

## Self-Replicating Tiles and Their Boundary

A. Vince

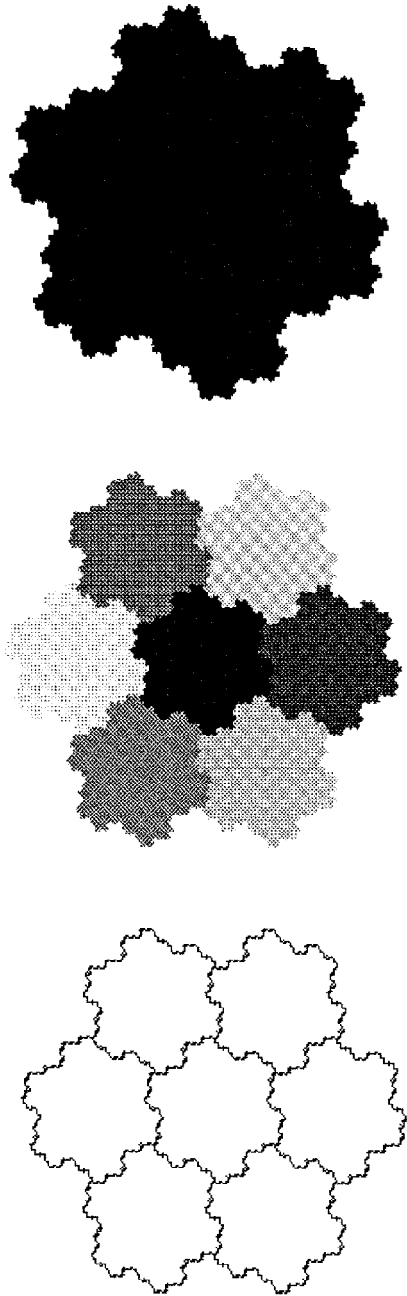
Department of Mathematics, University of Florida,  
Gainesville, FL 32611, USA  
vince@math.ufl.edu

**Abstract.** All self-replicating lattice tilings of the plane can be constructed using special iterated function systems (IFS). Certain self-replicating curves can be constructed using the recurrent set method (RS). A bijection between the IFS parameters and the RS parameters is such that the curve  $K$  produced by the RS parameters is the boundary of the tile  $T$  produced by the IFS parameters. The correspondence is algorithmic in that  $K$  can be drawn from the IFS data using turtle graphics and  $T$  can be drawn from the RS data using an IFS iteration.

### 1. Introduction

A nonempty compact subset  $T$  of  $\mathbb{R}^2$  is called *self-affine* if there exists a linear transformation  $A$  such that  $A(T)$  is the nonoverlapping union of translates of  $T$ . In this paper *nonoverlapping* means that the interiors are disjoint. If, in addition,  $A$  is a similarity, then  $T$  is called *self-similar*. A square is self-similar, but this is not a typical example. Fractal examples appear in Figs. 1 and 2. The Gosper “flowsnake” in Fig. 1 is self-similar and is tiled by seven smaller similar copies. Figure 2, which is tiled by nine smaller similar copies, illustrates that a self-similar set may not be simply connected.

A *tiling* is a collection  $\mathcal{T}$  of nonempty compact subsets of  $\mathbb{R}^2$ , called *tiles*, such that (1) each tile is the closure of its interior, (2) the union of the tiles in  $\mathcal{T}$  is  $\mathbb{R}^2$ , and (3) distinct tiles are nonoverlapping. A *lattice tiling* is a tiling by translates of a single tile by the points of a lattice. A *self-replicating tiling* (rep-tiling) is a tiling  $\mathcal{T}$  by translates of a single tile such that there exists a linear map  $A$  with the following property. For each tile  $T$  in the tiling, the image  $A(T)$  is, in turn, tiled by copies of tiles in  $\mathcal{T}$ . The tilings in Figs. 1 and 2 extend to lattice rep-tilings. Note that the definition of rep-tiling is more stringent than just requiring a lattice tiling by self-affine tiles. There exist tilings by self-affine tiles that are not lattice rep-tilings [14]. Numerous papers on self-replicating tilings can be found in the literature; see, for example, the work of Bandt [1], Gilbert [9], Gröchenig and Haas [10], Kenyon [13], Lagarias and Wang [14], and the author [20].



**Fig. 1.** The Gosper flowsnake—tiling and boundary.

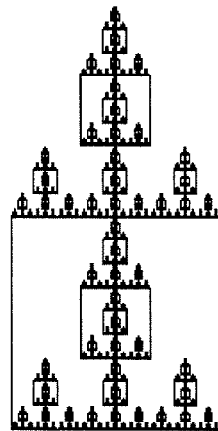
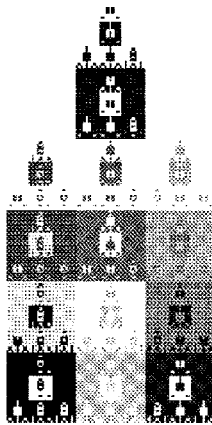
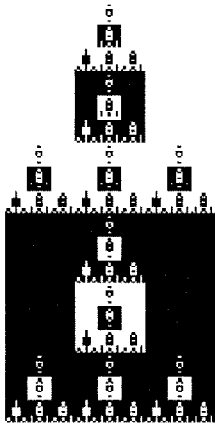


Fig. 2. A nonsimply connected self-affine tile.

In addition to their geometric appeal, there has been an interest in these tilings because of their application to the construction of wavelet bases; see, for example, the work of Gröchenig and Madych [11] and Strichartz [19]. For standard results on tilings see the textbook by Grünbaum and Shephard [12].

The main result of this paper is an explicit bijection between two known methods for constructing lattice rep-tilings of  $\mathbb{R}^2$ . The first method produces the tiles themselves; the second method produces the boundaries of the tiles. An algorithm based on the first method was used to produce the upper tiling in Fig. 1, while an algorithm based on the second method was used to produce the lower tiling; similarly for the tile in Fig. 2. In principle, the method is simple. Start with a parallelogram  $Q$  that is a fundamental domain for the lattice. Iterate  $Q$  in a certain way (using an iterated function system); iterate the boundary of  $Q$  in another way (using the recurrent set method). Under some general conditions the two iterations will converge to a tile  $T$  and a curve  $K$ , respectively, such that  $K$  is the boundary of  $T$ .

