

## TOPOLOGICAL BAR-CODES OF FRACTALS: A NEW CHARACTERIZATION OF SYMMETRIC BINARY FRACTAL TREES

TARA D. TAYLOR

*Department of Mathematics, Statistics and Computer Science  
St. Francis Xavier University  
Antigonish, Nova Scotia, Canada  
E-mail: ttaylor@stfx.ca*

**Abstract.** The goal of this paper is to provide foundations for a new way to classify and characterize fractals using methods of computational topology. The fractal dimension is a main characteristic of fractal-like objects, and has proved to be a very useful tool for applications. However, it does not fully characterize a fractal. We can obtain fractals with the same dimension that are quite different topologically. Motivated by techniques from shape theory and computational topology, we consider fractals along with their  $\epsilon$ -hulls as  $\epsilon$  ranges over the non-negative real numbers. In particular, we develop theory for the class of non-overlapping symmetric binary fractal trees that can be generalized to broader classes of fractals. We investigate various features of the  $\epsilon$ -hulls of the trees, based on the holes in these hulls. We determine the hole sequence of these trees together with the persistence intervals of the holes as the ‘topological bar-codes’ of these fractals. We provide quantitative results for a selection of specific trees to illustrate the theory. Finally, we prove that for non-overlapping symmetric binary fractal trees, the growth rate of holes in  $\epsilon$ -hulls is equal to the similarity dimension.

**1. Introduction, background and motivation.** To date, the main characteristic of a fractal-like object is its fractal dimension. The fractal dimension has proved to be a very useful tool for applications [8], [11]. However, the fractal dimension does not fully characterize a fractal, since it is possible to obtain two fractals with the same fractal dimensions but different topologies. To demonstrate this point, we consider the well-known Sierpiński gasket, displayed in Figure 1, and its relatives. One way to describe the

---

2000 *Mathematics Subject Classification*: Primary 28A80; Secondary 55N99.

*Key words and phrases*: fractals, fractal trees, topological bar-codes, epsilon-hulls, homology. Research of the author supported by the Natural Sciences and Engineering Research Council of Canada and St. Francis Xavier University. The initial research for this paper was done during doctoral research under the supervision of D. Pronk of Dalhousie University. The paper is in final form and no version of it will be published elsewhere.

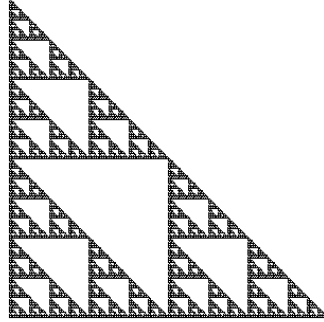


Fig. 1. The Sierpiński gasket

Sierpiński gasket is in terms of an *iterated function system* (IFS). Let

$$\begin{aligned} (1) \quad & f_1(x, y) = \frac{1}{2}(x, y), \\ (2) \quad & f_2(x, y) = \frac{1}{2}[(x, y) + (1, 0)], \\ (3) \quad & f_3(x, y) = \frac{1}{2}[(x, y) + (0, 1)]. \end{aligned}$$

The maps  $f_i$  are contractive similarities, and the IFS  $\{f_1, f_2, f_3\}$  has the Sierpiński gasket as its unique attractor [11]. It is straightforward to use the self-similarity of the Sierpiński gasket to calculate its fractal dimension to be  $\ln 3 / \ln 2 \approx 1.585$ .

In general, an iterated function system consists of a finite set  $\{\phi_1, \dots, \phi_N\}$  of contractive similarities of  $\mathbb{R}^d$  (for  $d = 1, 2, 3, \dots$ ). Let  $D$  be the unique real number for which

$$(4) \quad \sum_{i=1}^N r_i^D = 1.$$

There exists a unique nonempty compact subset  $K$  of  $\mathbb{R}^d$ , called the attractor, which is invariant with respect to the contractions [7]:

$$K = \bigcup_{i=1}^N \phi_i(K).$$

We say that  $K$  is strictly self-similar. For example, the Sierpiński gasket and its relatives (described in the next paragraph) are all strictly self-similar. The set  $K$  is said to satisfy the *open set condition* if there exists a bounded open set  $G \in \mathbb{R}^d$  such that

1.  $\phi_i(G) \subseteq G$  for each  $i$ ,
2.  $\phi_i(G) \cap \phi_j(G) = \emptyset$  for  $i \neq j$ .

When the open set condition is met,  $K$  has both Hausdorff and Minkowski dimensions equal to the solution  $D$  in Eq. (4) [3], [7].

The three maps of the IFS for the Sierpiński gasket each map the unit square to a square with sides of length  $1/2$ , while preserving the orientation of the square, as

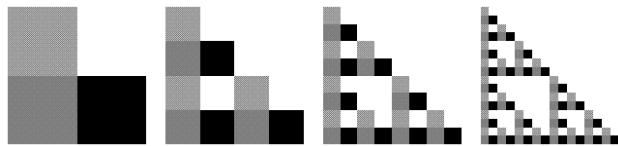


Fig. 2. First four iterations to obtain the Sierpiński gasket

displayed in Figure 2. A relative of the Sierpiński gasket is an attractor of an IFS that consists of three maps that all map the unit square to the three smaller squares with sides of length  $1/2$ , and each of the three maps may also involve a symmetry transformation of the square (reflections or rotations) [11]. An example of such a relative is displayed in Figure 3. The Sierpiński gasket and its 456 relatives (removing duplicates) [11] all have the same fractal dimension  $(\ln 3 / \ln 2)$ . Some relatives are totally disconnected, some are simply-connected, and some are multiply-connected, see Figure 4.



Fig. 3. A relative of the Sierpiński gasket

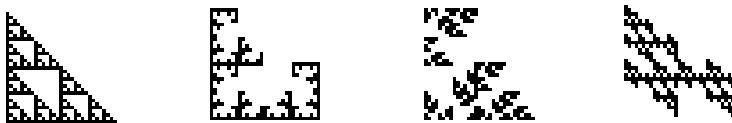


Fig. 4. The Sierpiński gasket and 3 relatives

One source for skepticism of the use of fractal geometry is the heavy reliance on the fractal dimension in the literature to date because it does not fully characterize a given fractal. Thus there is strong motivation to find new ways to classify and characterize fractals. In the case of the Sierpiński gasket and its relatives, the topological properties of the fractals are not constant: the homologies differ. In general, homology theory attempts to distinguish between spaces by constructing algebraic and numerical invariants that are related to the connectivity of the space. The most basic homology theory is simplicial homology, which is based on the triangulations of spaces, see [9]. Given a triangulation of a space, the homology groups can be calculated using an algorithm based on linear algebra, which in general has rather poor numerical behavior. However, in many applications it is the rank of the homology group that is needed, not the entire group structure. This rank is represented by the Betti number and is more easily computable.

The problem with many fractals is that they require infinitely many simplices in their triangulations, and thus at least one of their non-trivial homology groups would be infinite. This is not possible with simplicial homology, so something else needs to be

done for spaces like fractals which have infinitely detailed structure. Instead of applying homology theory to the original object  $X$ , one can apply it to derived spaces (arising from some kind of embedding of the original space) to obtain new information.

This idea has recently been used by Carlsson *et al.* [1], [17], [2]. Their research involves a “study of shape description using a marriage of geometric and topological techniques” [2]. Their derived spaces are constructed using tangential information about the underlying space  $X$  as a subset of  $\mathbb{R}^n$ . First, they have defined the *tangent complex* as the closure of the space of all tangents to all points of  $X$ . Homology of the tangent complex can be used to detect sharp features such as edges and corners. To distinguish the soft features, one needs the so-called *filtered tangent complex*. An invariant called a *persistence module* is obtained by applying homology to this filtered tangent complex. The filtered tangent complex can be used to distinguish between a circle and an ellipse, something that wasn’t possible with the ordinary tangent complex. Persistent homology is used to define a simple shape descriptor, called the *bar-code*. The bar-code is a combinatorial invariant that possesses information about the shape of an object. Ghrist [6] provides a wonderful survey of the current work in this area.

For our purposes, the derived spaces that we will use are the  $\epsilon$ -hulls of the space.

**DEFINITION 1.** The  $\epsilon$ -hull of  $A$  (or the closed  $\epsilon$ -neighborhood of  $A$ ) is defined as:

$$(5) \quad (A)_\epsilon := \{x \in \mathbb{R}^2 \mid d(x, A) \leq \epsilon\},$$

where  $d(x, y)$  is the usual Euclidean metric and  $d(x, A)$  is the infimum of  $d(x, a)$  as  $a$  ranges through  $A$ .

We assume that  $\epsilon \geq 0$ . Of course  $(A)_0 = A$ . If  $A$  is connected and compact, then the  $\epsilon$ -hulls are connected, compact subsets of  $\mathbb{R}^2$ , so determining the homology reduces to counting holes. The number of holes is equal to the rank of the first homology group (the first Betti number).

**1.1. Examples of  $\epsilon$ -hulls.** To illustrate what the analysis of  $\epsilon$ -hulls of a given set  $A$  can tell us, we consider a few preliminary examples. First we consider a circle. Let  $\Gamma$  be a circle with radius  $r$ , see Figure 5(a). Then for any  $\epsilon \in (0, r)$ ,  $(\Gamma)_\epsilon$  is an annulus. For  $\epsilon \in [r, \infty)$ ,  $(\Gamma)_\epsilon$  is a disk. Note that for  $\epsilon \in [0, r)$ , the hole of  $(\Gamma)_\epsilon$  naturally arises from the hole that is in the underlying set, so there does not seem to be any new information.

Second, consider a figure that is a circle with an arc missing, see Figure 5(b). The length between the endpoints of the gap is given by  $a$ . For  $\epsilon \in [r, \infty)$ , there are no holes. For  $\epsilon \in [a/2, r)$ , there is one hole. Finally, for  $\epsilon \in [0, a/2)$  there are no holes. This example demonstrates that it is possible for the underlying set to be simply-connected while there are values of  $\epsilon$  for which the  $\epsilon$ -hull is multiply-connected.

Now consider a slightly more interesting set, see Figure 6. We assume that

$$a_1/2 < a_2/2 < r_1 < r_2$$

Then for  $\epsilon \in [r_2, \infty)$ , there are no holes. For  $\epsilon \in [r_1, r_2)$ , there is 1 hole. For  $\epsilon \in [a_2/2, r_1)$ , there are 2 holes. For  $\epsilon \in [a_1/2, a_2/2)$ , there is 1 hole. Finally for  $\epsilon \in [0, a_1/2)$ , there are no holes. There are five different intervals of  $\epsilon$ -values for which the number and location of holes remains constant for every  $\epsilon$  in the interval. In terms of the holes, there are clearly

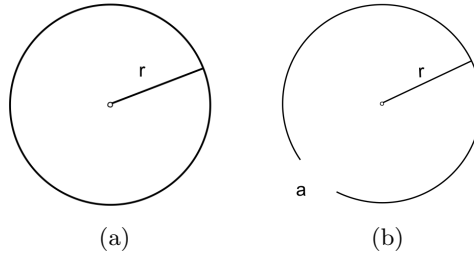


Fig. 5. (a) Circle and (b) large arc of circle

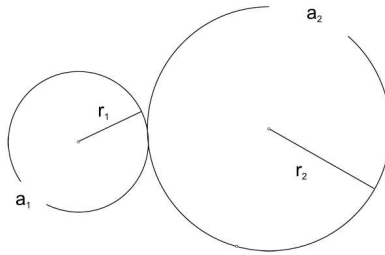


Fig. 6. Slightly more complicated figure

two different regions where there could be a hole. For example, there is a hole within the region bounded by the smaller arc with radius  $r_1$  for values of  $\epsilon$  in  $[a_1/2, r_1)$ . The value  $a_1/2$  is a minimum value for a hole to be in this region, and it is what will be referred to as a ‘contact’ value. The value  $r_1$  is the upper limit for  $\epsilon$ -values that yield a hole in this region, and it is what will be referred to as a ‘collapse’ value.

In the previous examples, we observe that there are intervals of values of  $\epsilon$  for which the hole structure of the  $\epsilon$ -hulls remains constant. For this reason, we are motivated to define a relation on the  $\epsilon$ -values. Given a connected, compact set  $A$ , we define a relation  $\sim_A$  on  $\epsilon$ -values as follows. We say that  $\epsilon_1 \sim_A \epsilon_2$  if for every  $\epsilon \in [\epsilon_{\min}, \epsilon_{\max}]$  (where  $\epsilon_{\min} = \min\{\epsilon_1, \epsilon_2\}$  and  $\epsilon_{\max} = \max\{\epsilon_1, \epsilon_2\}$ ), the hole structure of  $(A)_\epsilon$  is the same. By hole structure, we mean number and location of holes. This relation is an equivalence relation and partitions  $[0, \infty)$  into equivalence classes. Let  $N([\epsilon])$  denote the number of holes in  $(A)_{\epsilon'}$  for any  $\epsilon' \in [\epsilon]$ .

For a final preliminary example we consider the Sierpiński gasket. To completely determine the  $\epsilon$ -hulls, it suffices to consider the triangle with vertices  $(0, 0)$ ,  $(0, 1)$  and  $(1, 0)$ , shown in Figure 7(a).

The radius of the incircle is  $(\sqrt{2} - 1)/2$ . Let  $r_0$  be the radius of the incircle of the biggest hole of the Sierpiński gasket. By the scaling nature of the gasket,  $r_0 = (\sqrt{2} - 1)/4$ . Then the equivalence classes are  $[r_0/(2^k)]$  for  $k \geq 0$ , and

$$\begin{aligned} N([r_0]) &= 0, \\ N([r_0/2]) &= 1, \end{aligned}$$

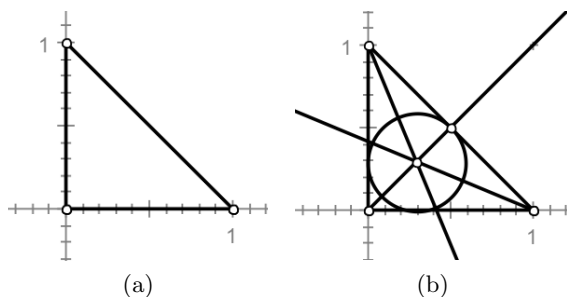


Fig. 7. (a) Triangle similar to the holes of the Sierpiński gasket and (b) the incircle of this triangle

$$N([r_0/4]) = 1 + 3 = 4,$$

$$N([r_0/(2^k)]) = 1 + 3 + \cdots + 3^{k-1} = \frac{3^k - 1}{2}.$$

It is straightforward to show that

$$\lim_{k \rightarrow \infty} \frac{\ln N([r_0/(2^k)])}{\ln 2^k} = \frac{\ln 3}{\ln 2}.$$

So the growth rate of the holes is just the similarity dimension.

At this point we may wonder if it is possible for  $\epsilon$ -hulls to give us any new information about fractals beyond the fractal dimension. We shall see that for the class of non-overlapping symmetric binary fractal trees the  $\epsilon$ -hulls do indeed provide new ways to classify and characterize the trees.

**1.2. Symmetric binary fractal trees.** Fractal trees were first introduced by Mandelbrot in “The Fractal Geometry of Nature” [8]. In general, fractal trees are compact connected subsets of  $\mathbb{R}^n$  (for some  $n \geq 0$ ) that exhibit some kind of branching pattern at arbitrary levels. The class of symmetric binary fractal trees were more recently studied by Mandelbrot and Frame [10] and the author [15], [16]. A symmetric binary fractal tree  $T(r, \theta)$  is defined by two parameters, the scaling ratio  $r$  (a real number between 0 and 1) and the branching angle  $\theta$  (an angle between  $0^\circ$  and  $180^\circ$ ). The trunk splits into two branches, one on the left and one on the right. Both branches have length equal to  $r$  times the length of the trunk and form an angle of  $\theta$  with the affine hull of the trunk. Each of these branches splits into two more branches following the same rule, and the branching is continued *ad infinitum* to obtain the fractal tree.

A *self-avoiding* tree has no self-intersection (see Figure 8). A *self-contacting* tree has self-intersection but no actual branch crossings (see Figure 9). For a given branching angle  $\theta$ , there is a unique scaling ratio  $r_{sc}(\theta)$  (or just  $r_{sc}$ ) such that the corresponding tree is self-contacting [10]. The values of  $r_{sc}$  as a function of  $\theta$  have been completely determined [10]. To determine the value of  $r_{sc}$ , find the smallest scaling ratio such that there is part of the tree besides the trunk that is on the linear extension of the trunk [10], [15]. Further details about the general class of self-contacting trees is available in other literature [10], [15], [16]. A *non-overlapping* tree is either self-avoiding or self-contacting.

