Applied and Computational Harmonic Analysis **7**, 211–238 (1999)

Self-Affine Tilings with Several Tiles, I

Karlheinz Gröchenig and Andrew Haas

Department of Mathematics, The University of Connecticut, Storrs, Connecticut, 06269-3009
E-mail: groch@math.uconn.edu; haas@math.uconn.edu

and

Albert Raugi

Départment de Mathématiques, Université de Rennes I, Campus de Beaulieu, F-35000 Rennes Cedex, France
E-mail: Albert.Raugi@univ-rennes1.fr

Communicated by W.R. Madych

Received December 24, 1997; revised July 22, 1998

The tilings of $\mathbb{R}^d$ by a finite number of lattice translates of self-affine prototiles are studied in their own right and as they relate to multiwavelet bases of $L^2(\mathbb{R}^d)$.

1. INTRODUCTION

The mathematical literature provides several notions of "self-similar tilings," which differ mainly by the group of motions that act on the prototiles [11, 15, 21, 22, 24]. The vague label is used to describe such strikingly different objects as the aperiodic Penrose tilings [21] and the periodic tilings obtained from the "twin-dragon fractal" [10, 19, 25].

In this paper we study self-affine tilings of $\mathbb{R}^d$ by a finite set of compact prototiles T_i, which tile $\mathbb{R}^d$ by translations in a lattice $\Lambda \subseteq \mathbb{Z}^d$. More precisely, a finite collection of sets $\mathcal{T} = \{T_i \subseteq \mathbb{R}^d\}_{i=1}^M$, consisting of sets that are either compact or empty, is said to Λ-tile $\mathbb{R}^d$ if

$$T_i \cap T_j \cong \emptyset \qquad \text{for } i \neq j, \tag{1}$$

setting $T = \bigcup_{i=1}^M T_i$

$$\bigcup_{k\in\Lambda} (k + T) \cong \mathbb{R}^d \tag{2}$$

and

$$T \cap (k + T) \cong \emptyset \qquad \text{for } k \in \Lambda,\ k \neq 0. \tag{3}$$

For two sets $X, Y \subseteq \mathbb{R}^d$ we write $X \cong Y$ if their symmetric difference $X \triangle Y = (X \setminus Y) \cup (Y \setminus X)$ has Lebesgue measure 0, i.e., $|X \triangle Y| = 0$. As usual, the sum of sets $X, Y \subseteq \mathbb{R}^d$

1063-5203/99 $30.00

is defined by $X + Y = \{z = x + y \mid x \in X,\ \ y \in Y\}$ if both X and Y are nonempty and $X + Y = \emptyset$ otherwise. Allowing prototiles to be empty makes it easier to access an important facet of the tiling problem, which is easily overlooked when all tiles are assumed to be nonempty. A compact set T satisfying (2) and (3) is called a Λ*-tile*. It follows from (3) that if T is a $\mathbb{Z}^d$ tile, then T has measure one.

Sets $T_i \subseteq \mathbb{R}^d$, $1 \leq i \leq M$, that are either compact or empty form a (A, Γ)*-self-affine collection* if there is an integer-valued $d \times d$-matrix A with all eigenvalues of modulus greater than one and finite (possibly empty) sets $\Gamma_{ij} \subseteq \mathbb{Z}^d$, $i, j = 1, \ldots, M$, so that

$$AT_i = \bigcup_{j=1}^{M} (\Gamma_{ij} + T_j) \qquad \text{for } i = 1, \ldots, M, \tag{4}$$

and for any i, j, k

$$(\beta + T_i) \cap (\gamma + T_j) \cong \emptyset \qquad \text{for } \beta \in \Gamma_{ki},\ \gamma \in \Gamma_{kj} \text{ and } i \neq j \text{ or } \beta \neq \gamma. \tag{5}$$

The matrix A is usually called a *dilation* or an expanding matrix and the set $\Gamma = \{\Gamma_{ij}\}$ is a *digit set*. Note that the essential disjointness in (5) needs only hold for different sets in the same equation of (4).

If sets $\mathcal{T} = \{T_i,\ i = 1, \ldots, M\}$ form a self-affine collection for some A and Γ and Λ-tile $\mathbb{R}^d$, then we call them a *self-affine* Λ*-tiling set with* M *prototiles*, abbreviated SAT.