The use of iterated function systems (IFSs) to produce “fractiles” is well known, with nice expositions, for example, by Barnsley [2] and Falconer [8]. Many of the illustrations of “fractiles” in the popular literature use IFS-based algorithms. The *recurrent set* method, less well known, but also elegant, is due to Dekking [4], [5]. It uses an algebraic construction related to string rewriting and L-systems. A referee has also brought to our attention an earlier connection between the two methods due to Bedford [3]. That paper has many aspects in common with our treatment, but in the context of constructing Markov partition boundaries for hyperbolic toral endomorphisms.

The paper is organized as follows. In Section 2 a certain triple  $(L, A, D)$ , called *tile data*, is defined and used to construct a tile  $T := T(L, A, D)$ . (Here  $L$  denotes a lattice in the plane.) The IFS method is used in this construction. In Section 3 a certain pair  $(\sigma, f)$ , called *boundary data*, is defined and used to construct a closed curve  $K := K(\sigma, f)$ . The recurrent set method is used in this construction. The main result in Section 4 (Theorem 2) is a correspondence  $(L, A, D) \leftrightarrow (\sigma, f)$  between tile data and boundary data with the following properties. If

$$\mathcal{T} = \{T + x \mid x \in L\}$$

is the set of translations of tile  $T$  by the lattice  $L$ , then under very modest conditions (Theorem 3 in Section 5), essentially that the curve  $K$  is not space filling,

- (1)  $K(\sigma, f)$  is the boundary of  $T(L, A, D)$ , and
- (2)  $\mathcal{T}$  is a self-replicating lattice tiling.

Moreover, every self-replicating lattice tiling can be obtained by this construction for some tile or boundary data.

The figures in this paper are, of course, not actually  $T$  and  $K$ . There are sequences of approximations  $\{T_n\}$  and  $\{K_n\}$  that converge to  $T$  and  $K$ , respectively, in the Hausdorff metric. Our correspondence between tile data and boundary data is algorithmic in the sense that, given tile data, an algorithm computes the boundary data that draws  $K_n$  using turtle graphics. Conversely, given the boundary data, an algorithm computes the tile data that draws  $T_n$  using finite iteration.

Calculation of the Hausdorff dimension of the boundary of a self-similar tile is addressed by Duvall et al. in [7], and for the Levy curve, in particular, in [6].

## 2. Tile Data

A set  $\{f_i\}_{i=1}^k$  of contractions from  $\mathbb{R}^2$  to  $\mathbb{R}^2$  is called an *iterated function system* (IFS). We restrict attention to  $\mathbb{R}^2$  because the methods of Section 3 hold only for the plane. The definitions and results of this section, however, generalize in an obvious way to  $\mathbb{R}^n$ . On the space  $H$  of compact subsets of  $\mathbb{R}^2$ , with respect to the Hausdorff metric, define  $f: H \rightarrow H$  by  $f(X) = \bigcup_{i=1}^k f_i(X)$ , for any compact set  $X$ . It is well known that  $f$  is a contraction on  $H$  and hence, by the contraction mapping theorem,  $f$  has a unique fixed point or *attractor*  $T$  satisfying

$$T = \bigcup_{i=1}^k f_i(T) \tag{2.1}$$

and given by

$$T = \lim_{n \rightarrow \infty} f^{(n)}(X_0), \tag{2.2}$$

where  $f^{(n)}$  denotes the  $n$ th iterate of  $f$ ,  $X_0$  is an arbitrary compact subset of  $\mathbb{R}^2$ , and the limit is with respect to the Hausdorff metric. The set  $T_n = f^{(n)}(X_0)$  is an  $n$ th approximation to  $T$  and is easy to express in algorithmic form. It is usually such an algorithm that is used to produce the fractal graphics that appear in many books and papers on the subject.

Consider the following special case of an IFS. Let  $L$  be a lattice, i.e., all integer linear combinations of two linearly independent vectors in  $\mathbb{R}^2$ . Further, let  $A: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be an expansive linear transformation that preserves the lattice  $L$ . *Expansive* means that all eigenvalues have modulus greater than one. Preserving the lattice means that  $A(L) \subset L$ . Let  $D = \{d_1, d_2, \dots, d_k\}$  be a set of representatives of the cosets in  $L/A(L)$ . Such a set  $D$  will be called a *digit set*. By standard results in algebra  $k = |\det(A)|$ . The triple  $(L, A, D)$  will comprise the “data” out of which a tile  $T := T(L, A, D)$  will be constructed.

To define  $T := T(L, A, D)$ , let  $Q$  be a parallelogram spanned by two vectors that generate the lattice  $L$ . Let  $T_0 = Q$  and let

$$T_n := T_n(L, A, D, Q) = \bigcup \left\{ A^{-n}(Q) + x \mid x \in \sum_{i=1}^n A^{-i}(D) \right\}, \tag{2.3}$$

where the sum is in the Minkowski sense  $X + Y = \{x + y \mid x \in X, y \in Y\}$ . Note that the union is nonoverlapping. Now let

$$T := T(L, A, D) = \lim_{n \rightarrow \infty} T_n = \left\{ x \mid x = \sum_{i=1}^{\infty} A^{-i} d_i, d_i \in D \right\}. \tag{2.4}$$

Since  $A$  is expansive, each  $f_i(x) = A^{-1}(x + d_i)$ ,  $i = 1, 2, \dots, k$ , is a contraction and, according to formula (2.2), the set  $T(L, A, D)$  is the attractor of the IFS  $\{f_i\}_{i=0}^{k-1}$ . Although  $T_n$  depends on the parallelogram  $Q$ , the limit set  $T$  does not.