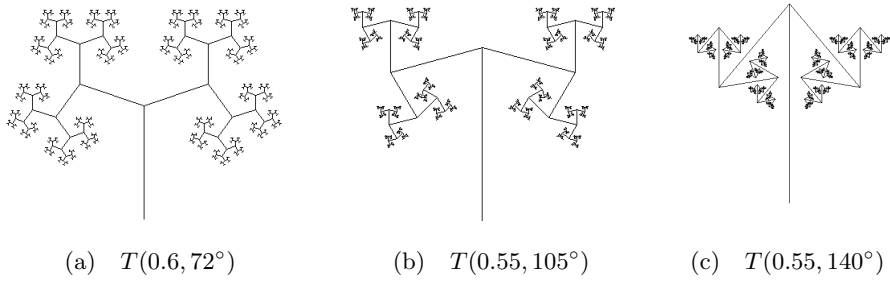


Fig. 8. Self-avoiding trees

A *self-overlapping* tree has branch crossings. This paper focuses on the non-overlapping trees.

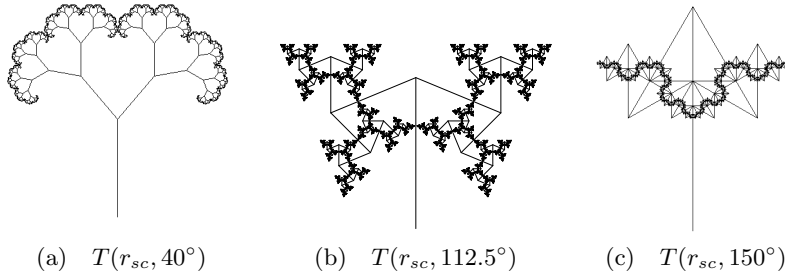


Fig. 9. Self-contacting trees

**1.3. Topology of non-overlapping symmetric binary fractal trees.** Mandelbrot and Frame classified all symmetric binary self-contacting fractal trees according to the topological type of the canopy of the tree. The two branching angles  $90^\circ$  and  $135^\circ$  are identified as being topologically critical [10]. Mandelbrot and Frame restrict their attention to the self-contacting trees, while our work studies both the self-avoiding and self-contacting trees. The self-avoiding symmetric binary fractal trees are all simply-connected and they are all topologically equivalent. On the other hand, the self-contacting trees are infinitely complicated in the sense that they have infinitely many holes (with the two exceptions  $T(r_{sc}, 90^\circ)$  and  $T(r_{sc}, 135^\circ)$ ). The homeomorphism classes of non-overlapping binary trees have been completely determined [15], [16]. The homology type of non-space-filling self-contacting trees is too complicated though.

It seems that from a topological point of view, symmetric binary fractal trees are either too trivial or too complex for topological invariants to describe them. For this reason, we are motivated to look at derived spaces, namely the  $\epsilon$ -hulls. The use of  $\epsilon$ -hulls allows us to give a finer, more interesting classification of symmetric binary fractal trees than the straight topological one given by Mandelbrot and Frame [10]. The self-avoiding fractal trees, which by themselves are all topologically equivalent, can now be associated with systems of  $\epsilon$ -hulls that possess vastly different topological properties, depending on

the branching angle and scaling ratio of the original tree. It is possible for an  $\epsilon$ -hull of a self-avoiding tree with a specific scaling ratio and branching angle to have non-trivial homology.

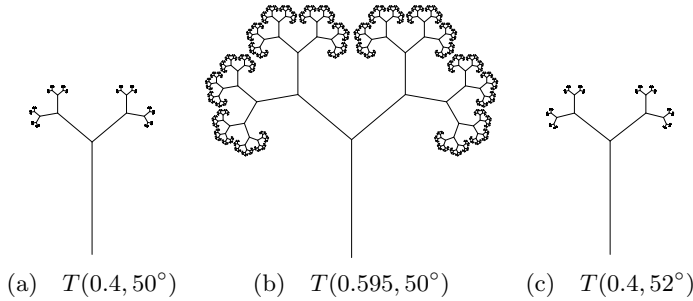


Fig. 10. Comparison of self-avoiding trees

Consider the images of three self-avoiding trees in Figure 10. The first two trees have the same branching angle but different scaling ratio, while the first and third trees have the same scaling ratio and slightly different branching angle. The three trees in Figure 10 each have trivial homology since they are contractible, and they are all topologically equivalent. We will develop a kind of classification of the trees in which the first and third would be of the same class, while the second tree is in a different class.

We will further refine our study of the homology of  $\epsilon$ -hulls by looking at different features of the holes. Various properties of the holes of the  $\epsilon$ -hulls offer different ways to characterize the trees. We first introduce the notion of a hole class, and define the persistence interval and persistence of a hole class. We define the important concept of level of a hole. We are able to define a ‘topological bar-code’ for a tree based on how the hole structure of the  $\epsilon$ -hulls change as a function of  $\epsilon$ .

It is important to remember that many of the constructions developed to analyze the computational topology of symmetric binary fractal trees will extend not only to more general fractal trees, but to other classes of fractals as well. With this thought in mind, many of the definitions in the following sections have been developed to be as general as possible.

**1.4.  $\epsilon$ -Hulls and fractals.** The  $\epsilon$ -hulls have recently been used by other researchers to study fractals. The main example that we reference is the research of Robins *et al.* [12], [13], [14]. Their research presents a study of the extrapolation of topological information about the structure of a space from a finite set of data points. They assume that the underlying set,  $X$ , is a compact metric space, and the data,  $S$ , are a finite set of points that approximate  $X$ . A finite set of points has trivial topological structure. The basic approach is to determine the topological properties of the  $\epsilon$ -hulls of  $S$  as  $\epsilon \rightarrow 0$ , and to extrapolate this information to investigate the connectivity and homology of the underlying set  $X$ .

Previous work by Robins *et al.* has focused on holes in the  $\epsilon$ -hulls that correspond to a hole in the underlying space. The problem is to identify which holes in the  $\epsilon$ -hulls do correspond to such holes. Persistent Betti numbers count the number of holes that



persist in the epsilon-hull for a certain range of epsilon values [12], [13]. When  $X$  has fractal structure, it is possible to see unbounded growth in the persistent Betti numbers  $\beta_k^0(\overline{X_\epsilon})$  (defined in [12]) as  $\epsilon \rightarrow 0$ , so one can characterize this growth by assuming an asymptotic power law. Not surprisingly, if the underlying set is a self-similar fractal, there is a connection between the similarity dimension and the number of holes as  $\epsilon \rightarrow 0$ . If  $\beta_k^0(\overline{X_\epsilon}) \rightarrow \infty$  as  $\epsilon \rightarrow 0$ , then it is natural to quantify the divergence by assuming an asymptotic power law:  $\beta_k^0(\overline{X_\epsilon}) \sim \epsilon^{\gamma_k}$ . Robins gives the following conjecture: If  $X$  is a self-similar fractal and a growth rate  $\gamma_k$  is non-zero, then it is equal to the similarity dimension.

Our own work supports this conjecture. We prove the conjecture for holes in the  $\epsilon$ -hulls of any tree. Specifically, we prove that for a non-simple tree with scaling ratio  $r$  (for which there exist  $\epsilon$ -hulls that are multiply-connected) and a sequence  $\{\epsilon_n\}$  defined by  $\epsilon_n = r^n \epsilon_0$  such that  $\epsilon_0 > 0$  and there are a finite number of holes in the  $\epsilon$ -hulls for all  $\epsilon_n$ , the growth rate of holes is given by the similarity dimension (Theorem 16 in Section 5).

Our basic assumptions are different from the work discussed above [12], [13]. Though the  $\epsilon$ -hulls do provide the basis for our study of fractals, we consider the  $\epsilon$ -hull of the actual fractal, not of a finite approximation to the fractal. This approach has been quite fruitful for the class of symmetric binary fractal trees, and future work will include a thorough study of the  $\epsilon$ -hulls of the relatives of the Sierpiński gasket. Contrary to the earlier work [12], [13], we are interested in all holes that arise in an  $\epsilon$ -hull, not just ones that are due to holes in the underlying fractal. Our goal is not just to classify the topology of the underlying fractal, but to use the  $\epsilon$ -hulls to gain finer invariants to make comparisons. We are also interested in persistence as another characteristic of the holes. In this case, persistence says something about the size of the holes, and it can be considered a topological way to describe ‘lacunarity’ (discussed further below). Persistence of holes also reflects how ‘space-filling’ the holes are, and this characteristic may be particularly useful for applications. Although the utilization of  $\epsilon$ -hulls is not new, our goals and methods are quite different from those used in the literature to date.

As mentioned at the beginning of this Introduction, it is possible for two fractals to have the same fractal dimensions but different topologies. One attempt to characterize this has been through ‘lacunarity’. Mandelbrot first used the term lacunarity to describe the texture of a fractal in terms of the holes or gaps of the fractal [8]. Generally, a fractal with high lacunarity has larger holes or gaps compared to a fractal with the same dimension but lower lacunarity. Mandelbrot suggested several ways to measure lacunarity. One method involves  $\epsilon$ -hulls so it is worthwhile to discuss here.

If  $K$  is a compact subset of  $\mathbb{R}^d$ , let  $|K(\epsilon)|$  denote the  $d$ -dimensional Lebesgue measure (length, area, volume, *etc.*) of the  $\epsilon$ -hull of  $K$ . Let  $D$  be the box-counting or Minkowski dimension of  $K$ . Then [5]

$$(6) \quad \lim_{\epsilon \rightarrow 0} \frac{\log |K(\epsilon)|}{\log \epsilon} = d - D.$$

In many cases we have the stronger condition that [8]

$$|K(\epsilon)| \sim L \epsilon^{d-D},$$

where  $L$  is a ‘prefactor’. Mandelbrot has suggested that  $1/L$  is a numerical measure of lacunarity [8].

For strictly self-similar sets, there are stronger results about lacunarity. Gatzouras presents a theorem that is a refinement to Eq. (6).

**THEOREM** (Gatzouras [5]). *Let  $\phi_1, \dots, \phi_N : \mathbb{R}^d \rightarrow \mathbb{R}^d$  be contracting similarities with similarity ratios  $r_1, \dots, r_N \in (0, 1)$ , and let  $K$  be the corresponding invariant set. Assume that the open set condition is met and that  $\sum_{i=1}^N r_i^D = 1$ .*

1. *If the numbers  $\log r_1, \dots, \log r_N$  do not concentrate on an additive subgroup of  $\mathbb{R}$  of the form  $\lambda\mathbb{Z} = \{\dots, -2\lambda, -\lambda, 0, \lambda, 2\lambda, \dots\}$  for some  $\lambda > 0$ , then the following limit exists and is finite:*

$$\lim_{\epsilon \rightarrow 0} \frac{|K(\epsilon)|}{\epsilon^{d-D}}.$$

2. *The following limit always exists and is finite:*

$$L := \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \frac{|K(e^{-t})|}{e^{-t(d-D)}} dt.$$

The number  $L$  is always defined, and  $1/L$  can be used as a numerical measure of lacunarity. If the condition for (1) is met, then this limit necessarily equals  $L$ . Note that the Sierpiński gasket and its relatives do not satisfy (1). The general idea of the theorem is that for values of  $t$  that are small, the gaps or holes of  $K$  are filled. As  $t$  increases, the holes become exposed and the Lebesgue measure drops. It is important to note that using lacunarity to compare two sets is only meaningful if the sets have the same Hausdorff dimension [4].

We shall see that our approach using  $\epsilon$ -hulls is broader than this notion of lacunarity. First of all, our analysis is on fractals that are not strictly self-similar. Secondly, the analysis considers all values of  $\epsilon$ , not just what happens in the limit as  $\epsilon$  goes to 0. Finally, and most importantly, our analysis will be useful for comparing fractals with different dimensions.

The organization of the remainder of the paper is as follows. Section 2 presents terminology and mathematical details of the symmetric binary fractal trees and their  $\epsilon$ -hulls. Section 3 presents deeper theory regarding the  $\epsilon$ -hulls of trees, including a discussion on hole classes and hole persistence, leading to the definition of the hole sequence of a tree, which provides a ‘topological bar-code’. To illustrate the utility of the new theory, we look at specific examples of trees and their  $\epsilon$ -hulls in Section 4. Following these examples, we are able to further discuss the theory in Section 5. This includes a proof that for non-simple trees the growth rate of holes is equal to the similarity dimension. Finally in Section 6 there is a brief discussion of some current and future work related this paper.

**2.  $\epsilon$ -Hulls of symmetric binary fractal trees.** Before proceeding with the analysis of the  $\epsilon$ -hulls of non-overlapping symmetric binary fractal trees, we first describe the fractal trees and their scaling properties more precisely by using a monoid representation. For any  $r \in (0, 1)$  and any  $\theta \in (0^\circ, 180^\circ)$ , the *generator maps*  $m_R(r, \theta)$  (or just  $m_R$ ) and

$m_L(r, \theta)$  (or just  $m_L$ ) are transformations on  $\mathbb{R}^2$ :

$$(7) \quad m_R \left( \begin{bmatrix} x \\ y \end{bmatrix} \right) = r \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

$$(8) \quad m_L \left( \begin{bmatrix} x \\ y \end{bmatrix} \right) = r \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

The generator maps  $m_R$  and  $m_L$  each consist of a rotation, contraction and translation. The image of any compact set  $U$  under a generator map is similar to the set  $U$  with contraction factor  $r$ .

An *address*  $\mathbf{A} = A_1 A_2 \cdots$  is a string (finite or infinite) of elements, with each element either  $R$  (for ‘right’) or  $L$  (for ‘left’). If an address is finite, then the *level* of the address is equal to the number of elements in the string. The only level 0 address is the empty address  $\mathbf{A}_0$ . An address map  $m_{\mathbf{A}}$  is a composition (finite or infinite) of generator maps. For finite addresses  $\mathbf{A} = A_1 A_2 \cdots A_k$ ,  $m_{\mathbf{A}} = m_{A_k} \circ \cdots \circ m_{A_2} \circ m_{A_1}$ .  $\mathcal{A}_k$  denotes the set of all level  $k$  addresses. An infinite address map is the limit of finite address maps, and is well defined because  $r \in (0, 1)$ .  $\mathcal{A}_{\infty}$  denotes the set of all infinite addresses. If  $C$  is any compact subset of  $\mathbb{R}^2$  and if  $m_{\mathbf{A}}$  is a level  $k$  address map, then the set  $m_{\mathbf{A}}(C)$  is a compact subset of  $\mathbb{R}^2$  that is similar to  $C$  with contraction factor  $r^k$ . The monoid  $M_{LR}$  consists of all address maps with composition as the binary operation.

The *trunk*  $T_0$  of any tree is the closed vertical line segment between the points  $(0, 0)$  and  $(0, 1)$ . The image  $m_R(T_0)$  is the first branch to the right of the trunk, *i.e.*, the closed line segment between the points  $(0, 1)$  and  $(r \sin \theta, r \cos \theta + 1)$ , and the image  $m_L(T_0)$  is the first branch on the left side. Given a level  $k$  address  $\mathbf{A}$ , the *level  $k$  branch*  $\mathbf{b} = b(\mathbf{A})$  is  $m_{\mathbf{A}}(T_0)$ . The *affine hull* of a branch is the line in  $\mathbb{R}^2$  that the branch is collinear with (also referred to as the *linear extension*). A *level  $k$  finite approximation tree*  $T_k(r, \theta)$  is the union of all branches up to and including level  $k$ . With the Hausdorff distance as metric on the space of compact subsets of  $\mathbb{R}^2$ , the sequence of compact sets  $\{T_k(r, \theta)\}$  has a limit as the number of levels of branching goes to infinity because  $r \in (0, 1)$ . The *symmetric binary fractal tree*  $T(r, \theta)$  is defined to be this limit.

Given an address  $\mathbf{A}$ , the *point with address  $\mathbf{A}$*  is denoted by  $P_{\mathbf{A}}$  and defined as follows. If  $\mathbf{A}$  is finite,  $P_{\mathbf{A}}$  is  $m_{\mathbf{A}}((0, 1))$ , *i.e.*, the endpoint of the branch  $b(\mathbf{A})$ , and is referred to as a *vertex* of the tree. A *tip point* of the tree  $T(r, \theta)$  is  $P_{\mathbf{A}}$  for some infinite address  $\mathbf{A}$ , and it equals  $m_{\mathbf{A}}((0, 1))$ . That is, for an infinite address  $\mathbf{A}$ ,  $P_{\mathbf{A}}$  is equal to the limit of the points  $P_{\mathbf{A}_i}$  as  $i$  goes to infinity, where  $\mathbf{A}_i$  is the address of length  $i$  that consists of the first  $i$  elements of  $\mathbf{A}$ . The *top points* of a tree are all points of the tree that have maximal  $y$ -value.

