vol. 224 (2018), pp. 39–55.
- [13] D. MARKER, *Model Theory: An Introduction*, Graduate Texts in Mathematics, Springer, New York, 2006.
- [14] R. D. MAULDIN and S. C. WILLIAMS, *Hausdorff dimension in graph directed constructions*. *Transactions of the American Mathematical Society*, vol. 309 (1988), pp. 811–829.
- [15] R. MCNAUGHTON, *Testing and generating infinite sequences by a finite automaton*. *Information and Control*, vol. 9 (1966), no. 5, pp. 521–530.
- [16] W. MERZENICH and L. STAIGER, *Fractals, dimension, and formal languages*. *RAIRO—Theoretical Informatics and Applications*, 1993, vol. 28, pp. 361–386.



[17] A. PARKER, K. YANCEY, and M. YANCEY, *Regular language distance and entropy*, **42nd International Symposium on Mathematical Foundations of Computer Science (MFCS2017)** (K. G. Larsen, H. L. Bodlaender, and J.-F. Raskin, editors), Leibniz International Proceedings in Informatics, Schloss Dagstuhl—Leibniz-Zentrum fuer Informatik, Dagstuhl, 2017, pp. 3:1–3:14.

[18] D. PERRIN and J. É. PIN, *Infinite Words: Automata, Semigroups, Logic and Games*, Elsevier, Amsterdam, 2004.

[19] C. E. SHANNON, *A mathematical theory of communication*. *The Bell System Technical Journal*, vol. 27 (1948), no. 3, pp. 379–423.

[20] L. STAIGER, *The entropy of finite  $\omega$ -languages*. *Problems of Control and Information Theory*, vol. 14 (1985), no. 5, pp. 383–392.

[21] ———, *Combinatorial properties of the Hausdorff dimension*. *Journal of Statistical Planning and Inference*, vol. 23 (1989), pp. 95–100.

DEPARTMENT OF MATHEMATICS  
THE OHIO STATE UNIVERSITY  
231 W. 18TH AVE.  
COLUMBUS, OH 43210, USA

*E-mail:* [blockgorman.1@osu.edu](mailto:blockgorman.1@osu.edu)

PURE MATHEMATICS  
UNIVERSITY OF WATERLOO  
200 UNIVERSITY AVENUE WEST  
WATERLOO, ON N2L 3G1  
CANADA

*E-mail:* [c2schulz@uwaterloo.ca](mailto:c2schulz@uwaterloo.ca)