The tuple  $(L, A, D)$  will be referred to as *tile data* and  $(L, A, D, Q)$  as *extended tile data* if

- (1)  $L$  is a lattice in  $\mathbb{R}^2$ ,
- (2)  $A: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is an expansive linear transformation,
- (3)  $D$  is a digit set, i.e., a complete set of coset representatives of  $L/A(L)$ , and
- (4)  $Q$  is a parallelogram spanned by two generators of  $L$ .

As the attractor of an IFS,  $T$  is compact and, according to (2.1), has the self-affine property

$$A(T) = \bigcup_{d \in D} (T + d), \tag{2.5}$$

where the union is nonoverlapping. Moreover, it is known that  $T(L, A, D)$  is the closure of its interior and its boundary has Lebesgue measure 0; see [10], [18], and [20]. For the tile in Fig. 1, for example, the lattice  $L$  is the hexagonal lattice;

$$A = \begin{pmatrix} \frac{5}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{5}{2} \end{pmatrix};$$

and  $D = \{0, 1, \omega, \omega^2, \omega^3, \omega^4, \omega^5\}$ , where  $\omega$  is a primitive sixth root of unity.

The following result is known [10], [20].

**Theorem 1.** *Given tile data  $(L, A, D)$  there is a self-replicating tiling of  $\mathbb{R}^n$  by translates of  $T(L, A, D)$  to a subset  $S$  of  $L$ . Every self-replicating lattice tiling of the plane is of the form  $\{T(L, A, D) + x \mid x \in L\}$  for some tile data  $(L, A, D)$ .*

**Remark 1.** The tiling in the first sentence of Theorem 1 may not be a lattice tiling. In example (2.6) below, the set  $S$  consists of all lattice points for which the sum of the coordinates is even:

$$L = \mathbb{Z}^2, \quad A = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}, \tag{2.6}$$

$$D = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \end{pmatrix} \right\}.$$

In example (2.7), due to Lagarias and Wang [14], the set  $S$  is not even a lattice:

$$L = \mathbb{Z}^2, \quad A = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}, \quad D = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \end{pmatrix} \right\}. \tag{2.7}$$

Assume that the origin lies in the interior of  $T$  and  $0 \in D$ . Let  $D_n = \sum_{i=0}^{n-1} A^i(D)$  be the set of lattice points representable using  $n$  digits and  $D_\infty = \bigcup_{i=0}^{\infty} D_n$ . Then it is known that  $\{T + x \mid x \in D_\infty\}$  is a self-replicating tiling, but  $D_\infty$  may not be the whole lattice  $L$  or even a sublattice of  $L$ .

In their impressive paper [17], Lagarias and Wang address the conjecture of Gröchenig and Haas that there always exists a lattice tiling by copies of  $T(L, A, D)$  (not necessarily

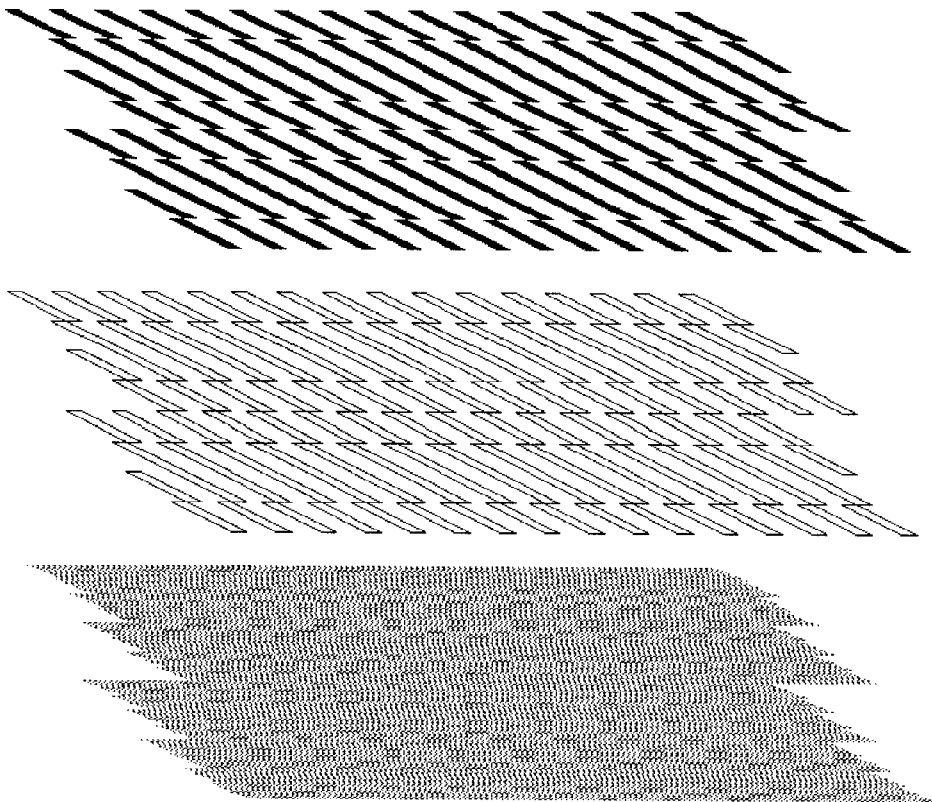


Fig. 3. Approximation of a tile with space-filling boundary.

by the lattice  $L$ , not necessarily self-replicating). They prove the conjecture for self-affine tiles in dimension 2 and for self-similar tiles in all dimensions.

**Remark 2.** The boundary  $\partial T$  of the tile  $T$  may behave badly. Although  $\lim_{n \rightarrow \infty} T_n = T$ , it may be the case that  $\lim_{n \rightarrow \infty} \partial T_n \neq \partial T$ . In fact, in examples (2.6) and (2.7) above  $\lim_{n \rightarrow \infty} \partial T_n$  is the whole tile  $T$ ; the boundary is space filling. The tile  $T$  and its fifth approximation  $T_5$  for example (2.7) are shown in Fig. 3.

The issues brought up in these two remarks are addressed in Theorem 3 of Section 5.