In general, given a set  $U$  that is a subset of a tree  $T(r, \theta)$  and an address  $\mathbf{A}$ ,  $m_{\mathbf{A}}(U)$  is also a subset of the tree, and we refer to it as a *descendant* of  $U$ . The level of a descendant is equal to the level of the address map  $\mathbf{A}$ . A particular class of descendants are subtrees. A *level  $k$  subtree* of a tree  $T(r, \theta)$  is defined to be  $m_{\mathbf{A}}(T)$  for some level  $k$  address  $\mathbf{A}$ . The subtree is denoted by  $S_{\mathbf{A}}(r, \theta)$ ,  $S_{\mathbf{A}}$ ,  $S_{\mathbf{b}}$ , or just  $S$  (the branch  $\mathbf{b} = b(\mathbf{A})$  acts as the trunk of the subtree). If  $U$  is a subset of a tree and  $m_{\mathbf{A}}$  is a level  $k$  address map, then  $m_{\mathbf{A}}(U)$  is similar to  $U$  with contraction factor  $r^k$ . In general, we write  $V \sim_r^k U$  if there exists a level  $k$  address map  $\mathbf{A}$  such that  $V = m_{\mathbf{A}}(U)$ .

For an address  $\mathbf{A} = A_1 A_2 \cdots$ , there exists a natural path on the tree that starts with the trunk and goes to  $P_{\mathbf{A}}$ . The path consists of the trunk along with all branches  $b(\mathbf{A}_i)$ , where  $\mathbf{A}_i = A_1 \cdots A_i$ . We denote this path  $p(\mathbf{A})$ , and because it starts at the trunk we consider it to be level 0. A level  $k$  path is the image of a level 0 path under some level  $k$  address map. Certain paths are relevant for various features of a tree, such as for identifying a hole. Paths that start with the trunk and are such that every other branch is vertical (paths which are as ‘straight’ as possible) form a class of paths that are particularly important. This warrants special symbols for the sets of addresses for such paths.

$$(9) \quad \mathcal{AL}_{2k} = \{A_1 A_2 \cdots A_{2k} \mid A_{2i-1} A_{2i} \in \{RL, LR\}, \ 1 \leq i \leq k\} \quad (k \geq 0),$$

$$(10) \quad \mathcal{AL}_{\infty} = \{A_1 A_2 \cdots \mid A_{2i-1} A_{2i} \in \{RL, LR\}, \ \forall i\}.$$

For example, the address  $RLRLRLRL$  is in  $\mathcal{AL}_8$ . The address  $(RLRL)^{\infty}$  is in  $\mathcal{AL}_{\infty}$ . The ‘ $\mathcal{AL}$ ’ refers to the fact that a level 0 path given by any such address is such that the path alternates between vertical branches (all even level branches) and branches which form an angle of  $\theta$  with the  $y$ -axis (all odd level branches).

From the definition of the generator maps (Eqs. (4) and (8)), it is straightforward to show that  $m_R$  is equal to  $m_L$  composed with reflection across the  $y$ -axis and vice versa. Consequently, each tree is equal to its reflection across the  $y$ -axis. Because of this left-right symmetry, we often restrict our attention to the right side of the tree. The *mirror image* of an address  $\mathbf{A}$  is the address  $\mathbf{A}^*$  that is obtained by switching each element of  $\mathbf{A}$  from  $L$  to  $R$  and  $R$  to  $L$ . For example, if  $\mathbf{A} = RRLR$ , then  $\mathbf{A}^* = LLRL$ . The mirror image of any object of the tree is the reflection of the object across the  $y$ -axis. It is often useful to employ the following concise notation.

$$(11) \quad \mathbf{y} = \{(0, y) \in \mathbb{R}^2 \mid y \in \mathbb{R}\}, \text{ i.e. the } y\text{-axis},$$

$$(12) \quad \mathbf{y}_I = \{(0, y) \in \mathbb{R}^2 \mid y \in I\}, \text{ where } I \text{ is any subset of } \mathbb{R}.$$

All self-contacting trees besides  $T(r_{sc}, 90^\circ)$  and  $T(r_{sc}, 135^\circ)$  have infinitely many holes. The boundary of a hole is a simple closed curve. The holes and their boundary curves were discussed in greater detail in [16]. For our purposes, it suffices to state that such a self-contacting tree has at least one hole whose boundary intersects  $\mathbf{y}$ . Any hole whose boundary intersects  $\mathbf{y}$  is considered to be level 0. Any hole  $H$  whose boundary does not intersect  $\mathbf{y}$  is such that there is a level 0 hole  $H'$  and some address map  $m_{\mathbf{A}}$  such that  $H = m_{\mathbf{A}}(H')$ . Thus the level of a hole or a curve is the level of the corresponding address map.

Certain points of a tree will be important for studying holes. We introduce notation for two particular addresses that are in  $\mathcal{AL}_{\infty}$ :

$$(13) \quad \mathbf{C}_R = RL(LR)^{\infty} \quad \text{and} \quad \mathbf{C}_L = LR(RL)^{\infty}.$$

We now describe the canopy points of a tree, which correspond to endpoints of gaps of trees whose top tip points are Cantor sets [12]. The *degree 0 top canopy interval*, denoted by  $I_{tc}$ , is the closed line segment bounded by the point  $P_{\mathbf{C}_R}$  with address  $\mathbf{C}_R = RL(LR)^{\infty}$  and its mirror image  $P_{\mathbf{C}_L}$  with address  $\mathbf{C}_L = LR(RL)^{\infty}$  (see Figure 11). The point  $P_{\mathbf{C}_R}$  is the *right endpoint* of the canopy interval, and  $P_{\mathbf{C}_L}$  is the *left endpoint*. There are no

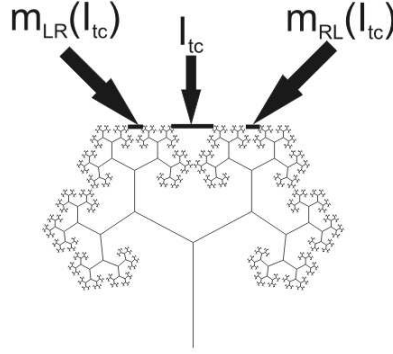


Fig. 11. Level 0, degrees 0 and 1 canopy intervals of  $T(r_{sc}, 62^\circ)$

top points between  $P_{C_R}$  and  $P_{C_L}$ . The interval  $I_{tc}$  intersects the tree only in the points  $P_{C_R}$  and  $P_{C_L}$ , so it indicates a gap in the top of the tree. The image  $m_{\mathbf{A}}(I_{tc})$  is a closed line segment whose endpoints are with addresses  $\mathbf{A}C_R$  and  $\mathbf{A}C_L$ , and the interior of the line segment is disjoint from the subtree (and the entire tree). A *degree  $k$  canopy interval* is the image of  $I_{tc}$  under an address map of the form  $m_{\mathbf{A}}$ , where  $\mathbf{A} \in \mathcal{A}_{2k}$ . Figure 11 displays the two degree one (level 0) canopy intervals. The *top canopy intervals* are the collection of all top canopy intervals of degree  $k$ , as the degree ranges over  $k \geq 0$ . The *top canopy points* are the collection of endpoints of top canopy intervals. The degree of a top canopy point is the degree of the interval that it is an endpoint of. Right canopy points of degree  $k$  are at addresses  $\mathbf{A}C_R$  for some  $\mathbf{A} \in \mathcal{A}_{2k}$ , and left canopy points of degree  $k$  are at addresses of the form  $\mathbf{A}C_L$  for some  $\mathbf{A} \in \mathcal{A}_{2k}$ . In general, a *canopy point* is a top canopy point, or the image of a top canopy point under  $m_{\mathbf{A}}$  on some subtree  $S_{\mathbf{A}}$ . Likewise, a *canopy interval* on a subtree  $S_{\mathbf{A}}$  is the image of a top canopy interval under  $m_{\mathbf{A}}$ .

### 2.1. $\epsilon$ -Hulls of trees

**DEFINITION 2.** Let  $\theta \in (0^\circ, 180^\circ)$  and  $r \in (0, 1)$  be given, and let  $\epsilon \geq 0$ . We denote the  $\epsilon$ -hull of the tree with scaling ratio  $r$  and branching angle  $\theta$  by  $E(r, \theta, \epsilon)$ , and it is equal to  $(T(r, \theta))_\epsilon$ :

$$(14) \quad E(r, \theta, \epsilon) = \{(x, y) \in \mathbb{R}^2 \mid d((x, y), T(r, \theta)) \leq \epsilon\}.$$

The boundary of  $E(r, \theta, \epsilon)$  is denoted by  $\partial E(r, \theta, \epsilon)$  or just  $\partial E$ . Thus

$$(15) \quad \partial E = \{(x, y) \in \mathbb{R}^2 \mid d((x, y), T(r, \theta)) = \epsilon\}.$$

The new notation for the  $\epsilon$ -hulls of trees is used to reflect the fact that the topological and geometrical properties of a given  $\epsilon$ -hull are a function of  $r$ ,  $\theta$  and  $\epsilon$ .

**REMARK.** For any  $r$  and  $\theta$ ,  $E(r, \theta, 0) = \partial E(r, \theta, 0) = T(r, \theta)$ . For any  $\epsilon \geq 0$ ,  $T(r, \theta) \subseteq E(r, \theta, \epsilon)$ .

**DEFINITION 3.** Let  $r \in (0, 1)$  and  $\theta \in (0^\circ, 180^\circ)$  be given. Let  $\epsilon \in [0, \infty]$ . Let  $S = S_{\mathbf{A}}(r, \theta)$  be a subtree of  $T(r, \theta)$ , where  $\mathbf{A} \in \mathcal{A}_k$  for some  $k \geq 1$ . Let  $\mathbf{b} = b(\mathbf{A})$ . We define the  $\epsilon$ -hull

of the subtree  $S$  with scaling ratio  $r$ , branching angle  $\theta$  and trunk  $b(\mathbf{A})$ , denoted by  $E_{\mathbf{A}}(r, \theta, \epsilon)$ ,  $E_S(\epsilon)$ ,  $E_S$ , or  $E_{\mathbf{b}}$ , to be

$$(16) \quad E_{\mathbf{A}}(r, \theta, \epsilon) = \{(x, y) \in \mathbb{R}^2 \mid d((x, y), S_{\mathbf{A}}(r, \theta)) \leq \epsilon\}.$$

PROPOSITION 1. *Let  $r, \theta$  be given. Let  $S = S_{\mathbf{A}}(r, \theta) \in \mathcal{S}_k$ , where  $\mathbf{A} \in \mathcal{A}_k$ , for some  $k \geq 1$ . Then for any  $\epsilon \in [0, \infty)$ ,  $E_{\mathbf{A}}(r, \theta, r^k \epsilon) = m_{\mathbf{A}} E(r, \theta, \epsilon)$ . Consequently  $E_{\mathbf{A}}(r, \theta, r^k \epsilon) \sim_r^k E(r, \theta, \epsilon)$  (i.e., similar with factor  $r^k$  via  $m_{\mathbf{A}}$ ).*

Intuitively, for a given tree  $T(r, \theta)$  and a given  $\epsilon \geq 0$ , one considers a hole in  $E(r, \theta, \epsilon)$  to be some open, contractible (i.e. simply-connected) set that is disjoint from the  $\epsilon$ -hull, and such that the boundary of the hole is a subset of the boundary of the  $\epsilon$ -hull. For a given tree  $T(r, \theta)$  with  $r \in (0, 1)$  and  $\epsilon \geq 0$ , let  $E^C(r, \theta, \epsilon)$ , or  $E^C$ , denote the complement of the  $\epsilon$ -hull in  $\mathbb{R}^2$ .

$$(17) \quad E^C(r, \theta, \epsilon) = \mathbb{R}^2 \setminus E(r, \theta, \epsilon) = \{(x, y) \in \mathbb{R}^2 \mid d((x, y), T(r, \theta)) > \epsilon\}.$$

In general, we use a superscript of ‘ $C$ ’ to denote the complement of a given set in  $\mathbb{R}^2$ .

For any  $r, \theta$ , and  $\epsilon$ ,  $E^C$  is an open set. So if  $E^C$  is non-empty, then the components of  $E^C$  are all open. There is one unbounded component whose first homology group is  $\mathbb{Z}$ , and any other components are contractible. The holes, if they exist, are the bounded components. They are also the insides of simple, closed curves of the boundary of the  $\epsilon$ -hull.

Let  $\Gamma$  denote the set of the simple, closed curves in  $\mathbb{R}^2$ . For  $\gamma \in \Gamma$ , the inside of a curve is denoted  $O(\gamma)$ . We also define the set of simple, closed curves of tree  $T(r, \theta)$ , denoted  $\Gamma(r, \theta)$  (discussed in greater detail in [16]). A hole of a tree  $T(r_{sc}, \theta)$  is equal to  $O(\gamma)$  for some  $\gamma$  such that  $O(\gamma)$  is disjoint from  $T(r_{sc}, \theta)$ . Now we define the set of simple, closed curves of the boundary of an  $\epsilon$ -hull.

DEFINITION 4. Let  $r \in (0, 1)$  and  $\theta \in (0^\circ, 180^\circ)$  be given. Let  $\epsilon \in [0, \infty]$ . A *simple, closed curve of the boundary of the  $\epsilon$ -hull*  $E(r, \theta, \epsilon)$  is a simple, closed curve  $\gamma$  such that  $\gamma$  is a subset of  $\partial E(r, \theta, \epsilon)$  and  $O(\gamma)$  is disjoint from the  $\epsilon$ -hull  $E(r, \theta, \epsilon)$ , i.e.  $O(\gamma) \subset E^C$ . Let  $\Gamma(r, \theta, \epsilon)$  be the collection of all simple, closed curves of the boundary of the  $\epsilon$ -hull  $E(r, \theta, \epsilon)$ .

DEFINITION 5. Let  $r \in (0, 1)$  and  $\theta \in (0^\circ, 180^\circ)$  be given. Let  $\epsilon \in [0, \infty]$ . A *hole in the closed  $\epsilon$ -hull*  $E(r, \theta, \epsilon)$  is a region of the form  $O(\gamma)$  for some  $\gamma \in \Gamma(r, \theta, \epsilon)$ . Alternately, a set  $H$  is a hole if it is an open, simply-connected subset of  $\mathbb{R}^2$  such that  $H$  is a component of the complement of  $E(r, \theta, \epsilon)$  in  $\mathbb{R}^2$ . Denote the set of all such holes by  $\mathcal{H}(r, \theta, \epsilon)$ , or just  $\mathcal{H}(\epsilon)$ .

Let  $\epsilon \geq 0$ . Let  $N(r, \theta, \epsilon)$  be the number of holes in the  $\epsilon$ -hull  $E(r, \theta, \epsilon)$  of the tree  $T(r, \theta)$ . Often we write  $N(\epsilon)$ . For a given  $\epsilon \geq 0$ ,  $N(r, \theta, \epsilon)$  may be 0, a finite number, or infinite, depending on  $r$  and  $\theta$ . For any tree,  $N(r, \theta, \infty) = 0$ . For any  $\theta$ , if  $r < r_{sc}$ , then  $N(r, \theta, 0) = 0$ . That is, the number of holes in any self-avoiding tree is 0 since there are no simple, closed curves in  $\partial E(r, \theta, 0) = T(r, \theta)$ , because any self-avoiding tree is contractible.

PROPOSITION 2. For any  $\theta$  except  $90^\circ$  or  $135^\circ$ ,  $N(r_{sc}, \theta, 0) = \infty$ . That is, a self-contacting, non-space-filling tree has an infinite number of holes.

*Proof.* There must exist at least one closed, simple curve in the tree  $T(r_{sc}, \theta)$  such that  $O(\gamma)$  is disjoint from  $T(r_{sc}, \theta)$  (see [16] for actual constructions of such curves). Without loss of generality, let  $\gamma \in \Gamma(r, \theta, 0)$  such that  $\gamma$  intersects  $\mathbf{y}$  (such a curve necessarily exists [16]). For each  $k \geq 1$ , the set  $m_{R^k}(\gamma)$  is a simple closed curve, by the similarity of the tree. Let  $\gamma_k$  denote  $m_{R^k}(\gamma)$ . Also by the similarity of the tree, we have  $O(\gamma_k)$  disjoint from  $T(r_{sc}, \theta)$ , so  $O(\gamma_k)$  is a hole of the tree. Each  $O(\gamma_k)$  is distinct, because each  $\gamma_k$  is similar to  $\gamma$  with contraction factor  $r_{sc}^k$ . Therefore the corresponding holes to these simple, closed curves as  $k$  ranges through the positive integers are all distinct, and so there must be an infinite number of distinct holes. ■

DEFINITION 6. A tree  $T(r, \theta)$  is a *simple tree* if  $N(r, \theta, \epsilon) = 0$  for every  $\epsilon \in [0, \infty]$ . Thus every  $\epsilon$ -hull of the tree is contractible (there are no holes).

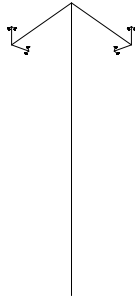


Fig. 12. Simple tree  $T(0.25, 125^\circ)$

Many simple trees do exist [15]. For example, the space-filling trees  $T(r_{sc}, 90^\circ)$  and  $T(r_{sc}, 135^\circ)$  are contractible, and any  $\epsilon$ -hull of either tree is also contractible. These two trees are special cases. In general, the simple trees occur for small scaling ratios. In fact, for any branching angle, there exist scaling ratios such that the corresponding trees are simple (see [15] for a proof). Figure 12 displays a self-avoiding tree that is simple.

DEFINITION 7. A tree  $T(r, \theta)$  is a *non-simple tree* if there exists an  $\epsilon \geq 0$  such that  $N(r, \theta, \epsilon) > 0$ .

To continue the study of non-simple trees, more theory regarding the holes themselves is needed.

**3. Persistence of holes, hole classes and bar-codes.** In this section we present theory regarding the hole structure of  $\epsilon$ -hulls of the trees, leading to the definition of ‘topological bar-codes’ for the trees.

**3.1. Hole classes and persistence.** Consider the self-contacting tree with branching angle  $145^\circ$ , shown in Figure 13(a). The tree itself contains infinitely many holes. For example, there is a hole  $H_1$  (the grey region) whose boundary contains the vertical line segment between the point with address  $RR$  and the top of the trunk, and part of the branch

$b(R)$ . For values of  $\epsilon > 0$  that are sufficiently small, there will still be exactly one hole in this region of the  $\epsilon$ -hull. There will not always be a hole in this region, because for sufficiently large  $\epsilon$ , the region is completely covered by the  $\epsilon$ -hull. This leads us to wonder for what range of values of  $\epsilon$  the holes persist.

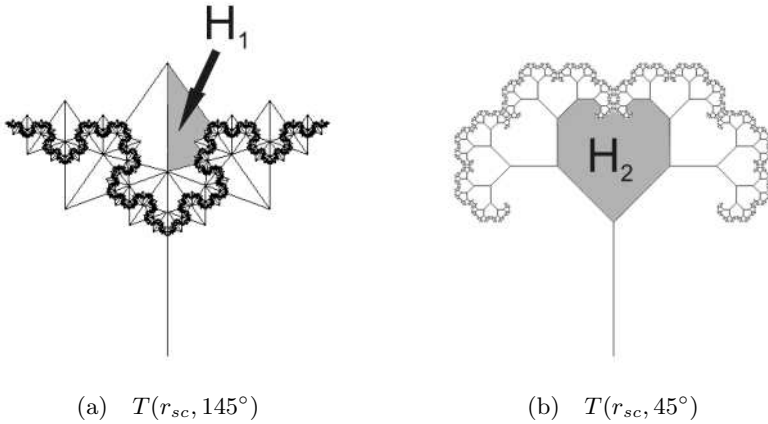


Fig. 13. Self-contacting trees with holes

Consider the tree  $T(r_{sc}, 45^\circ)$  shown in Figure 13(b). There is a hole  $H_2$  bounded in part by the point  $(0, 1)$ , the lowest tip point on the  $y$ -axis (with address  $RL^3(LR)^\infty$ ), and the branches  $b(R)$  and  $b(L)$ . It is possible to determine the smallest  $\epsilon$  for which this hole is covered by the  $\epsilon$ -hull of the tree [15]. For any  $\epsilon$  between 0 and this ‘collapse’ value, there is exactly one hole in the  $\epsilon$ -hull that has non-empty intersection with  $H_2$ .

This is not always the case. Consider the hole  $H_1$  in the tree  $T(r_{sc}, 145^\circ)$  again. Now consider the point with address  $RRRL$ . When  $\epsilon$  is half the  $x$ -coordinate of this point, the original hole splits into two holes. Because of this possibility of a hole splitting, we have to be careful with the definition of the persistence interval. We would like the definition to be such that for any two holes, their corresponding persistence intervals are either the same or they are disjoint, because we do not want one to be a proper subset of the other (as would be the case for one of the two holes that is a result of a larger hole splitting in two).

**DEFINITION 8.** Let  $T(r, \theta)$  be a non-simple tree, let  $\epsilon \geq 0$  be such that there exists a hole  $H$  in  $E(r, \theta, \epsilon)$ . Then the *persistence interval* of  $H$ , denoted by  $P(H)$ , is the maximal interval  $I$  such that for any  $\epsilon_1 \in I$ , there is a unique hole  $H_1$  of  $E(r, \theta, \epsilon_1)$  that has non-empty intersection with  $H$ , and  $H_1$  obeys the following condition. For any  $\epsilon_2 \in I$ , there exists exactly one hole  $H_2$  of  $E(r, \theta, \epsilon_2)$  that has non-empty intersection with  $H_1$ .

**DEFINITION 9.** Let  $T(r, \theta)$  be a non-simple tree, let  $\epsilon_0 \geq 0$  be such that there exists a hole  $H$  of  $E(r, \theta, \epsilon_0)$ . The *hole class*  $[H]$  is the set of holes that have non-empty intersection with  $H$  as  $\epsilon$  ranges through  $P(H)$ .

We often use just the symbol  $H$  to denote the hole class  $[H]$  when the context is clear. For any  $H' \in [H]$ , we have  $P(H') = P(H)$  (as a result of the definition of persistence



interval), so  $P(H)$  is often considered to be the persistence interval of the hole class  $[H]$ , and is also denoted by  $P([H])$ .

DEFINITION 10. The *persistence* of a hole class  $[H]$  is defined to be the length of the persistence interval  $P([H])$ , and is denoted by  $|P([H])|$  or  $|P(H)|$ .

We shall see that it is possible for a hole class to have persistence equal to 0.

DEFINITION 11. Let  $T(r, \theta)$  be a non-simple tree, and let  $H$  be a hole of some  $\epsilon$ -hull of the tree. The *contact value* of the hole class  $[H]$  is the infimum of the persistence interval of the hole class, and it is denoted by  $\underline{\epsilon}_{[H]}$  or just  $\underline{\epsilon}_H$ . The *collapse value* of the hole class  $[H]$  is the supremum of the persistence interval, and it is denoted by  $\overline{\epsilon}_{[H]}$  or just  $\overline{\epsilon}_H$ . Thus

$$(18) \quad \underline{\epsilon}_H = \inf(P([H])), \quad \overline{\epsilon}_H = \sup(P([H])).$$

The collapse value marks the end of a certain hole class. It could mean that the maximal hole of the class is now completely covered by the  $\epsilon$ -hull, but it could also mean that the hole class has split into more than one hole class (so the region is not covered).

DEFINITION 12. Let  $[H]$  be a hole class. The *maximal hole of the hole class*  $[H]$  is denoted  $[H]_{max}$ , and it is the element of  $[H]$  that occurs at  $\underline{\epsilon}_H$ . Every other hole in the hole class is a subset of this hole.

If a hole class has contact value equal to 0, it is considered to be a *self-contacting hole class*, because the maximal hole of the class is a hole of the actual tree. More will be said regarding persistence intervals once we have introduced the concept of the level of a hole. We will present quantitative results about persistence intervals of hole classes for the  $\epsilon$ -hulls of specific trees.

**3.2. Symmetry and levels of holes.** Each symmetric binary fractal tree has the  $y$ -axis, which we have denoted by  $\mathbf{y}$ , as an axis of symmetry. We now investigate how this symmetry affects the structure of the holes. It follows immediately that those holes which intersect  $\mathbf{y}$  are symmetric in  $\mathbf{y}$ , and other holes which do not intersect  $\mathbf{y}$  have a counterpart reflected across  $\mathbf{y}$ . It follows from self-similarity that many holes will be symmetric in lines which form linear extensions of branches on the tree.

Thus it seems plausible that there are holes in some  $\epsilon$ -hulls that are also symmetric, in the sense that a hole is symmetric if there is an axis of symmetry through the hole that is an axis of symmetry for the hole itself.

DEFINITION 13. For given values of  $r$  and  $\theta$  and for  $\epsilon \in [0, \infty)$ , a hole  $H \in \mathcal{H}(r, \theta, \epsilon)$  is a *symmetric hole* if there exists an axis of symmetry through it. Otherwise it is a *non-symmetric hole*.

For example, the hole  $H_2$  of the tree  $T(r_{sc}, 45^\circ)$  described above (see Figure 13(b)) is symmetric about  $\mathbf{y}$ . The hole  $H_4$  of  $T(r_{sc}, 45^\circ)$  described below (see Figure 14(b)) is symmetric about the affine hull of the branch  $b(RRLL)$ . The hole  $H_1$  of the tree  $T(r_{sc}, 145)$  described above (see Figure 13(a)) is not symmetric. However, the mirror image of  $H_1$  is also a hole. The hole  $H_3$  of  $T(r_{sc}, 145)$  described below (see Figure 14(a)) is not symmetric.

For given values of  $r$  and  $\theta$  and for  $\epsilon \in [0, \infty]$  it is clear that both the  $\epsilon$ -hull of the tree and its boundary are symmetric about  $\mathbf{y}$ . Suppose  $\gamma \in \Gamma(r, \theta, \epsilon)$  is such that  $\gamma$  does not intersect  $\mathbf{y}$ . Then the mirror image  $\gamma^*$  ( $\gamma$  reflected across  $\mathbf{y}$ ) is an element of  $\Gamma(r, \theta, \epsilon)$  distinct from  $\gamma$ .

It is possible to have a hole whose boundary is entirely contained in the boundary of the  $\epsilon$ -hull of a subtree. Recall that a level  $k$  subtree is the image of  $T(r, \theta)$  under some level  $k$  address map. This motivates the definition of the ‘level’ of a hole. Before we give the definition, we present some preliminary results about holes and subtrees.

**THEOREM 3 (Subtree Theorem).** *Let  $\theta$  be given, let  $r \leq r_{sc}$ , and let  $\epsilon \geq 0$  be such that  $\mathcal{H}(r, \theta, \epsilon)$  is non-empty. Let  $H \in \mathcal{H}(r, \theta, \epsilon)$ . Then there exists a unique integer  $k \geq 0$  such that:*

1. *For every  $m \leq k$ , there exists a unique level  $m$  subtree  $S$  such that  $\partial H \subset \partial E_S$ .*
2. *For all integers  $l > k$ , there are no level  $l$  subtrees  $S'$  such that  $\partial H \subset \partial E_{S'}$ .*

*Proof.* Suppose  $\theta$  is given,  $r \leq r_{sc}$ , and  $\epsilon \geq 0$  such that  $\mathcal{H}(r, \theta, \epsilon)$  is non-empty. Let  $H \in \mathcal{H}(r, \theta, \epsilon)$ .

If  $\partial H$  was completely included in the boundary of more than one subtree, then these two subtrees must have been overlapping at branch interiors, and this contradicts that  $r \leq r_{sc}$ . So if  $\partial H$  is contained in the boundary of the  $\epsilon$ -hull of some level  $m$  subtree, then this subtree is the unique such subtree of level  $m$ .

Trivially,  $\partial H$  is a subset of the boundary of the  $\epsilon$ -hull of the level 0 subtree (which is just the tree itself). Now for any  $m \geq 0$  such that there exists a unique level  $m$  subtree whose  $\epsilon$ -hull boundary contains  $\partial H$ , then the same is true for all integers  $l$  such that  $0 \leq l \leq m$ . Let  $l$  be such that  $0 \leq l \leq m$ . Then there is a unique level  $l$  subtree  $S'$  that is a superset of  $S$  (since there is a unique level  $l$  branch that is an ancestor of the trunk of the subtree  $S$ ), and hence  $\partial H \subset \partial E_{S'}$ .

Finally, there exists an integer  $m$  such that there are no level  $m$  subtrees whose  $\epsilon$ -hull boundaries entirely contain  $\partial H$ . If  $\epsilon = 0$ , then  $\partial E = T(r, \theta)$ , and  $\partial H$  must partially consist of branches of the tree. Any integer that is higher than the level of a branch that is part of  $\partial H$  could be taken as  $m$ . If  $\epsilon > 0$ , then there exists an integer  $m$  such that the  $\epsilon$ -hull of any subtree of level  $m$  or higher is contractible (since the size of the subtrees is decreasing as the levels increase), and so such an  $m$  would be such that there are no level  $m$  subtrees whose  $\epsilon$ -hull boundaries entirely contain  $\partial H$ .

Thus there must exist a maximal non-negative integer  $k$  such that there exists a level  $k$  subtree  $S$  such that  $\partial H \subset \partial E_S$ , and this level  $k$  subtree is unique and for all integers  $l > k$ , there are no level  $l$  subtrees  $S'$  such that  $\partial H \subset \partial E_{S'}$ . ■

**DEFINITION 14.** Let  $\theta$  and  $r \leq r_{sc}$  be given, let  $\epsilon \geq 0$  be such that  $\mathcal{H}(r, \theta, \epsilon) \neq \emptyset$ . Let  $H \in \mathcal{H}(r, \theta, \epsilon)$ , let  $H' = [H]_{max}$  and  $\epsilon' = \underline{\epsilon_H}$ . Then  $H$  is a *level  $k$  hole* for some  $k \geq 0$ , if  $k$  is the largest integer such that there exists a level  $k$  subtree  $S$ , whose  $\epsilon$ -hull boundary  $\partial E_S(\epsilon')$  contains the boundary  $\partial H'$  of  $H'$ . For a given non-negative integer  $k$ , denote the set of all level  $k$  holes by  $\mathcal{H}_k(r, \theta, \epsilon)$ , or just  $\mathcal{H}_k(\epsilon)$ .

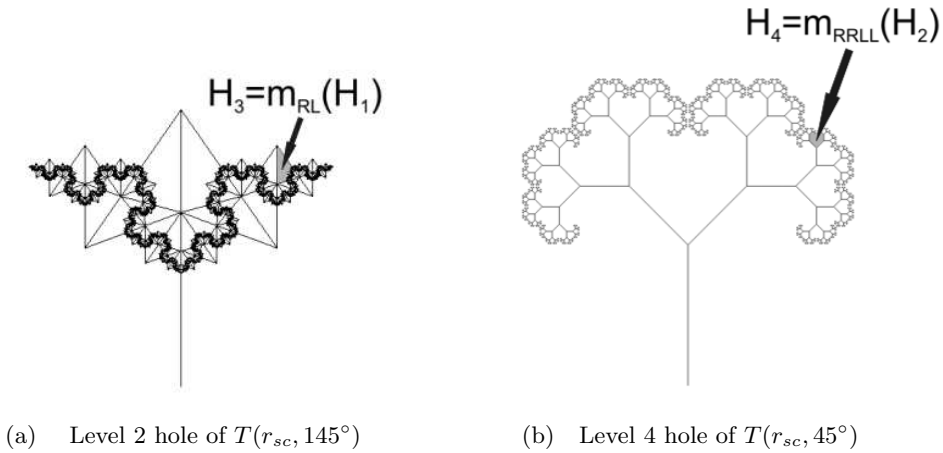


Fig. 14. Examples of holes of different levels

Examples of holes with levels higher than 0 are displayed in Figures 14(a) and (b). Figure 14(a) displays the level 2 hole  $H_3 = m_{RL}(H_1)$  of  $T(r_{sc}, 145^\circ)$ . Figure 14(b) displays the level 4 hole  $m_{RRL}(H_2)$  of  $T(r_{sc}, 45^\circ)$ .

**REMARKS.** An immediate consequence of this definition of hole level is that the elements of any hole class are all at the same level. In addition, because  $\partial E_S(\epsilon')$  contains the boundary  $\partial H'$  of  $H'$  we also have that  $\partial E_S(\epsilon)$  contains the boundary  $\partial H$  of  $H$ . Let  $\theta$  and  $r \leq r_{sc}$  be given, and let  $\epsilon \geq 0$ . Let  $k \geq 0$ . Then  $N_k(r, \theta, \epsilon)$ , or just  $N_k(\epsilon)$ , is the number of level  $k$  holes of the corresponding  $\epsilon$ -hull.

The following theorem will be extremely useful for analyzing the characteristics of hole classes.