### 3. Boundary Data

In this section we summarize the recurrent set method for constructing a certain closed curve. Let  $G := G \langle a, b \rangle$  be the free group on two generators  $a$  and  $b$ . So  $G$  consists of all words in the letters  $\{a, b, a^{-1}, b^{-1}\}$ , including the empty word  $e$ . The operation is concatenation, and the only relations are  $aa^{-1} = e = a^{-1}a$  and  $bb^{-1} = e = bb^{-1}$ .

Consider an endomorphism  $\sigma: G \rightarrow G$  and a homomorphism  $f: G \rightarrow \mathbb{R}^2$  such that neither  $\sigma(a)$  nor  $\sigma(b)$  is the empty word and  $(f(a), f(b))$  is positively oriented and linearly independent. In other words, if  $f(a) = (a_1, a_2)$  and  $f(b) = (b_1, b_2)$ , then

$$\det \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} > 0.$$

Note that both  $\sigma$  and  $f$  are determined by their action on  $a$  and  $b$ . The pair  $\{\sigma, f\}$  is the “data” from which a closed curve  $K(\sigma, f)$  will be constructed.

To define  $K(\sigma, f)$ , note the existence of a unique linear map  $A = A_{(\sigma, f)}$  that yields a commutative diagram:

$$\begin{array}{ccc} G & \xrightarrow{f} & \mathbb{R}^2 \\ \sigma \downarrow & & A_{(\sigma, f)} \downarrow \\ G & \xrightarrow{f} & \mathbb{R}^2 \end{array}$$

In fact, the matrix for  $A_{(\sigma, f)}$ , with respect to the basis  $(f(a), f(b))$ , is

$$M_\sigma = \begin{pmatrix} m_{aa} & m_{ab} \\ m_{ba} & m_{bb} \end{pmatrix}, \tag{3.1}$$

where  $m_{\alpha\beta}$  is the number of occurrences of  $\alpha$  in  $\sigma(\beta)$ , counting  $\alpha^{-1}$  as occurring  $-1$  time. Here  $\alpha$  and  $\beta$  are each either  $a$  or  $b$ . Let  $w = \alpha_1\alpha_2 \cdots \alpha_n$  be a word in which each  $\alpha_i$  is an element of  $\{a, b, a^{-1}, b^{-1}\}$ , and consider the sequence of points  $x_i \in \mathbb{R}^2$ ,  $i = 0, 1, \dots, n$ , given by  $x_0 = (0, 0)$  and  $x_i = f(\alpha_1\alpha_2 \cdots \alpha_i) = f(\alpha_1) + f(\alpha_2) + \cdots + f(\alpha_i)$ ,  $i \geq 1$ . Join the points  $x_0, x_1, \dots, x_n$  sequentially by directed line segments to obtain a directed polygonal path  $p(w)$  and let

$$K_n := K_n(\sigma, f) = A^{-n} p(\sigma^{(n)}(w_0)),$$

where  $w_0 = aba^{-1}b^{-1}$ . Then  $K_n$  is a closed, directed, polygonal path in  $\mathbb{R}^2$ . It is known [5] that if  $A_{(\sigma, f)}$  is expansive, then the (undirected) sequence  $\{K_n\}$  converges, with respect to the Hausdorff metric, to a closed curve  $K := K(\sigma, f)$ . The curve that is the boundary of the tile in Fig. 1, for example, is  $K(\sigma, f)$  where

$$\begin{aligned} f(a) &= (1, 0), & \sigma(a) &= a^2b, \\ f(b) &= \left(\frac{5}{2}, \frac{\sqrt{3}}{2}\right), & \sigma(b) &= ba^{-1}b^2. \end{aligned}$$

Some line segments may be traversed by  $K_n$  more than one time. We impose the convention that each traversal of a line segment in one direction cancels a traversal of that line segment in the opposite direction. Thus  $K_n$  can consist of several closed curves, and hence  $K_n$ , and also  $K$ , may be disconnected.

The directed sides of the polygon  $A^n(K_n)$  are all translates of the vectors  $\mathbf{u} = f(a)$ ,  $\mathbf{v} = f(b)$ ,  $-\mathbf{u}$ , or  $-\mathbf{v}$ . Given a point  $x$  in the plane, count the number of  $\mathbf{v}$  sides of  $A^n(K_n)$  that intersect a ray with endpoint at  $x$  in the  $\mathbf{u}$  direction ( $+\mathbf{v}$  counts  $+1$ ;  $-\mathbf{v}$



counts  $-1$ ). It is easy to show that this number is the same as the number of  $\mathbf{u}$  sides that lie above  $x$  in the  $-\mathbf{v}$  direction. Call this common number the *winding number* of  $A^n(K_n)$  about the point  $x$ . The *winding number* of  $K_n$  about a point  $x$  is defined as the winding number of  $A^n(K_n)$  about the point  $A^n(x)$ . Note that the winding number makes sense even in the case that  $K_n$  fails to be connected.

It can happen that the winding number of  $K_n$  about a point is more than 1 or negative. In this case there is no well-defined region enclosed by  $K_n$ . The following result makes this situation easy to detect. Its proof is contained as part of the proof of Theorem 2 in Section 4.

**Lemma.** *If the winding number of  $K_1$  about every point of  $\mathbb{R}^2 \setminus K_1$  is either 0 or 1, then the same is true of  $K_n$ ,  $n > 1$ .*

The pair  $\{\sigma, f\}$  used to produce the closed curve  $K(\sigma, f)$  will be referred to as *boundary data* if

- (1)  $\sigma: G \langle a, b \rangle \rightarrow G \langle a, b \rangle$  is an endomorphism,
- (2)  $f: G \rightarrow \mathbb{R}^2$  is a homomorphism such that  $(f(a), f(b))$  is positively oriented and linearly independent,
- (3)  $A_{(\sigma, f)}$  is expansive, and
- (4) the winding number of  $K_1$  about every point of  $\mathbb{R}^2 \setminus K_1$  is either 0 or 1.

**Remark.** Let  $T_n := T_n(\sigma, f)$  denote the closure of the region(s) bounded by  $K_n := K_n(\sigma, f)$ . It will be seen in Section 4 that  $T_n$  also converges, with respect to the Hausdorff metric, to a compact set, say  $T$ . The following issue arises. Although  $K_n$  is the boundary of  $T_n$ , it may be the case that  $K$  is not the boundary of  $T$ . In fact, it is even possible that  $K$  is space-filling. In the following example, it is actually the case that  $K = T$ :

$$\begin{aligned} f(a) &= (1, 0), & \sigma(a) &= a^2, \\ f(b) &= (0, 1), & \sigma(b) &= a^2 b^2 a^{-1} b^{-2} a^{-1} b b a. \end{aligned}$$

In fact, for this example,  $K_n$  is the boundary of the tile  $T_n$  in example (2.7) of Section 2. The tile  $T_5$  and its boundary  $K_5$  are shown in Fig. 3, as well as the whole tile  $T$ . Since there is little hope of relating  $T$  and its boundary in such pathological situations, this problem is addressed in Theorem 3 of Section 5.

#### 4. The Correspondence between Tile and Boundary Data

In Section 2 a tile is constructed from tile data  $(L, A, D)$ , and in Section 3 a closed curved is constructed from boundary data  $(\sigma, f)$ . In this section the two constructions are related to each other.

Given boundary data  $(\sigma, f)$ , let  $L$  be the lattice generated by  $f(a)$  and  $f(b)$ ; let  $A = A_{(\sigma, f)}$ ; and let  $Q$  be the parallelogram spanned by the vectors  $f(a)$  and  $f(b)$ . For a lattice point  $x$ , let  $Q_x$  be the translate to  $Q$  to  $x$ . Let  $D = D_{(\sigma, f)}$  be the set of lattice points  $x$  such that  $Q_x$  is enclosed by  $A(K_1(\sigma, f))$ . Given boundary data  $(\sigma, f)$ , we have now produced extended tile data  $(L, A, D, Q)$ .

**Theorem 2.** *The mapping  $\Theta: (\sigma, f) \mapsto (L, A, D, Q)$  given above induces a bijection from the collection of all boundary data to the collection of all extended tile data such that:*

- (1)  $\partial T_n(L, A, D) = K_n(\sigma, f)$ .
- (2)  $\{T_n + x \mid x \in L\}$  is a tiling of the plane.
- (3)  $A(T_{n+1}) = \bigcup_{d \in D} (T_n + d)$ , where the union is nonoverlapping.

*Proof.* All unions in this proof are nonoverlapping unions. Concerning statement (2), note that, since  $D$  is a set of coset representatives for  $L/A(L)$  we have  $L = D + A(L)$ . Iterating this equation yields  $L = \sum_{i=0}^{n-1} A^i(D) + A^n(L)$ . Letting  $D_n = \sum_{i=0}^{n-1} A^i(D)$ , we have  $\mathbb{R}^2 = \bigcup\{Q + x \mid x \in A^n(L) + D_n\}$ . This implies that  $\mathbb{R}^2 = A^{-n}(\mathbb{R}^2) = \bigcup\{A^{-n}(Q) + x \mid x \in L + A^{-n}(D_n)\} = \bigcup\{T_n + x \mid x \in L\}$ , the last equality by definition (2.3).

Concerning statement (3), definition (2.3) implies  $A(T_{n+1}) = \bigcup\{A^{-n}(Q) + x \mid x \in \sum_{i=1}^n A^{-i}(D) + D\} = \bigcup\{T_n + d \mid d \in D\}$ .

Concerning statement (1), we next prove the following for each  $n \geq 1$ . Recall that  $w_0 = aba^{-1}b^{-1}$ .

$$p(\sigma^{(n)}(w_0)) \text{ bounds } A^n(T_n). \tag{4.1}$$

$$p(\sigma^{(n)}(w_0)) = p(w_1\sigma^{(n-1)}(w_0)w_1^{-1} \dots w_k\sigma^{(n-1)}(w_0)w_k^{-1}), \tag{4.2}$$

where  $p(w_i)$  is a path from the origin to  $A^{n-1}(d_i)$ . We use the term *bounds* in (4.1) to mean that the winding number of  $p(\sigma^{(n)}(w_0))$  about each point in the interior of  $A^n(T_n)$  is 1 and the winding number about each point in the complement of  $A^n(T_n)$  is 0. Note that the lemma in Section 3, as well as statement (1) in Theorem 3, is a consequence of statement (4.1).

Statements (4.1) and (4.2) will be proved by induction. Both are true by definition for  $n = 1$ ; assume they are true for  $n = N$ . Applying  $\sigma$  to the words on both sides of (4.2) yields

$$p(\sigma^{(N+1)}(w_0)) = p(\sigma(w_1)\sigma^{(N)}(w_0)(\sigma(w_1))^{-1} \dots \sigma(w_k)\sigma^{(N)}(w_0)(\sigma(w_k))^{-1}).$$

However, using the commutative diagram in Section 3,  $p(\sigma(w_i))$  is a path from the origin to  $A^N(d_i)$ ; therefore (4.2) is true for  $n = N + 1$ .

By statement (3) of Theorem 2,  $A^{N+1}(T_{N+1}) = \bigcup_{d \in D} (A^N(T_N) + A^N(d))$ . By the induction hypothesis  $A^N(T_N)$  is bounded by  $p(\sigma^{(N)}(w_0))$ . Since the union  $\bigcup_{d \in D} (A^N(T_N) + A^N(d))$  is nonoverlapping, it is bounded by  $p(w_1\sigma^{(N)}(w_0)w_1^{-1} \dots w_k\sigma^{(N)}(w_0)w_k^{-1})$ , where  $p(w_i)$  is a path from the origin to  $A^N(d_i)$ . By (4.2), the set  $A^{N+1}(T_{N+1})$  is bounded by  $p(\sigma^{N+1}(w_0))$ , which proves (4.1) for  $n = N + 1$ .