**THEOREM 4 (Hole Level Theorem).** *Let  $\theta$  be given, and let  $r \leq r_{sc}$ . Let  $\epsilon \geq 0$  such that  $\mathcal{H}_k(r, \theta, \epsilon)$  is non-empty, for some  $k \geq 1$ . Let  $H \in \mathcal{H}_k(r, \theta, \epsilon)$ . Then for  $\epsilon' = r^{-k}\epsilon$ , there exists a hole  $H' \in \mathcal{H}_0(r, \theta, \epsilon')$  and a level  $k$  address map  $m_{\mathbf{A}}$  such that  $H = m_{\mathbf{A}}(H')$ , and thus  $H \sim_r^k H'$ .*

*Proof.* Let  $\theta$  be given, let  $r \leq r_{sc}$ , and let  $\epsilon \geq 0$ . Let  $H \in \mathcal{H}_k(r, \theta, \epsilon)$  for some  $k \geq 0$ . Let  $\epsilon' = r^{-k}\epsilon$ . There exists a subtree  $S = S_{\mathbf{A}}(r, \theta)$ , for some  $\mathbf{A} \in \mathcal{A}_k$ , such that  $\partial H \subset \partial E_S$  (by definition of a level  $k$  hole). Now  $S$  is a level  $k$  subtree, so  $E_{\mathbf{A}}(r, \theta, \epsilon) \sim_k E(r, \theta, \epsilon')$  since  $E_{\mathbf{A}}(r, \theta, \epsilon) = m_{\mathbf{A}}(E(r, \theta, \epsilon'))$ . Now  $\partial H$  is a simple, closed curve that is a subset of the boundary  $\partial E_S$  of  $E_S$ .  $\partial E_S \sim_r^k \partial E$ , via the address map  $m_{\mathbf{A}}$ , so there exists  $\gamma \in \partial E(r, \theta, \epsilon')$  such that  $m_{\mathbf{A}}(\gamma) = H$  and so  $\partial H \sim_r^k \gamma$ . Let  $H' = O(\gamma) \in \mathcal{H}(r, \theta, \epsilon')$ . Then  $m_{\mathbf{A}}(H') = H$  and  $H \sim_r^k H'$ . Note that the boundary of  $E_S$  may intersect the boundaries of other level  $k$  subtrees, but the intersection can't affect  $\partial H$  because that would contradict  $H$  being a level  $k$  hole. ■

The idea of this theorem is that any hole must correspond to a level 0 hole. For example, the level 2 hole  $H_3 = m_{RL}(H_1)$  in Figure 14(a) clearly corresponds to the level

0 hole  $H_1$  in Figure 13(a). It is always possible to find an address and a level 0 hole to obtain the given hole.

The converse of the previous theorem is also true and is stated below in the Descendant Theorem (Theorem 7). The Hole Level Theorem may seem obvious due to the scaling nature of the trees. However, the  $\epsilon$ -hulls of subtrees may overlap in the formation of the  $\epsilon$ -hull of the tree, so it is more complicated than just considering the  $\epsilon$ -hulls of subtrees as disjoint sets. The Hole Level Theorem tells us that any hole in any  $\epsilon$ -hull is the image of a level 0 hole under a level  $k$  address map, *i.e.*, it is a result of an action of the free monoid. This is a very special property, because this implies that the level 0 holes yield information about holes at *any level*. We can restrict our attention to the level 0 holes, and this is the same thing as taking the fundamental domain under the action of the free monoid. However, it is interesting to see how the critical values and scaling ratios interact.

We now direct our attention to the properties of the level 0 holes, and then derive more general results about higher level holes. Self-contact for a given branching angle  $\theta$  occurs at the smallest scaling ratio such that either a tip point intersects  $\mathbf{y}_{(1, y_{\max}]}$ , or a tip point or branch endpoint intersects the trunk [15]. This implies that  $r_{sc}$  is the smallest scaling ratio such that either  $S_{RL}$  and  $S_{LR}$  intersect or  $S_{RR}$  and  $T_0$  intersect. In terms of holes of the tree  $T(r, \theta) = E(r, \theta, 0)$ , self-contact occurs for the smallest scaling ratio such that there is a level 0 hole in the tree. For  $\epsilon > 0$ , the conditions for a hole to be level 0 are similar.

**THEOREM 5 (Level 0 Hole Criteria).** *Let  $\theta$  be given, and let  $r \leq r_{sc}$ . Let  $\epsilon \geq 0$ . A hole  $H \in \mathcal{H}(r, \theta, \epsilon)$  is level 0 if and only if one of the following conditions holds:*

1. *The boundary of  $H$  has non-empty intersection with both  $E_{RL}(\epsilon)$  and  $E_{LR}(\epsilon)$ .*
2. *The boundary of  $H$  has non-empty intersection with both  $E_{RR}(\epsilon)$  and  $T_0(\epsilon)$ .*
3. *The boundary of  $H$  has non-empty intersection with both  $E_{LL}(\epsilon)$  and  $T_0(\epsilon)$ .*

*Proof.* See [15]. ■

**COROLLARY 6.** *A hole is level 0 if and only if it intersects  $\mathbf{y}_{(1, \infty)}$  (in the first case of Theorem 5) or it does not intersect  $\mathbf{y}$  but needs the trunk for formation (in the second and third cases).*

**THEOREM 7 (Descendant Theorem).** *Let  $\theta$  be given, and let  $r \leq r_{sc}$ . Let  $\epsilon \geq 0$  be such that there exists a level 0 hole  $H \in \mathcal{H}(r, \theta, \epsilon)$ . For any positive integer  $j$ , there exist  $2^j$  corresponding distinct level  $j$  holes in  $\mathcal{H}(r, \theta, r^j \epsilon)$ , namely the holes of the form  $m_{\mathbf{A}}(H)$ , where  $\mathbf{A} \in \mathcal{A}_j$ .*

*Proof.* See [15]. First one proves that  $m_R(H)$  and  $m_L(H)$  are distinct level 1 holes in  $(r, \theta, r\epsilon)$ . Then proceed via induction. ■

Due to the nice scaling properties of holes, we have some immediate results about hole classes. To determine the persistence interval of any hole, it suffices to determine the persistence interval of its corresponding level 0 hole.

**PROPOSITION 8.** *Let  $T(r, \theta)$  be a tree. Let  $H$  be a level 0 hole, and let  $H_{\mathbf{A}}$  be the corresponding level  $k$  hole for some level  $k$  address map  $\mathbf{m}_{\mathbf{A}}$ . Then  $\epsilon \in P([H])$  if and only if  $r^k \epsilon \in P([H_{\mathbf{A}}])$ .*

COROLLARY 9. Let  $T(r, \theta)$  be any tree, and let  $H \in \mathcal{H}_0(r, \theta, \epsilon)$  for some  $\epsilon \geq 0$ . Let  $\mathbf{A} \in \mathcal{A}_k$  for some  $k \geq 1$ . Let  $H_{\mathbf{A}} = m_{\mathbf{A}}(H) \in \mathcal{H}_k(r, \theta, r^k \epsilon)$ . Then

$$(19) \quad \underline{\epsilon}_{H_{\mathbf{A}}} = r^k \underline{\epsilon}_H \quad \text{and} \quad \overline{\epsilon}_{H_{\mathbf{A}}} = r^k \overline{\epsilon}_H.$$

To study the holes in  $\mathcal{H}(r, \theta, \epsilon)$  as  $\epsilon$  ranges through  $[0, \infty)$ , as a result of Theorems 4 and 7 it suffices to determine the critical values for the level 0 holes. Any level  $k$  hole is the image of a level 0 hole under a suitable level  $k$  address map. Conversely, any level 0 hole has  $2^k$  corresponding level  $k$  holes via the address maps. To get the complete picture for the persistent homology, we need to know how the critical values and the scaling ratios interact. Another important factor regarding holes is symmetry. The  $\epsilon$ -hulls all possess left-right symmetry, so it suffices to consider holes that are not disjoint from the right side of the  $y$ -axis. This will suffice to give information about holes on both sides.

**3.4. The hole sequence.** We now investigate one way to characterize a tree, based on what can happen in the  $\epsilon$ -hulls of a specific tree as  $\epsilon$  ranges through  $[0, \infty]$ . First we discuss the critical values of  $\epsilon$  for a tree, that is, for a pair  $r$  and  $\theta$ . Then we will define a relation on the  $\epsilon$ -values, and this will enable us to define a hole partition and sequence. The hole sequence is the new characteristic of a tree that is a kind of ‘topological bar-code’ in the sense of Carlsson *et al.* [2].

Let  $r, \theta$  be such that  $T(r, \theta)$  is not a simple tree. The set  $Con(r, \theta)$  denotes the set of contact values for all possible hole classes (of any level). The set  $Col(r, \theta)$  denotes the set of collapse values for all possible hole classes (of any level). Because of the scaling nature of the contact and collapse values and because we can restrict our attention to holes that are the maximal hole of their class, we have

$$(20) \quad Con(r, \theta) = \{r^k \underline{\epsilon}_H | k \geq 0, H \text{ is a maximal level 0 hole}\},$$

$$(21) \quad Col(r, \theta) = \{r^k \overline{\epsilon}_H | k \geq 0, H \text{ is a maximal level 0 hole}\}.$$

DEFINITION 15. For a pair  $(r, \theta)$ , a *critical  $\epsilon$ -value* of  $(r, \theta)$  is any contact or collapse value. The set of all critical values is denoted by  $Crit(r, \theta)$ . If a tree is simple, then we say  $Crit(r, \theta)$  is empty.

If  $Crit(r, \theta)$  is non-empty, then it has infinitely many elements. Given a level 0 hole class  $H$ , its collapse value  $\overline{\epsilon}_H$  is non-zero. For each  $k \geq 1$ , there is a level  $k$  hole  $H_k$  (corresponding to some level  $k$  address map) whose collapse value is  $r^k \overline{\epsilon}_H$ . Thus there are infinitely many distinct critical values. It is generally straightforward to determine  $\underline{\epsilon}_H$  for a level 0 hole class. Unfortunately,  $\overline{\epsilon}_H$  is generally not as easy to determine. Often a lower and upper bound are sufficient for determining the hole sequence. We now present notation for the bounds that we will consider.

A subscript of ‘gh’ refers to a lower bound for a critical value of  $\epsilon$ . The ‘gh’ stands for ‘guaranteed hole’. A subscript of ‘gc’ refers to an upper bound for a critical value of  $\epsilon$ . The ‘gc’ stands for ‘guaranteed collapse’.

The lower bounds we consider are denoted by  $\epsilon_H^{gh}$ , where we require that there exists a hole of  $[H]$  at  $\epsilon_H^{gh}$ . Thus  $\underline{\epsilon}_H \leq \epsilon_H^{gh} < \overline{\epsilon}_H$ . The upper bounds we consider are denoted by  $\epsilon_H^{gc}$ , where we require that there are no holes of  $[H]$  for any  $\epsilon \geq \epsilon_H^{gc}$ . Thus  $\overline{\epsilon}_H \leq \epsilon_H^{gc}$ . We will use these bounds when dealing with specific examples.

For any tree, if  $\epsilon$  is sufficiently large, then the entire  $\epsilon$ -hull is simply-connected. As  $\epsilon$  decreases to 0, the number of holes may become non-zero (if the tree is non-simple). We now investigate the number of holes as a function of  $\epsilon$ . We are concerned not just with the number of holes, but also the classes of holes.

**DEFINITION 16.** Let  $\theta \in (0^\circ, 180^\circ)$  and let  $r \leq r_{sc}$ . The *hole congruence relation of the pair*  $(r, \theta)$ , denoted by  $\sim_{r, \theta}$ , is defined on the set  $[0, \infty]$  of  $\epsilon$ -values as follows.  $\epsilon_1 \sim_{r, \theta} \epsilon_2$  if the number of holes and the hole classes remain constant for  $\epsilon$  between  $\epsilon_1$  and  $\epsilon_2$  (inclusively).

The hole congruence relation  $\sim_{r, \theta}$  is obviously an equivalence relation.

**DEFINITION 17.** Let  $\theta \in (0^\circ, 180^\circ)$  and let  $r \leq r_{sc}$ . The *hole congruence partition of the pair*  $r, \theta$ , denoted  $\mathcal{HP}(r, \theta)$ , is the partition of  $[0, \infty]$  into the equivalence classes with respect to the hole congruence relation  $\sim_{r, \theta}$ . An equivalence class is denoted by  $[\epsilon]_{r, \theta}$ , or just  $[\epsilon]$  when  $r$  and  $\theta$  are understood.

We can order the equivalence classes of the hole partition as follows. We say that  $[\epsilon_1] \leq [\epsilon_2]$  if and only if  $\epsilon_1 \geq \epsilon_2$ . This reverse ordering is used because when we consider actual trees, we consider what happens as  $\epsilon$  decreases from  $\infty$ , we don't start by looking at small values of  $\epsilon$ . In addition, we use this reverse ordering to define the hole sequence.

**THEOREM 10.** Let  $\epsilon \in \text{Crit}(r, \theta)$ . Then  $[\epsilon] = [\epsilon, \epsilon']$  if  $\epsilon'$  is the next highest element of  $\text{Crit}(r, \theta)$  (so there are no elements of  $\text{Crit}(r, \theta)$  between  $\epsilon$  and  $\epsilon'$ );  $[\epsilon] = [\epsilon, \infty]$  if  $\epsilon$  is the largest value of  $\text{Crit}(r, \theta)$ ; otherwise  $[\epsilon] = \{\epsilon\}$  (if there are critical values higher than  $\epsilon$ , but no next highest).

*Proof.* See [15]. ■

**PROPOSITION 11.** For any non-simple tree  $T(r, \theta)$ , the equivalence class of 0 with respect to the hole congruence relation is the singleton set  $\{0\}$ .

*Proof.* See [11]. ■

Finally we can define the hole sequence of a tree.

**DEFINITION 18.** For a pair  $(r, \theta)$  such that  $r \leq r_{sc}$ , the *hole sequence of the pair*  $(r, \theta)$ , or the *hole sequence of the tree*  $T(r, \theta)$ , is the ordered set of numbers

$$(22) \quad \{N_\epsilon(r, \theta)\} = \{N(r, \theta, \epsilon) \mid \epsilon \in \text{Crit}(r, \theta) \text{ or } \epsilon = 0\},$$

where the set is ordered according to decreasing values of  $\epsilon$ .

We shall see that for many trees the hole sequence is order isomorphic to the natural numbers (*i.e.*, can be indexed by the natural numbers). However, is not true in general. The complications arise when there are non-zero values of  $\epsilon$  that have singleton equivalence classes. An example of how this can occur is with the tree  $T(r_{sc}, 90^\circ)$ , as we shall see in the next section.

**4. Specific examples.** In this section of the paper we look at specific examples of non-overlapping trees and their  $\epsilon$ -hulls. We give one complete detailed example, and then we discuss certain aspects of other trees.

**4.1. Detailed example:  $T(r_{sc}, 45^\circ)$ .** We begin with a detailed example of a specific tree, the self-contacting tree with branching angle  $45^\circ$  (see Figure 15(a)). This tree is a nice example to start with because the geometric calculations are more straightforward than with other angles. This example is worked through in detail to demonstrate the methods that are used for other trees. First, we need to determine  $r_{sc}$ . Using methods detailed in [10], [15] and [16],  $r_{sc}$  is the root of the polynomial  $r^3 + \sqrt{2}r^2 - 1/\sqrt{2}$  that is in  $(0, 1)$ , hence  $r_{sc} \approx 0.59347$ . We claim that the self-contacting hole classes are the only hole classes for the tree. Further discussion is available in [15].

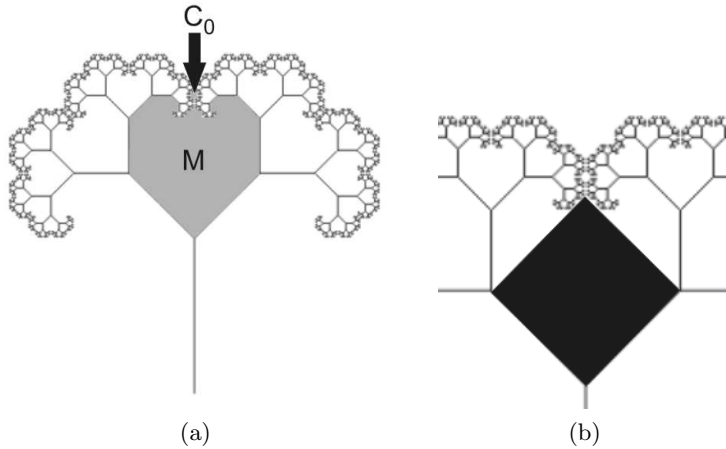


Fig. 15. (a) Level 0 holes of  $T(r_{sc}, 45^\circ)$ , (b) main level 0 hole of  $T(r_{sc}, 45^\circ)$

As discussed in [16], we can identify a self-contacting hole class by giving its location in terms of the upper and lower limit of where the boundary of the maximal hole intersects the  $y$ -axis  $\mathbf{y}$ . The largest hole class is located by the pair  $(\mathbf{A}_0, RL^3(LR)^\infty)$ , and we denote this class by  $M$ . All other hole classes are located by pairs of the form  $(\mathbf{A}C_L, \mathbf{A}C_R)$ , where  $\mathbf{A} \in \mathcal{A}_{2k}$  (see Equation 9) for some  $k \geq 0$ . Recall that  $\mathbf{C}_L$  denotes the address  $LR(RL)^\infty$  and  $\mathbf{C}_R$  denotes the address  $RL(LR)^\infty$ . Let  $C_0$  denote the hole class identified by  $(RL^3\mathbf{C}_L, RL^3\mathbf{C}_R)$ . All other hole classes (except for  $M$ ) are similar to  $C_0$  and are referred to as canopy holes because their boundaries consist entirely of tip points on the canopies of the subtrees  $S_{RL^3}$  and  $S_{LR^3}$ , see Figure 16(a).