To show that the mapping  $\Theta$  is onto, consider extended tile data  $(L, A, D, Q)$ . Let homomorphism  $f: G \rightarrow \mathbb{R}^2$  be defined by letting  $f(a)$  and  $f(b)$  be the lattice points (positively oriented) that generate parallelogram  $Q$ . We claim that there exists an endomorphism  $\sigma: G \rightarrow G$  such that  $A_{(\sigma, f)} = A$  and  $D_{(\sigma, f)} = D$ . To see this, let

$$M = \begin{pmatrix} m_{aa} & m_{ab} \\ m_{ba} & m_{bb} \end{pmatrix}$$

be the matrix of  $A$  with respect to the basis  $(f(a), f(b))$ . Then  $\tau(a) = a^{m_{aa}}b^{m_{ba}}$ ,  $\tau(b) = a^{m_{ab}}b^{m_{bb}}$  defines an endomorphism with the property that  $D_{(\tau, f)}$  is a set of coset representatives for  $L/A(L)$ . However, this digit set may not be  $D$ . Proceed in steps to find an endomorphism  $\sigma$  such that  $D_{(\sigma, f)} = D$ . At each step replace some digit  $d$  in the current digit set by either  $d \pm A(f(a))$  or  $d \pm A(f(b))$ . Any digit set for  $L$  and  $A$  can be obtained by a chain of such replacements. Each such replacement can be accomplished by changing the current  $\tau(a) = \alpha$ ,  $\tau(b) = \beta$  to the altered endomorphism of the form  $\tau'(a) = \alpha$ ,  $\tau'(b) = waba^{-1}b^{-1}w^{-1}\beta$  or  $\tau'(a) = wbab^{-1}a^{-1}w^{-1}\alpha$ ,  $\tau'(b) = \beta$ , where  $p(w)$  is a path from the origin to  $d$ . The validity of this procedure follows from the commutative diagram in Section 3. (It is possible that the winding number of the curve  $K_1(\tau, f)$  about points of  $\mathbb{R}^2 \setminus K_1(\tau, f)$  is other than 0 or 1. It is nevertheless the case that the lattice points enclosed by  $A(K_1(\tau, f))$ , counted with multiplicity given by the winding number, give a digit set in the following sense. Let  $\varphi: L \rightarrow \mathbb{Z}$  be the function that counts this multiplicity. Then  $\sum_{x \in C} \varphi(x) = 1$  for each coset  $C$  of  $L/A(L)$ . This extended notion of digit set allows the arguments above to go through without change.)

To show that  $\Theta$  is one-to-one, assume that boundary data  $(\sigma, f)$  and  $(\tau, g)$  map to the same extended tile data under  $\Theta$ . Since  $(f(a), f(b))$  and  $(g(a), g(b))$  span the same parallelogram  $Q$  and are both positively oriented, we have  $f = g$ . Next assume that  $\sigma(a) = \alpha$ ,  $\sigma(b) = \beta$ ,  $\tau(a) = \gamma$ ,  $\tau(b) = \delta$ . Because  $A_{(\sigma, f)} = A_{(\tau, f)}$  we have from (3.1) that  $f(\alpha) = f(\gamma)$  and  $f(\beta) = f(\delta)$ . However, then  $D_{(\sigma, f)} = D_{(\tau, f)}$  implies that  $\alpha = \gamma$  and  $\beta = \delta$  because  $p(\alpha\beta\alpha^{-1}\beta^{-1})$  and  $p(\gamma\delta\gamma^{-1}\delta^{-1})$  enclose the same set of digits. Now we have  $\sigma = \tau$  since  $\sigma$  and  $\tau$  are determined by their values on  $a$  and  $b$ . □

### 5. The Limit Tile and Its Boundary

In light of Theorem 2, fix particular extended tile data  $(L, A, D, Q)$  and corresponding boundary data  $(\sigma, f)$ . We use the notation  $T_n$  for the tile and  $K_n$  for its boundary. The sequences  $\{T_n\}$  and  $\{K_n\}$  converge in the Hausdorff metric. Let

$$K = \lim_{n \rightarrow \infty} K_n,$$

$$T = \lim_{n \rightarrow \infty} T_n.$$

According to statements (1) and (2) of Theorem 2, we know that  $\partial T_n = K_n$  for each  $n$  and that translates of  $T_n$  by the lattice  $L$  yield a tiling of  $\mathbb{R}^2$ . As pointed out in the remarks at the end of Sections 2 and 3, these results may not extend to the limit tile and limit boundary. It may not, in general, be true that  $\partial T = K$  or that  $\{T + x \mid x \in L\}$  is a tiling of  $\mathbb{R}^2$ . The main result of this section provides conditions under which the pathological situations do not occur. In what follows  $T^\circ$  denotes the interior of  $T$ ;  $m$  denotes Lebesgue measure on  $\mathbb{R}^2$ ; and  $|L|$  denotes the area of the fundamental domain of the lattice  $L$ .

**Theorem 3.** *Let  $T$  be a tile constructed from tile data  $(L, A, D)$  and let  $K$  be the curve constructed from corresponding boundary data. Let  $\mathcal{T} = \{T + x \mid x \in L\}$ . The following statements are equivalent:*

- (1)  $m(T) = |L|$ .

- (2)  $K = \partial T$ .
- (3)  $K$  is not space filling.
- (4)  $\mathcal{T}$  is a lattice rep-tiling of the plane.

**Remark 1.** By a result of Lagarias and Wang [14],  $m(T)$  is an integral multiple of  $|L|$ , whether or not the other conditions in Theorem 3 are satisfied.

**Remark 2.** Theorem 3 is mainly concerned with conditions that involve the tile and its boundary. Additional equivalent conditions are given by Gröchenig and Haas [10] and by the author [20]. These conditions are algorithmic, and together with Theorem 3, they provide an effective method to check that the curve generated is actually the boundary of the self-affine tile.

The condition in [20] is in terms of addresses of lattice points and is most easily stated when  $0 \in D$  and  $0 \in T^o$ . A lattice point  $x$  is said to have a *finite address* if it can be expressed in the form  $x = \sum_{i=1}^n A^i d_i$  where  $d_i \in D$ .

**Theorem [20].** *There is a disk  $B$  centered at the origin, with radius depending only on  $A$  and  $D$ , such that  $\mathcal{T}$  is a lattice rep-tiling if and only if each lattice point in  $B$  has a finite address.*

In [20] there is an effective calculation of the radius of  $B$  and an easy algorithm to determine whether a lattice point has a finite address.

The condition in [10] is in terms of a contact matrix. Fix a basis  $e_1, e_2$  of the lattice  $L$  and set  $S_0 = \{\pm e_1, \pm e_2\}$ . Recursively define the sets

$$S_n = \{x \in L \mid (Ax + D) \cap (y + D) \neq \emptyset \text{ for some } y \in S_{n-1}\}$$

and set  $S = \bigcup_{n=1}^\infty S_n$ . It turns out that  $S$  is finite because  $S_n$  eventually stabilizes. Define the entries of an  $(|S| \times |S|)$ -matrix  $C$ , called the *contact matrix*, by

$$c_{x,y} = |(Ax + D) \cap (y + D)| \quad \text{for } x, y \in S.$$

**Theorem [10].**  *$\mathcal{T}$  is a lattice rep-tiling if and only if the contact matrix has spectral radius less than  $|\det A|$ .*

**Remark 3.** The proof of the equivalence of statements (1) and (4) in Theorem 3 is due to Gröchenig and Madych [11], but is so short we include it here for completeness.

*Proof of Theorem 3.* The following easy result about convergence in the Hausdorff metric is used several times in the proof. Assume  $\lim_{n \rightarrow \infty} X_n = X$  in the Hausdorff metric.

- (a) If  $x \notin X$ , then there exists a disk  $B$  centered at  $x$  such that  $X_n \cap B = \emptyset$  for  $n$  sufficiently large.
- (b) If  $x \in X$ , then for any disk  $B$  centered at  $x$  we have  $X_n \cap B \neq \emptyset$  for  $n$  sufficiently large.

(2)  $\Rightarrow$  (4) It is always the case that  $\mathbb{R}^2$  is covered by the tiles in  $\mathcal{T}$  for the following reason. Since  $D$  is a set of coset representatives for  $L/A(L)$  we have  $L = D + A(L)$ . Iterating this equation yields  $L = \sum_{i=0}^{n-1} A^i(D) + A^n(L)$  which implies  $A^{-n}(L) = \sum_{i=1}^n A^{-i}(D) + L$ . Since  $\bigcup_{n=1}^\infty A^{-n}(L)$  is dense in  $\mathbb{R}^2$  and  $\lim_{n \rightarrow \infty} (\sum_{i=1}^n A^{-i}(D)) = T$ , this implies that  $\mathbb{R}^2 = \bigcup\{T + x \mid x \in L\}$ .

The self-replicating property follows directly from property (2.5).

To show that the tiles are nonoverlapping, assume by way of contradiction, the existence of points  $x, y \in L, x \neq y$ , such that  $(T + x)^o \cap (T + y)^o \neq \emptyset$ . By statement (2) we have  $\lim_{n \rightarrow \infty} \partial T_n = \partial T$ . By statement (a) above, there exists a disk  $B$  centered at some point in  $(T + x)^o \cap (T + y)^o$  such that  $B \cap \partial(T_n + x) = \emptyset = B \cap \partial(T_n + y)$  for  $n$  sufficiently large. However, because  $\lim_{n \rightarrow \infty} T_n = T$  we also know, by statement (b), that  $B \cap (T_n + x) \neq \emptyset$  and  $B \cap (T_n + y) \neq \emptyset$  for  $n$  sufficiently large. We now claim that  $B \subseteq (T_n + x) \cap (T_n + y)$ , which contradicts statement (2) of Theorem 2. To prove the claim, assume that  $B$  is not contained in  $T_n + x$ . Then  $B$  would contain points in both  $T_n + x$  and its complement; hence  $B$  would contain a point of  $\partial(T_n + x)$ , a contradiction; similarly for  $B$  being contained in  $T_n + y$ .

(4)  $\Rightarrow$  (2) We first show that  $\partial T \subseteq K$ . Assume that  $x \in \partial T \subset T$ . Because  $\lim_{n \rightarrow \infty} T_n = T$ , statement (b) implies that any disk  $B$  centered at  $x$  must contain a point in  $T_n$  for  $n$  sufficiently large. Since  $x \in \partial T$  the interior of the disk  $B$  contains a point not in  $T$ . So by statement (a) disk  $B$  contains a point in the complement of  $T_n$  for  $n$  sufficiently large. Since  $B$  contains points in both  $T_n$  and its complement, we have  $B \cap K_n = B \cap \partial T_n \neq \emptyset$  for  $n$  sufficiently large. By statement (a) this implies that  $x \in K$  because  $\lim_{n \rightarrow \infty} K_n = K$ .

To prove that  $K \subseteq \partial T$ , assume, by way of contradiction, that there exists a point  $x \in K \setminus \partial T$ . Then  $x \in T^o$ ; otherwise  $x \notin T$  and  $\lim_{n \rightarrow \infty} T_n = T$  would imply, by statement (a), the existence of a disk  $B$  centered at  $x$  such that  $T_n \cap B = \emptyset$  for  $n$  sufficiently large. However, this is impossible because  $x \in K = \lim_{n \rightarrow \infty} \partial T_n$  implies, by statement (b), that  $\partial T_n \cap B \neq \emptyset$  for  $n$  sufficiently large.

Now consider any disk  $B$  centered at  $x$  contained in  $T$ . As mentioned in the paragraph above,  $B$  must contain a point of  $\partial T_n$  for  $n$  sufficiently large. Since  $\{T_n + x \mid x \in L\}$  is a polygonal tiling by Theorem 2, there is another tile  $T_n + y_n, y_n \in L$ , such that  $(T_n + y_n) \cap B \neq \emptyset$ . By taking a subsequence of  $\{y_n\}$ , there exists a single lattice point  $y$  such that  $(T_n + y) \cap B \neq \emptyset$  for arbitrarily large values of  $n$ . Since  $\lim_{n \rightarrow \infty} (T_n + y) = T + y$ , it follows from statement (a) that  $x \in T + y$ . Now  $x \in T^o \cap (T + y)$  implies  $T^o \cap (T + y)^o \neq \emptyset$  because  $T$  is the closure of its interior. However, this contradicts the assumption that  $\{T + x \mid x \in L\}$  is a tiling.