Since any canopy hole class is the image of the hole class  $C_0$ , and hence similar to  $C_0$ , we only need to determine the critical  $\epsilon$ -values for  $M$  and  $C_0$ . In general, there are  $2^j$  level 0 canopy holes that are similar to  $C_0$  with contraction factor  $r_{sc}^{2j}$ , via the address maps from addresses in  $\mathcal{A}_{2j}$ . We denote these holes  $C_j$ . The persistence interval for each hole begins with 0, since each hole class is a self-contacting hole class. The exact collapse values of the persistence intervals can also be determined. All hole classes here are symmetric about the  $y$ -axis, so to determine the collapse values we determine the smallest  $\epsilon$  so that the interval between the two locator points of the class is within  $\epsilon$  of the tree.

First we calculate the collapse value  $\overline{\epsilon}_M$  of the largest hole. To do this, we determine which point on the  $y$ -axis inside the hole is furthest from the boundary of the hole.

Consider Figure 15(b), which is a close-up view of the main hole, along with a square formed from the branches  $b(R)$  and  $b(L)$ , and two other new line segments. The top right line segment intersects the tree at the top of the subtree  $S_{RL^4}$ . Further details for the collapse value of the hole class  $M$  are found in [15], and  $\overline{\epsilon}_M \approx 0.32266$ .

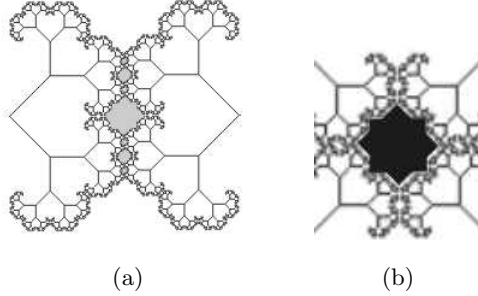


Fig. 16. (a) Close-up of the intersection of  $S_{RL^3}$  and  $S_{LR^3}$  of  $T(r_{sc}, 45^\circ)$ , (b) Canopy Holes Boundary Approximation for  $T(r_{sc}, 45^\circ)$

We can use a similar method to determine the value of  $\overline{\epsilon}_{C_0}$ . Consider the largest canopy hole  $C_0$ . The boundary of this hole can be approximated by a polygon of the form shown in Figure 16(b). Again, further details are available in [11], and we find  $\overline{\epsilon}_{C_0} \approx 0.0175$ .

Now we are able to completely determine the persistence intervals, the hole partition and the hole sequence for this tree.

**PROPOSITION 12.** *The level 0 critical  $\epsilon$ -values for  $T(r_{sc}, 45^\circ)$  are:*

$$\begin{aligned} \epsilon_M &= 0, & \overline{\epsilon}_M &\approx 0.3227, \\ \epsilon_{C_0} &= 0, & \overline{\epsilon}_{C_0} &\approx 0.0175, \\ \epsilon_{C_j} &= 0, & \overline{\epsilon}_{C_j} &= r_{sc}^{2j} \overline{\epsilon}_{C_0} \quad j \geq 0. \end{aligned}$$

To order these critical values in decreasing order, we compare  $\overline{\epsilon}_{C_0}$  to  $\overline{\epsilon}_M$ . For these values we have

$$r_{sc}^6 \overline{\epsilon}_M < \overline{\epsilon}_{C_0} < r_{sc}^5 \overline{\epsilon}_M.$$

The first five non-trivial equivalence classes are given in Table 1. Any other equivalence class is of the form  $[r_{sc}^j \overline{\epsilon}_{C_0}, r_{sc}^{5+j} \overline{\epsilon}_M)$ , where  $j \geq 0$ , or of the form  $[r_{sc}^{6+j} \overline{\epsilon}_M, r_{sc}^j \overline{\epsilon}_{C_0})$ , where  $j \geq 0$ . We will look at these two forms of classes, and we consider the cases when  $j$  is odd and when  $j$  is even.

First we will look at equivalence classes of the form  $[r_{sc}^{6+2k} \overline{\epsilon}_M, r_{sc}^{2k} \overline{\epsilon}_{C_0})$ , where  $k \geq 0$ . Holes of class  $M$  are present at levels 0 through  $6 + 2k - 1$ , thus  $N_M = 2^{6+2k} - 1$ . Holes of class  $C_0$  are present at levels 0 through  $2k$  and  $N_{C_0} = 2^{2k+1} - 1$ . Holes of class  $C_1$  are present at levels 0 through  $2(k - 1)$  and  $N_{C_1} = 2(2^{2k-1} - 1)$ . In general, for integers  $l$  such that  $0 \leq l \leq k$ , holes of type  $C_l$  are present at levels 0 through  $2(k - l)$ . We have  $N_{C_l} = 2^l(2^{2k-2l+1} - 1) = 2^{2k-l+1} - 2^l$ . There are no other classes of holes. The total



**Table 1.** Summary of the first twelve persistence intervals and numbers of holes in the hole partition for  $T(r_{sc}, 45^\circ)$  (the " means that the values from the row above are also included, the superscripts in the middle column represent the level of the hole class)

Equivalence class	Description of holes present	$N([\epsilon])$
$[\overline{\epsilon_M}, \infty)$		0
$[r_{sc}\overline{\epsilon_M}, \overline{\epsilon_M})$	4 $M$	1
$[r_{sc}^2\overline{\epsilon_M}, r_{sc}\overline{\epsilon_M})$	", 2 $M^1$	3
$[r_{sc}^3\overline{\epsilon_M}, r_{sc}^2\overline{\epsilon_M})$	", 4 $M^2$	7
$[r_{sc}^4\overline{\epsilon_M}, r_{sc}^3\overline{\epsilon_M})$	", 8 $M^3$	15
$[r_{sc}^5\overline{\epsilon_M}, r_{sc}^4\overline{\epsilon_M})$	", 16 $M^4$	31
$[\overline{\epsilon_{C_0}}, r_{sc}^5\overline{\epsilon_M})$	", 32 $M^5$	63
$[r_{sc}^6\overline{\epsilon_M}, \overline{\epsilon_{C_0}})$	", 1 $C_0$	64
$[r_{sc}\overline{\epsilon_{C_0}}, r_{sc}^6\overline{\epsilon_M})$	", 64 $M^6$	128
$[r_{sc}^7\overline{\epsilon_M}, r_{sc}\overline{\epsilon_{C_0}})$	", 2 $C_0^1$	130
$[r_{sc}^2\overline{\epsilon_{C_0}}, r_{sc}^7\overline{\epsilon_M})$	", 128 $M^7$	258
$[r_{sc}^8\overline{\epsilon_M}, r_{sc}^2\overline{\epsilon_{C_0}})$	", 2 $C_1$	260

number of holes for this equivalence class is then

$$N([r_{sc}^{6+2k}\overline{\epsilon_M}, r_{sc}^{2k}\overline{\epsilon_{C_0}})) = 2^{6+2k} - 1 + \sum_{l=0}^k (2^{2k-l+1} - 2^l) = 2^{6+2k} + 2^{k+1}(2^{k+1} - 2).$$

Next are equivalence classes of the form  $[r_{sc}^{6+2k+1}\overline{\epsilon_M}, r_{sc}^{2k+1}\overline{\epsilon_{C_0}})$ , where  $k \geq 0$ . Holes of class  $M$  are present at levels 0 through  $6 + 2k$ , thus  $N_M = 2^{7+2k} - 1$ . In general, for integers  $l$  such that  $0 \leq l \leq k$ , holes of class  $C_l$  are present at levels 0 through  $2(k-l) + 1$ . We have  $N_{C_l} = 2^l(2^{2k-2l+2} - 1) = 2^{2k-l+2} - 2^l$ . There are no other classes of holes. The total number of holes for this equivalence class is then

$$N([r_{sc}^{6+2k+1}\overline{\epsilon_M}, r_{sc}^{2k+1}\overline{\epsilon_{C_0}})) = 2^{7+2k} + 2^{k+1}(2^{k+2} - 3).$$

Similarly, we determine that

$$N([r_{sc}^{2k}\overline{\epsilon_{C_0}}, r_{sc}^{5+2k}\overline{\epsilon_M})) = 2^{6+2k} + 2^k(2^k - 2)$$

and

$$N([r_{sc}^{2k+1}\overline{\epsilon_{C_0}}, r_{sc}^{5+2k+1}\overline{\epsilon_M})) = 2^{7+2k} + 2^{k+1}(2^{k+1} - 2).$$

We conclude:

**THEOREM 13.** *For  $T(r_{sc}, 45^\circ)$ , the equivalence classes with respect to hole congruence are completely determined, and thus the hole sequence is also completely determined. We have*

$$N_0 = 0, \quad N_1 = 1, \quad N_2 = 3, \quad N_3 = 7, \quad N_4 = 15, \quad N_5 = 31, \quad N_6 = 63$$

and for  $k \geq 0$ ,

$$\begin{aligned} N_{7+4k} &= N([r_{sc}^{6+2k}\overline{\epsilon_M}, r_{sc}^{2k}\overline{\epsilon_{C_0}})) = 2^{6+2k} + 2^{k+1}(2^{k+1} - 2), \\ N_{8+4k} &= N([r_{sc}^{2k+1}\overline{\epsilon_{C_0}}, r_{sc}^{5+2k+1}\overline{\epsilon_M})) = 2^{7+2k} + 2^{k+1}(2^{k+1} - 2), \end{aligned}$$

$$N_{9+4k} = N([r_{sc}^{6+2k+1}\overline{\epsilon_M}, r_{sc}^{2k+1}\overline{\epsilon_{C_0}})) = 2^{7+2k} + 2^{k+1}(2^{k+2} - 3),$$

$$N_{10+4k} = N([r_{sc}^{2(k+1)}\overline{\epsilon_{C_0}}, r_{sc}^{5+2(k+1)}\overline{\epsilon_M})) = 2^{6+2(k+1)} + 2^{k+1}(2^{k+1} - 2).$$

REMARKS. The hole sequence is order-isomorphic to the natural numbers. The hole sequence is monotonically increasing.

**4.1. Golden 108  $T(r_{sc}, 108^\circ)$ .** We now consider another self-contacting tree:  $T(r_{sc}, 108^\circ)$ . It can be shown that  $r_{sc} = (-1 + \sqrt{5})/2 = 1/\phi \approx 0.618$  [15]. Here  $\phi$  denotes the golden ratio. There are exactly three other self-contacting trees that scale with the golden ratio (corresponding to the angles  $60^\circ$ ,  $120^\circ$  and  $144^\circ$ ) [15].

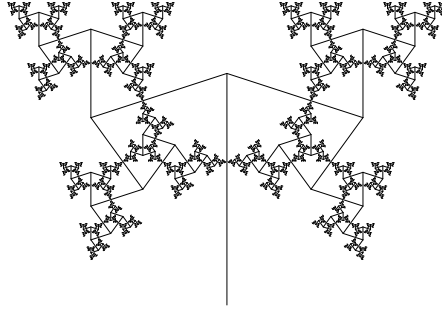


Fig. 17.  $T(1/\phi, 108^\circ)$

This tree has two different types of holes. The main holes are self-contacting holes. The second type arises from contact between two of the top tip points, namely the points  $C_R = RL(LR)^\infty$  and  $C_L = LR(RL)^\infty$ .

The main level 0 hole class is identified by the pair  $(\mathbf{A}_0, RRR(LR)^\infty)$ . Let  $M$  denote this hole class. Let  $S$  denote the level 0 secondary hole class. Consider the top of the subtree  $S_{RRR}$ . The gap between the canopy points  $P_{RRRC_R}$  and  $P_{RRRC_L}$  yields a hole for some non-zero contact value (which equals half the distance between the two points). There are other hole classes that correspond to the other canopy pairs of this subtree, and the contact values decrease towards 0 as the degree of the canopy pairs gets higher. Thus, for any  $\delta > 0$ , there is some canopy hole class whose contact value is less than  $\delta$ , and the persistence of  $M$  must be 0.

Let  $M_k$  denote the main mixed hole class that has contact value  $r_{sc}^k \epsilon_S$  for  $k \geq 1$ . Then, for  $k \geq 2$ , the collapse value of  $M_k$  is the contact value of  $M_{k-1}$ . The collapse value of  $M_1$  corresponds to when the entire region is covered. Without going into the technical details (available in [15]), it is possible to show that

$$\underline{\epsilon_{M_1}} = r_{sc}^2 \epsilon_S, \quad \overline{\epsilon_{M_1}} < \epsilon_{M_1}^{gc} \approx 0.119006,$$

$$\underline{\epsilon_S} = 0.224514, \quad 0.224514 < \overline{\epsilon_S} < \epsilon_S^{gc} \approx 0.236068.$$

This implies  $r_{sc} \overline{\epsilon_{M_1}} < r_{sc}^2 \epsilon_S < r_{sc}^2 \overline{\epsilon_S} < \overline{\epsilon_{M_1}}$ . Thus

$$N_0 = N([\overline{\epsilon_S}, \infty]) = 0,$$

$$N_1 = N([\underline{\epsilon_S}, \overline{\epsilon_S})) = 1 \text{ (level 0 secondary),}$$

$$\begin{aligned} N_2 &= N([r_{sc}\overline{\epsilon_S}, \underline{\epsilon_S}]) = 0, \\ N_3 &= N([r_{sc}\underline{\epsilon_S}, r_{sc}\overline{\epsilon_S}]) = 2 \text{ (level 1 secondary)}, \\ N_4 &= N([\overline{\epsilon_{M_1}}, r_{sc}\underline{\epsilon_S}]) = 0, \end{aligned}$$

and in general:

$$\begin{aligned} N_{5+3k} \ (k \geq 0) &= N([r_{sc}^{k+2}\overline{\epsilon_S}, r_{sc}^k\overline{\epsilon_{M_1}}]) = 2^{k+2} - 2, \\ N_{6+3k} \ (k \geq 0) &= N([r_{sc}^{k+2}\underline{\epsilon_S}, r_{sc}^{k+2}\overline{\epsilon_S}]) = 2^{k+3} - 2, \\ N_{7+3k} \ (k \geq 0) &= N([r_{sc}^{k+1}\overline{\epsilon_{M_1}}, r_{sc}^{k+2}\overline{\epsilon_S}]) = 2^{k+2} - 2. \end{aligned}$$

REMARKS. The hole sequence is order-isomorphic to the natural numbers. The hole sequence is not monotonically increasing. The interesting thing to note here is that in the  $\epsilon$ -hull for  $\epsilon = r_{sc}^2 \underline{\epsilon_S}$ , there are holes in level 2 and level 0, but no holes of level 1. This provides a counter-example to a conjecture that it is not possible to have a discontinuity in the levels of holes possible for any specific  $\epsilon$ .

**4.3. Golden 144  $T(r_{sc}, 144^\circ)$ .** As mentioned in the previous example, another self-contacting tree that scales with the golden ratio is  $T(r_{sc}, 144^\circ)$ . Figure 18 displays an image of the tree.

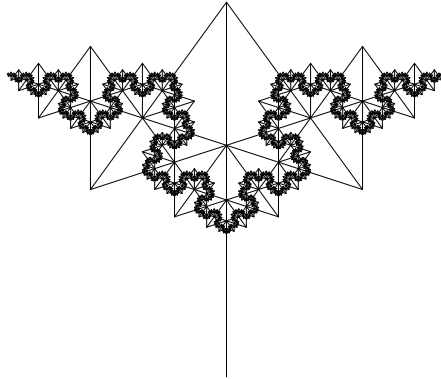


Fig. 18.  $T(1/\phi, 144^\circ)$

The self-contacting hole classes of this tree are the main hole class  $M$  located by  $\mathbf{A}_0$  and  $RR$ ; and the vertex classes  $V_k$  located by  $RR(LR)^k$  and  $RR(LR)^{k+1}$  for  $k \geq 0$ . Let  $P_1 = (x_1, y_1)$  be the point with address  $RRRL$ . Then  $x_1 \approx 0.085757$  and  $y_1 \approx 1 - 0.263932$ . There are non-self-contacting hole classes that correspond to the self-contacting hole classes splitting. For example, consider the point  $P_{RRRL}$ . When there is contact between this point and the trunk, the main self-contacting hole class splits. Thus the main class splits when  $\epsilon = x_1/2$ . The vertex class  $V_k$  splits when  $\epsilon$  is half the value of the  $x$ -coordinate of the point with address  $RR(LR)^{k+1}RL$ . There are no other hole classes [15].

The region of the tree that is bounded by the two branches  $b(RR(LR)^k L)$  and  $b(RR(LR)^{k+1})$ , and the portion of the trunk between the starting point of  $b(RR(LR)^k L)$  and the endpoint of  $b(RR(LR)^{k+1})$ , is similar to the region of the tree bounded by the

branch  $b(R)$ , the branch  $b(RR)$  and the portion of the trunk between  $(0, 1)$  and the end-point of  $b(RR)$ , with a contraction factor of  $r_{sc}^{2(k+1)}$ . This fact reduces the set of critical values that we need to determine.

Let  $MS$  denote the hole class with contact value  $x_1/2$  that is above the point  $P_1$  (so the upper part of the hole class  $M$  that remains after it splits). We refer to it as the ‘main split’ class. Let  $VS$  denote the hole class with contact value  $x_1/2$  that is below the point  $P_1$  (so the lower part of the hole class  $M$  after it splits). We refer to this as the ‘vertex split’ class.

The contact values we have so far are  $\underline{\epsilon}_M = \underline{\epsilon}_{V_k} = 0$  and  $\underline{\epsilon}_{MS} = \underline{\epsilon}_{VS} = x_1/2$ . The collapse values we have so far are  $\overline{\epsilon}_M = x_1/2$  and  $\overline{\epsilon}_{V_k} = r_{sc}^{2(k+1)}x_1/2$ . To find an upper bound for the hole class  $MS$ , we determine the  $\epsilon$ -value such that there is a point in the region that is  $\epsilon$  away from the trunk, the branch  $b(R)$  and the point  $P_1$ . This gives  $\epsilon_{MS}^{gc} \approx 0.0655123$  [15]. To find an upper bound for the hole class  $VS$ , we find the  $\epsilon$ -value such that the point  $(\epsilon, (y_1 + y_2)/2)$  is  $\epsilon$  away from the point  $P_1$ . Solving for this value gives  $\epsilon_{VS}^{gc} \approx 0.0474052$ .

To summarize:  $r_{sc}\overline{\epsilon}_{MS} < \underline{\epsilon}_{VS} < \overline{\epsilon}_{VS} < \overline{\epsilon}_{MS}$ . The hole partition and sequence is given by

$$\begin{aligned} N_0 &= N([\overline{\epsilon}_{MS}, \infty]) = 0, \\ N_1 &= N([\overline{\epsilon}_{VS}, \overline{\epsilon}_{MS})) = 2 \text{ (level 0 main split),} \\ N_2 &= N([\underline{\epsilon}_{VS}, \overline{\epsilon}_{VS})) = 4 \text{ (2 level 0 main split, 2 level 0 vertex split),} \\ N_3 &= N([r_{sc}\overline{\epsilon}_{MS}, \underline{\epsilon}_{VS})) = 2 \text{ (level 0 main),} \\ N_4 &= N([r_{sc}\overline{\epsilon}_{VS}, r_{sc}\overline{\epsilon}_{MS})) = 6 \text{ (2 level 0 main split, 4 level 1 main split),} \\ N_5 &= N([r_{sc}\underline{\epsilon}_{VS}, r_{sc}\overline{\epsilon}_{VS})) = 10 \text{ (same as } N_4 \text{ plus 4 level 1 vertex split).} \end{aligned}$$

In general, given  $k \geq 0$ : For intervals of the form  $[r_{sc}^{2k+1}\overline{\epsilon}_{VS}, r_{sc}^{2k+1}\overline{\epsilon}_{MS})$ , there are main split holes at levels 0 through  $2k$ , with the total number of holes given by

$$N([r_{sc}^{2k}\overline{\epsilon}_{VS}, r_{sc}^{2k}\overline{\epsilon}_{MS})) = \sum_{i=1}^{k+1} 2^{2i} - 2 = \frac{4}{3}(4^{k+1} - 3) - 2(k+1).$$

For intervals of the form  $[r_{sc}^{2k}\overline{\epsilon}_{VS}, r_{sc}^{2k}\overline{\epsilon}_{MS})$ , there are main split holes at levels 0 through  $2k+1$ , with the total number of holes given by

$$N([r_{sc}^{2k+1}\overline{\epsilon}_{VS}, r_{sc}^{2k+1}\overline{\epsilon}_{MS})) = \frac{8}{3}(4^{k+1} - 3) - 2(k+1).$$

For intervals of the form  $[r_{sc}^{2k}\underline{\epsilon}_{VS}, r_{sc}^{2k}\overline{\epsilon}_{VS})$ , there are main and vertex split holes at levels 0 through  $2k$ , with

$$N([r_{sc}^{2k}\underline{\epsilon}_{VS}, r_{sc}^{2k}\overline{\epsilon}_{VS})) = \frac{4}{3}(4^{k+1} - 3) - 2(k+1) + \frac{2}{3}(4^{k+1} - 1).$$

For intervals of the form  $[r_{sc}^{2k+1}\underline{\epsilon}_{VS}, r_{sc}^{2k+1}\overline{\epsilon}_{VS})$ , there are main and vertex split holes at levels 0 through  $2k+1$ , with

$$N([r_{sc}^{2k+1}\underline{\epsilon}_{VS}, r_{sc}^{2k+1}\overline{\epsilon}_{VS})) = \frac{8}{3}(4^{k+1} - 3) - 2(k+1) + \frac{4}{3}(4^{k+1} - 1).$$

For intervals of the form  $[r_{sc}^{2k+1}\overline{\epsilon_{MS}}, r_{sc}^{2k}\underline{\epsilon_{VS}})$ , there are main holes in levels 0 through  $2k$ , and

$$N([r_{sc}^{2k+1}\overline{\epsilon_{MS}}, r_{sc}^{2k}\underline{\epsilon_{VS}})) = \frac{4}{3}(4^{k+1} - 3) - 2(k+1).$$

For intervals of the form  $[r_{sc}^{2k+2}\overline{\epsilon_{MS}}, r_{sc}^{2k+1}\underline{\epsilon_{VS}})$ , there are main holes in levels 0 through  $2k+1$ , and

$$N([r_{sc}^{2k+2}\overline{\epsilon_{MS}}, r_{sc}^{2k+1}\underline{\epsilon_{VS}})) = \frac{8}{3}(4^{k+1} - 3) - 2(k+1).$$

REMARKS. The hole sequence is order-isomorphic to the natural numbers. The main hole class has persistence greater than zero, though it does split. This is different from the other three golden trees where the main hole class has persistence equal to 0.

**4.4. Self-avoiding tree  $T(0.5, 90^\circ)$ .** For self-avoiding trees with branching angle  $90^\circ$ , any hole classes that are not disjoint from the right side of the  $y$ -axis arise from canopy pairs of the subtree  $S_{RLL}$  or the subtree  $S_{RRR}$ . For this particular tree with scaling ratio  $r = 0.5$ , there are no holes above the trunk, so we just need to look at canopy pairs of the subtree  $S_{RRR}$ . Any point with address of the form  $RRR\mathbf{A}$ , where  $\mathbf{A} \in \mathcal{AL}_\infty$ , corresponds to a point on  $S_{RRR}$  that has minimal distance to the trunk. Any such point has the same  $x$ -component, so let  $x_1$  denote this value. With  $r = 0.5$ , we have  $x_1 = 1/3$ . The tip point with address  $RRR(RL)^\infty$  is the highest point on  $S_{RRR}$  that has minimal distance to the trunk, and it is at a distance of  $rx_1 = 1/6$  from the branch  $b(R)$ . Thus every canopy interval will locate a hole because they are all sufficiently far away from the branch  $b(R)$  when  $\epsilon = x_1/2 = 1/6$ . This means that there are infinitely many level 0 holes at  $\epsilon = \frac{1}{6}$ .

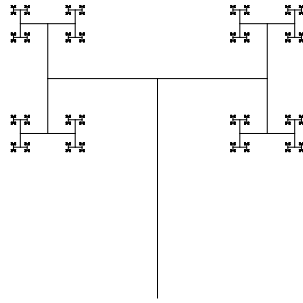


Fig. 19.  $T(0.5, 90^\circ)$

Let  $C_k$  denote the hole classes corresponding to degree  $k$  canopy intervals. It can be shown that for any  $k \geq 0$  [11]:

$$\underline{\epsilon_{C_k}} = \frac{1}{6}, \quad \overline{\epsilon_{C_k}} = \frac{1}{6} + \frac{1}{6(2^{6+4k})}.$$

The collapse values decrease as  $k$  increases, and  $\lim_{k \rightarrow \infty} \overline{\epsilon_{C_k}} = 1/6$ . We also have  $r\overline{\epsilon_{C_0}} \approx 0.0846354 < 1/6$ , so there can be holes in at most one level at a time.

For the hole sequence we will use a double index. The first index denotes the level and the second denotes the highest degree of the canopy interval possible, plus 1. We have

$$\begin{aligned} N_{00} &= N([\overline{\epsilon_{C_0}}, \infty]) = 0, \\ N_{01} &= N([\overline{\epsilon_{C_1}}, \overline{\epsilon_{C_0}})) = 2, \\ N_{02} &= N([\overline{\epsilon_{C_2}}, \overline{\epsilon_{C_1}})) = 6, \end{aligned}$$

and in general, for level 0 and  $k \geq 1$ :

$$N_{0k} = N([\overline{\epsilon_{C_k}}, \overline{\epsilon_{C_{k-1}}})) = 2^{k+1} - 2.$$

The value  $1/6$  is not hole congruent to any other real number, because  $\lim_{k \rightarrow \infty} \overline{\epsilon_{C_k}} = 1/6$ . Thus

$$N_{0\infty} = N(\{\underline{\epsilon_{C_0}}\}) = \infty.$$

We use this notation to signify that there are infinitely many critical values before  $\underline{\epsilon_{C_0}}$ , and that it is the limit of these values. For general levels  $j \geq 1$  and  $k \geq 1$ :

$$\begin{aligned} N_{j0} &= N([r^j \overline{\epsilon_{C_0}}, r^{j-1} \underline{\epsilon_{C_0}})) = 0, \\ N_{jk} &= N([r^j \overline{\epsilon_{C_k}}, r^j \overline{\epsilon_{C_{k-1}}})) = 2^j(2^{k+1} - 2), \\ N_{j\infty} &= N(\{r^j \underline{\epsilon_{C_0}}\}) = \infty. \end{aligned}$$

REMARKS. The hole sequence is not order-isomorphic to the natural numbers, we need to use a double index to order the sequence. There are non-zero values of  $\epsilon$  for which there are infinitely many holes in the corresponding  $\epsilon$ -hull, and these values of  $\epsilon$  form singleton sets in the hole partition. For any specific  $\epsilon$ -value, the  $\epsilon$ -hull of the tree can have holes in at most one level. The hole classes have small persistence. For the level 0 hole classes, the persistence of the class  $C_k$  is

$$P([C_k]) = \frac{1}{6(2^{6+4k})}.$$

For example, for  $k = 5$ ,  $P([C_5]) \approx 2.4835 \times 10^{-9}$ . This small persistence is related to the space-filling nature of the angle. In fact, the persistence of such hole classes decreases as the scaling ratio gets closer to the self-contacting scaling ratio  $1/\sqrt{2}$ .

**4.5. Self-avoiding trees  $T(0.4, 50^\circ)$  and  $T(0.4, 52^\circ)$ .** Figure 10(a) displays  $T(0.4, 50^\circ)$  and Figure 10(c) displays  $T(0.4, 52^\circ)$ .

First we consider the tree  $T(0.4, 50^\circ)$ . It can be shown to have only the main type of holes. Let  $M$  denote the main class of holes.  $P_c = (x_c, y_c)$  denotes the contact point with address  $RL^3(RL)^\infty$ . Then  $x_c \approx 0.21804$  and  $y_c \approx 1 + .46080$ . We have  $\underline{\epsilon_M} = x_c$ . To estimate the collapse value, we determine when a point on the  $y$ -axis is equidistant from the branch  $b(R)$  and the point  $P_c$ . This gives  $\epsilon_M^{gc} \approx 0.76540$ , which suffices to show that  $r\overline{\epsilon_M} < \underline{\epsilon_M}$ . The hole sequence and partition is given by

$$\begin{aligned} N_0 &= N([\overline{\epsilon_M}, \infty]) = 0, \\ N_1 &= N([\underline{\epsilon_M}, \overline{\epsilon_M})) = 1, \\ N_2 &= N([r\overline{\epsilon_M}, \underline{\epsilon_M})) = 0, \end{aligned}$$

$$N_3 = N([r\epsilon_M, r\overline{\epsilon_M})) = 2,$$

$$N_4 = N([r^2\overline{\epsilon_M}, r\epsilon_M)) = 0,$$

and in general, for  $j \geq 1$ ,

$$N_{2j} = N([r^j\overline{\epsilon_M}, r^{j-1}\epsilon_M)) = 0,$$

$$N_{2j+1} = N([r^j\epsilon_M, r^j\overline{\epsilon_M})) = 2^j.$$

REMARKS. For a specific  $\epsilon$ -value, the  $\epsilon$ -hull of the tree can have holes in at most one level. The hole sequence is not monotonically increasing, and every number in the sequence is either 0 or a power of 2. For any sequence of  $\epsilon$ -values where  $\epsilon_n = r\epsilon_{n-1}$ , we have the following growth rate:

$$\lim_{n \rightarrow \infty} \frac{\log(N([\epsilon_n]))}{\log(1/\epsilon_n)} = \frac{\log 2}{\log 1/r}.$$

We will not give all the details for the tree  $T(0.4, 52^\circ)$ . Although the critical values are different than for  $T(0.4, 50^\circ)$ , the hole sequence is the same. The hole sequence for the tree  $T(0.595, 50^\circ)$  in Figure 10(b) is much more complicated because it is possible for holes to be present in more than one level for a specific  $\epsilon$ -value. Thus the hole sequence gives a classification of trees in such a way that  $T(0.4, 50^\circ)$  and  $T(0.4, 52^\circ)$  are of the same class while  $T(0.595, 50^\circ)$  is in a separate class.

**5. Discussion.** In Sections 2 and 3, we developed new concepts and theory to study symmetric binary fractal trees, based on an analysis of their  $\epsilon$ -hulls as  $\epsilon$  ranges through the non-negative real numbers. The specific examples presented in the previous section have shown how rich this extra structure can be. We now revisit the theory. In particular, we discuss the persistence intervals of hole classes and critical  $\epsilon$ -values for a tree and the hole sequences of trees.

**5.1. Persistence intervals.** In studying the  $\epsilon$ -hulls of a tree as  $\epsilon$  ranges over the non-negative real numbers, we are interested in various aspects of the holes (if holes exist at all), not just the number of holes. One important feature is the persistence interval (the interval of  $\epsilon$ -values for which there is exactly one hole that has non-empty intersection with the maximal hole of the hole class) and the persistence (length of the persistence interval) of a hole class. One of the main results of our theory is that every hole class can be obtained from some level 0 hole class via a suitable address map, and the persistence is equal to the persistence of the level 0 hole class scaled by some power of the scaling ratio  $r$ . Once we have determined the persistence intervals of the level 0 holes, we then know all the persistence intervals for the tree, and thus all critical  $\epsilon$ -values.

As seen in the examples, there are trees that have certain hole classes with relatively small persistence (compared to other hole classes associated with the tree). So relatively small holes can make a big difference in our characterizations of the trees. For example, consider the three self-contacting trees  $T(r_{sc}, 108^\circ)$ ,  $T(r_{sc}, 112.5^\circ)$  and  $T(r_{sc}, 120^\circ)$  shown in Figure 20.