(2)  $\Rightarrow$  (3) By statement (2) we have  $\partial T = K$ . If  $K$  is space filling, then there is a disk  $B \subseteq K = \partial T \subseteq T$ . For the center of  $B$  to be in  $\partial T$ , disk  $B$  must contain points that are not in  $T$ , a contradiction.

(3)  $\Rightarrow$  (4) Assume that  $\mathcal{T}$  is not a tiling. As shown at the very beginning of this proof,  $\mathbb{R}^2$  is covered by the tiles in  $\mathcal{T}$  and the self-replicating property holds. So it must be the case that two tiles  $T + x$  and  $T + y$  overlap. If it can be shown that  $(T + x) \cap (T + y) \subseteq K + x$ , then we are done. Assume not. Then because  $\lim_{n \rightarrow \infty} K_n = K$ , statement (a) implies that there is a disk  $B$  centered at some point of  $[(T + x) \cap (T + y)] \setminus (K + x)$  such that  $B \cap (K_n + x) = \emptyset$  for  $n$  sufficiently large. Because  $\lim_{n \rightarrow \infty} T_n = T$ , by statement (b) disk  $B$  must contain points of both  $T_n + x$  and  $T_n + y$  for  $n$  sufficiently large. Since

$\{T_n + x \mid x \in L\}$  is a tiling by Theorem 2,  $B$  must contain a point of  $\partial(T_n + x) = K_n + x$  for each  $n$  sufficiently large, a contradiction.

(1)  $\Rightarrow$  (4) As already shown,  $\mathcal{T}$  covers  $\mathbb{R}^2$  and  $\mathcal{T}$  has the self-replicating property. It is sufficient to show that tiles do not overlap. Define a function on  $\mathbb{R}^2$  by  $f(x) = \sum_{y \in L} \chi_T(x - y)$ , where  $\chi_T$  is the characteristic function of  $T$ . Because  $\mathcal{T}$  is a covering,  $f(x) \geq 1$  for all  $x$ . If  $Q$  denotes a fundamental parallelogram for the lattice  $L$ , then  $\int_Q f(x) dx = \sum_{y \in L} \int_Q \chi_T(x - y) dx = \int_{\mathbb{R}^2} \chi_T(x) dx = m(T) = |L|$ . This implies that  $f(x) = 1$  almost everywhere, which implies the desired result.

(4)  $\Rightarrow$  (1) Using the same notation as above, if  $\mathcal{T}$  is a tiling, then  $f(x) = 1$  almost everywhere and  $m(T) = \int_{\mathbb{R}^2} \chi_T(x) dx = \int_Q f(x) dx = m(Q) = |L|$ .  $\square$

## References

1. C. Bandt, Self-similar sets 5. Integer matrices and fractal tilings of  $\mathbb{R}^n$ , *Proc. Amer. Math. Soc.* **112** (1991), 549–562.
2. M. Barnsley, *Fractals Everywhere*, Academic Press, Boston, 1988.
3. T. Bedford, Generating special Markov partitions for hyperbolic toral automorphisms using fractals, *Ergodic Theory Dynamical Systems* **6** (1986), 325–333.
4. F.M. Dekking, Recurrent sets, *Adv. in Math.* **44** (1982), 78–103.
5. F.M. Dekking, Replicating superfigures and endomorphisms of free groups, *J. Combin. Theory Ser. A* **32** (1982), 315–320.
6. P. Duvall and J. Keesling, The Hausdorff dimension of the Lévy dragon, preprint.
7. P. Duvall, J. Keesling, and A. Vince, The Hausdorff dimension of the boundary of a self-similar tile, preprint.
8. K. Falconer, *Fractal Geometry*, Wiley, Chichester, 1990.
9. W. J. Gilbert, Fractal geometry derived from complex bases, *Math. Intelligencer* **4** (1982), 78–86.
10. K. Gröchenig and A. Haas, Self-similar lattice tilings, *J. Fourier Anal. Appl.* **1** (1994), 131–170.
11. K. Gröchenig and W.R. Madych, Multiresolution analysis, Haar bases, and self-similar sets, *IEEE Trans. Inform. Theory* **38** (1994), 556–568.
12. B. Grünbaum and G. S. Shephard, *Tilings and Patterns*, Freeman, New York, 1987.
13. R. Kenyon, Self-replicating tilings, in *Symbolic Dynamics and Its Applications* (P. Walters, ed.), Contemporary Mathematics, Vol. 135, Birkhäuser, Boston, 1992, pp. 239–264.
14. J. Lagarias and Y. Wang, Self-affine tiles in  $\mathbb{R}^n$ , *Adv. in Math.* **121** (1996), 21–49.
15. J. Lagarias and Y. Wang, Integral self-affine tiles in  $\mathbb{R}^n$ . I. Standard and nonstandard digit sets., *J. London Math. Soc.* **54** (1996), 161–179.
16. J. Lagarias and Y. Wang, Integral self-affine tiles in  $\mathbb{R}^n$ . II. Lattice tilings, *J. Fourier Anal. Appl.* **3** (1997), 83–102.
17. J. Lagarias and Y. Wang, Tiling the line with translates of one tile, *Invent. Math.* **124** (1996), 341–365.
18. A. Schief, Separation properties for self-similar sets, *Proc. Amer. Math. Soc.* **122** (1994), 111–115.
19. R. S. Strichartz, Wavelets and self-affine tilings, *Constr. Approx.* **9** (1993), 327–346.
20. A. Vince, Replicating tessellations, *SIAM J. Discrete Math.* **6** (1993), 501–521.
21. A. Vince, Rep-tiling Euclidean space, *Aequationes Math.* **50** (1995), 191–215.

Received April 15, 1997, and in revised form November 13, 1997, and April 6, 1998.