The tree  $T(r_{sc}, 112.5)$  has only holes of the main type, while the other two have other types of holes as well. In the case of  $T(r_{sc}, 108^\circ)$ , there are other hole classes above the trunk.  $T(r_{sc}, 120^\circ)$  has the main type of holes along with other classes of holes that arise

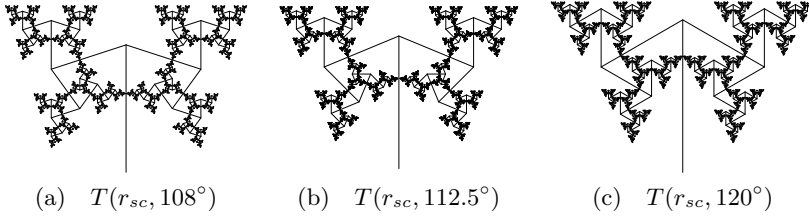


Fig. 20. Three relatively similar looking self-contacting trees with different hole sequences

due to the canopy gaps at the top of the subtree  $S_{RR}$ . The persistence of the secondary contact holes in the case of  $T(r_{sc}, 108^\circ)$  is relatively small compared with the main holes, as is the persistence of the mixed canopy holes of  $T(r_{sc}, 120^\circ)$ . So these small persistence holes make a big difference in the hole sequences of the trees. If we didn't want to make such a distinction between the trees, perhaps we could restrict our attention to level 0 hole classes that have some minimum persistence, so a persistence cutoff value. We have not studied this idea in detail yet, but it is definitely worth investigating. Depending on the persistence cutoff value, a tree with a complicated hole sequence based on our original theory could have a more straightforward hole sequence. However, from a theoretical point of view, it is interesting that our theory does distinguish between trees such as  $T(r_{sc}, 108^\circ)$ ,  $T(r_{sc}, 112.5^\circ)$  and  $T(r_{sc}, 120^\circ)$ , though the distinction may be too fine for the sake of applications.

Another issue is that a persistence cutoff value would eliminate hole classes that have 0 persistence. We have seen trees whose main holes have 0 persistence. To include these types of holes, perhaps we could consider a new definition of hole class and persistence interval. For example, one could define the persistence interval of a hole to be the range of  $\epsilon$ -values for which there is at least one hole that has non-empty intersection with the original hole, as opposed to our requirement that there is exactly one. This would change the nature of persistence intervals, because they would no longer be independent of the hole chosen (because of splitting holes). The set of critical  $\epsilon$ -values would remain the same, but the hole classes would be different. Another definition of persistence could make a distinction between level 0 hole classes that split into hole classes that are all level 0 (*e.g.*  $T(r_{sc}, 144^\circ)$ ) and level 0 hole classes that split into more than one hole class, but only one hole class is still level 0 (*e.g.*  $T(r_{sc}, 120^\circ)$ ). We are currently looking into other definitions of persistence, but have no major results yet. Future work includes a general definition for persistence of holes in  $\epsilon$ -hulls of any set in  $\mathbb{R}^2$ , not just symmetric binary fractal trees.

**5.2. Hole partitions and hole sequences.** The hole sequence is one way to characterize a fractal tree, and it reflects various features of the tree. It gives a 'topological bar-code', a term coined by Carlsson *et al.* [2]. Some of the examples in the previous section were such that the hole sequence was order-isomorphic to the naturals, that is, it could be indexed by the natural numbers. However, this is not true in general, since other sequences are such that they could be indexed by non-standard ordinals, using two indices that each came from the natural numbers. The tree  $T(0.5, 90^\circ)$  is an example of such a tree.



As demonstrated in the examples, the hole partitions and hole sequences have a wide range of complexity. As  $\epsilon$  decreases from  $\infty$ , the hole partition eventually comes to a block of intervals that is repeated (only scaled by a factor of  $r$  after one cycle of the block). For example, a self-contacting tree  $T(r_{sc}, \theta)$  with only main holes has a straightforward hole partition. The first non-trivial interval is  $[r_{sc}\overline{\epsilon_M}, \overline{\epsilon_M})$ , and every other interval is equal to this interval scaled by some factor of  $r_{sc}$ , so the length of the block that is repeated is just 1. On the other hand, a specific example where the length of the repeated block is infinite is the tree  $T(0.5, 90^\circ)$ . So perhaps the length of the repeated block is a way to characterize ‘complexity’ of a tree.

An important feature that the hole sequences indicate is that the growth rate of holes is always equal to the similarity dimension of the tree. A general conjecture for all self-similar fractals regarding growth rates of holes was put forth by Robins in [9]. Our proof is for the analogous result of non-overlapping non-simple symmetric binary fractal trees. The fact that it does not work for non-overlapping trees is not a counter-example, because they are not strictly self-similar. However, it does show that Robins’ conjecture would not work for all fractals with condensation. First we give a lemma that deals with the growth rate of holes for a specific hole class.

LEMMA 14. *Let  $T(r, \theta)$  be a non-simple tree. Let  $[H]$  be any level 0 hole class for the tree. Let  $\epsilon_0 \in P([H])$  be such that  $\epsilon_0 > 0$  and the only holes of type  $[H]$  in  $E(r, \theta, \epsilon_0)$  are level 0. Let  $\{\epsilon_n\}$  be a sequence of  $\epsilon$ -values for  $n \geq 0$  such that  $\epsilon_n = r^n \epsilon_0$ . Then*

$$(23) \quad \lim_{n \rightarrow \infty} \frac{\log(N_{[H]}([\epsilon_n]))}{\log(1/\epsilon_n)} = \frac{\log 2}{\log 1/r}$$

where  $N_{[H]}([\epsilon_n])$  denotes the number of holes of type  $[H]$  for any  $\epsilon$  in the equivalence class  $[\epsilon_n]$  (so the number of holes of the hole class  $[H]$  and its descendant hole classes).

*Proof.* If  $H$  is above the line  $y = 1$ , then  $N_{[H]}([\epsilon_0]) = 1$  and if  $H$  is below the line  $y = 1$ , then  $N_{[H]}([\epsilon_0]) = 2$ . Let  $N'_n = N_{[H]}([\epsilon_n])$ .

First suppose  $[H]$  is a self-contacting hole class, so that  $r^n \underline{\epsilon_H} = 0$  for all  $n \geq 0$ . For any  $n \geq 0$ , we have

$$N'_n = N'_0(1 + 2 + \cdots + 2^n) = N'_0(2^{n+1} - 1).$$

Then

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\log N'_n}{\log(1/\epsilon_n)} &= \lim_{n \rightarrow \infty} \frac{\log N'_0(2^{n+1} - 1)}{\log(1/r^n \epsilon_0)} = \lim_{n \rightarrow \infty} \frac{\log(2^{n+1} - 1) + \log(N'_0)}{\log(1/r^n) + \log(1/\epsilon_0)} \\ &= \lim_{n \rightarrow \infty} \frac{\log(2^{n+1} - 1)}{\log(1/r^n) + \log(1/\epsilon_0)} + \frac{\log(N'_0)}{\log(1/r^n) + \log(1/\epsilon_0)} \\ &= \lim_{n \rightarrow \infty} \frac{\log(2^{n+1} - 1)}{\log(1/r^n) + \log(1/\epsilon_0)} = \lim_{n \rightarrow \infty} \frac{\log(2^{n+1} - 1)}{\log(1/r^n)} \\ &= \lim_{n \rightarrow \infty} \frac{\log 2^{n+1}}{n \log(1/r)} = \lim_{n \rightarrow \infty} \frac{n \log 2 + \log 2}{n \log(1/r)} \\ &= \lim_{n \rightarrow \infty} \frac{n \log 2}{n \log(1/r)} = \frac{\log 2}{\log(1/r)}. \end{aligned}$$

Now suppose the hole class is not a self-contacting hole class. There is a finite integer  $d$  such that there are no level 0 holes of type  $H$  for any  $\epsilon_n$  where  $n \geq d$ , because the contact value of the hole class is non-zero. For any  $\epsilon_n$  where  $n \geq d$ , there are holes of levels  $n - d + 1$  through  $n$ , and

$$N'_n = N'_0(2^{n-d+1} + \dots + 2^n).$$

Then we can use a similar argument as for the self-contacting hole classes to show that

$$\lim_{n \rightarrow \infty} \frac{\log N'_n}{\log(1/\epsilon_n)} = \frac{\log 2}{\log(1/r)}. \blacksquare$$

**COROLLARY 15.** *Let  $T(r, \theta)$  be a non-simple tree. Let  $[H]$  be any hole class for the tree. Let  $\epsilon_0 \in P([H])$  be such that  $\epsilon_0 > 0$ . Let  $\{\epsilon_n\}$  be a sequence of  $\epsilon$ -values for  $n \geq 0$  such that  $\epsilon_n = r^n \epsilon_0$ . Then*

$$\lim_{n \rightarrow \infty} \frac{\log(N_{[H]}([\epsilon_n]))}{\log(1/\epsilon_n)} = \frac{\log 2}{\log 1/r}$$

where  $N_{[H]}([\epsilon_n])$  denotes the number of holes of type  $[H]$  for any  $\epsilon$  in the equivalence class  $[\epsilon_n]$ .

*Proof.* This follows directly from the previous lemma. See [11] for further details.  $\blacksquare$

**THEOREM 16 (Growth Rate of Holes).** *Let  $T(r, \theta)$  be a non-simple tree. Let  $\epsilon_0 > 0$  be such that  $E(r, \theta, r^n \epsilon_0)$  has a finite number of hole classes for all  $n \geq 0$ . For the sequence  $\{\epsilon_n\}$  defined by  $\epsilon_n = r^n \epsilon_0$ , the growth rate of holes is given by*

$$(24) \quad \lim_{n \rightarrow \infty} \frac{\log N([\epsilon_n])}{\log(1/\epsilon_n)} = \frac{\log 2}{\log 1/r}.$$

*Proof.* First, there exists  $m \geq 0$  such that for any  $n \geq m$ , the hole classes of  $E(r, \theta, \epsilon_n)$  are all descendants of the hole classes of  $E(r, \theta, \epsilon_m)$ , otherwise the assumption that  $E(r, \theta, \epsilon_n)$  has a finite number of hole classes for all  $n$  is contradicted. Let  $M$  be the number of hole classes in  $E(r, \theta, \epsilon_m)$ . We can label the hole classes  $[H_i]$  for  $1 \leq i \leq M$ . For  $n \geq m$ , we have

$$N([\epsilon_n]) = \sum_{i=1}^M N_{H_i}([\epsilon_n]),$$

where  $N_{H_i}([\epsilon_n])$  represents the number of holes that descend from  $[H_i]$ . Then

$$\lim_{n \rightarrow \infty} \frac{\log N([\epsilon_n])}{\log(1/\epsilon_n)} = \lim_{n \rightarrow \infty} \frac{\log \sum_{i=1}^M N_{H_i}([\epsilon_n])}{\log(1/\epsilon_n)}.$$

We will evaluate this limit by using the Squeeze Theorem. Set  $N'_M = \sum_{i=1}^M N_{H_i}([\epsilon_m])$ . For  $n \geq m$ , we have

$$2^{n-m} N'_M \leq \sum_{i=1}^M N_{H_i}([\epsilon_n]) \leq (1 + 2 + \dots + 2^{n-m}) N'_M.$$

The lower limit is the value that would correspond to each hole class being such that there can be holes in at most one level for any  $\epsilon$  while the upper limit corresponds to the

hole classes all being self-contacting. As in the proof of Lemma 14, we have

$$\lim_{n \rightarrow \infty} \frac{\log 2^{n-m} N'_M}{\log(1/\epsilon_n)} = \frac{\log 2}{\log 1/r}$$

and

$$\lim_{n \rightarrow \infty} \frac{\log(1 + \dots + 2^{n-m}) N'_M}{\log(1/\epsilon_n)} = \frac{\log 2}{\log 1/r}.$$

Therefore, we have

$$\lim_{n \rightarrow \infty} \frac{\log N([\epsilon_n])}{\log(1/\epsilon_n)} = \frac{\log 2}{\log 1/r}. \blacksquare$$

**6. Conclusions and future work.** The goal of this paper has been to provide foundations for a new way to characterize and classify fractals. Because it is possible for two fractals to have the same fractal dimensions but different topologies, it seems natural to look at topological features of the fractals. To deal with the complex structure of fractals, we considered fractals along with their  $\epsilon$ -hulls.

The main part of this paper provided definitions, terminology, theoretical results and specific examples for the class of non-overlapping symmetric binary fractal trees. The persistence intervals and hole sequence are detailed features of a tree that endow the tree with extra structure. We have a new classification of symmetric binary fractal trees based on the hole sequence. The hole sequences are interesting from a theoretical point of view, but may be difficult to use in applications. Thus some modifications may be needed to apply the theory to natural systems that can be modelled with fractal trees. There are other aspects of the  $\epsilon$ -hulls of trees that provide other types of classifications, and this will be discussed in future work.

One particular aspect is ‘complexity’, where we define complexity of a tree to be related to the highest number of levels of holes that are possible in the  $\epsilon$ -hull of a given tree for any specific  $\epsilon$ -value. Self-contacting trees that are not space-filling have infinite complexity, while all self-avoiding trees have finite complexity. Complexity also allows for the identification of critical angles. A critical angle in this sense marks a change in complexity. The two angles  $90^\circ$  and  $135^\circ$ , which were identified by Mandelbrot and Frame as being topologically critical [10], are interesting in terms of complexity. We have determined that for the branching angle  $90^\circ$ , non-overlapping trees have complexity at most 2, and for the branching angle  $135^\circ$ , the upper limit is 4 [15]. However, it is not true that each angle has a finite upper limit, because there is no upper limit for  $45^\circ$  [15]. We are still working on the details for other branching angles.

Another aspect is location—where on a non-simple tree do the hole classes appear? How does this depend on the parameters? For example, the tree  $T(r_{sc}, 112.5)$  (in Figure 20(b)) is interesting because it indicates a change in the location of holes. All self-contacting trees with branching angle less than  $112.5^\circ$  have some kind of hole class that is above the trunk, while all self-contacting trees with branching angle greater than or equal to  $112.5^\circ$  have no hole classes above the trunk. Thus the angle  $112.5^\circ$  is critical in terms of the location of the holes. We are still investigating the location of hole classes as a function of branching angle, along with critical angles that indicate a change in location.

In conclusion, the hole sequence, or ‘topological bar-code’, does provide new information about a fractal. This approach has proved to be quite fruitful for the class of non-overlapping symmetric binary fractal trees, and should prove to be fruitful for other classes of fractals as well.

**Acknowledgements.** The author wishes to thank the referee for helpful comments. The author also wishes to thank the organizers of the conference. The author is grateful for financial support from the Natural Sciences and Engineering Research Council of Canada.

### References

- [1] E. Carlsson, G. Carlsson and V. de Silva, *An algebraic topological method for feature identification*, Intl. J. Computational Geometry and Applications 16 (2006), 291–314.
- [2] G. Carlsson, A. Zomorodian, A. Collins and L. Guibas, *Persistence barcodes for shapes*, Intl. J. Shape Modeling 11 (2005), 149–187.
- [3] K. Falconer, *Fractal Geometry. Mathematical Foundations and Applications*, Wiley, Chichester, 1990.
- [4] M. Frantz, *Lacunarity, Minkowski content and self-similar sets in  $\mathbb{R}^n$* , in: Fractal Geometry and Applications: a jubilee of Benoît Mandelbrot. Part 1, Proc. Sympos. Pure Math. 72, Part 1, Amer. Math. Soc., Providence, RI, 2004, 77–91.
- [5] D. Gatzouras, *Lacunarity of self-similar and stochastically self-similar sets*, Trans. Amer. Math. Soc. 352 (2000), 1953–1983.
- [6] R. Ghrist, *Barcodes: The persistent topology of data*, Bull. Amer. Math. Soc. 45 (2008), 61–75.
- [7] J. Hutchinson, *Fractals and self-similarity*, Indiana Univ. Math. J. 30 (1981), 713–747.
- [8] B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman, New York, 1982.
- [9] J. R. Munkres, *Elements of Algebraic Topology*, Addison-Wesley, 1984.
- [10] B. Mandelbrot and M. Frame, *The canopy and shortest path in a self-contacting fractal tree*, The Mathematical Intelligencer 21 (1999), 18–27.
- [11] H. O. Peitgen, S. Jürgens and D. Saupe, *Chaos and Fractals: New Frontiers of Science*, Springer-Verlag, New York, 1992.
- [12] V. Robins, *Towards computing homology from finite approximations*, Topology Proceedings 24 (1999), 503–532.
- [13] V. Robins, *Computational topology at multiple resolutions: foundations and applications to fractals and dynamics*, Ph.D. Thesis, University of Colorado, 2000.
- [14] V. Robins, J. D. Meiss and E. Bradley, *Computing connectedness: An exercise in computational topology*, Nonlinearity 11 (1998), 913–922.
- [15] T. D. Taylor, *Computational topology and fractal trees*, Ph.D. Thesis, Dalhousie University, 2005.
- [16] T. D. Taylor, *Homeomorphism classes of self-contacting symmetric binary fractal trees*, Fractals 15 (2007), 9–25.
- [17] A. Zomorodian and G. Carlsson, *Computing persistent homology*, Discrete and Computational Geometry 33 (2005), 249–274.