The stipulation that the prototiles are positioned in the tiling by translation in a lattice is rather restrictive and excludes many self-similar (self-affine) tilings that appear in the literature [11, 15, 21, 22, 24].

The SATs with one prototile have been well studied. Interest in them became more intense after the discovery of a connection to wavelet theory [10]. It is known through papers of Gröchenig and Haas, Lagarias and Wang, and Conze et al. [2, 3, 9, 18] in dimensions $d = 1,\ d = 2,\ d \geq 3$, respectively, that if the single set $\Gamma = \Gamma_{11}$ is a complete set of coset representatives for the group $\mathbb{Z}^d/A\mathbb{Z}^d$, then there is a compact self-affine set Q solving (4) and a lattice $\Lambda \subseteq \mathbb{Z}^d$, for which Q is a Λ-tile. In dimension $d = 1$ this lattice is explicitly $\Lambda = n\mathbb{Z}$, where n is the greatest common divisor of elements in Γ (where we assumed without loss of generality that $0 \in \Gamma$). In dimension $d \geq 2$ it is known how to determine Λ in principle [3, 18], but the problem of characterizing Λ explicitly in terms of A and Γ is still unresolved. Also, given a dilation matrix A it is not yet known if there exists a digit set Γ for which $\Lambda = \mathbb{Z}^d$ [17]. In short, even in the simplest case $M = 1$ of a single prototile there are still large gaps in the theory.

The focus of this paper is the class of SATs with $M > 1$ prototiles. Given A and $\Gamma = \{\Gamma_{ij},\ i, j = 1, \ldots, M\}$, we show there exist finitely many (A, Γ)-self-affine collections and we give necessary conditions for a collection to be a Λ-tiling, with special attention to the case $\Lambda = \mathbb{Z}^d$. It is then possible to establish a connection to wavelet theory, similar to the one in [10]; that is, between SATs with several prototiles and multiwavelet bases. We show that in general an SAT determines a multiwavelet basis of $L^2(\mathbb{R}^d)$ and vice versa. This provides the first systematic construction of multiwavelet bases in higher dimensions with arbitrary dilation matrices.

Solutions of the dilation equations (4) can be described by means of digit expansions, in which the allowable sequence of digits in an expansion resembles the orbit of a point under a subshift of finite type. As in the one tile case, $M = 1$, (4) has exactly one solution with all compact prototiles. In contrast to the one tile case, when some prototiles are allowed to be empty there may exist several solutions. The familiar approach of constructing a self-similar set as the fixed point of an associated iterated function system will only produce the solution of equations (4) with all prototiles compact, but a slight modification, for which it is possible to have multiple attracting fixed points and also attracting periodic orbits, will suffice to generate all solutions.

We define the notion of a standard digit set Γ, resembling the standard digit sets introduced in [10, 16]. Elementary arguments show that for a self-affine collection to be a $\mathbb{Z}^d$-tiling, Γ must be a standard digit set. As is shown by many examples, this condition is far from sufficient. The theory of Markoff chains over a finite state space is used to analyze the self-affine collection associated to a standard digit set and results in a necessary condition for $\mathbb{Z}^d$-tiling in terms of a Markoff chain determined by the overall structure of the dilation equations. Other conditions are given that take account of both the overall structure of the equations and the specific digits involved.

If the characteristic functions $\chi_{T_i}(x)$ are combined in a column vector $\Phi(x)$, then the dilation equation (4) can be written as

$$\Phi(x) = \sum_{k \in \mathbb{Z}^d} C_k \Phi(Ax - k),$$

where C_k is an $M \times M$ matrix with entries being either 0 or 1. Such equations are among the main objects in wavelet theory and are called vector-valued scaling relations or vector-valued refinement equations. If a self-affine collection is a $\mathbb{Z}^d$-tiling, we show that Φ gives rise to a multiresolution analysis with multiplicity M. Conversely, to any multiresolution analysis whose basis functions are characteristic functions corresponds a self-affine $\mathbb{Z}^d$-tiling. To every self-affine $\mathbb{Z}^d$-tiling we then construct a particular orthonormal basis of $L^2(\mathbb{R}^d)$, a so-called wavelet basis. These results complement the theory of multiwavelets [7, 8, 13] with concrete examples and extend the work in [10] to higher multiplicities.

The paper is organized as follows: The first section is an example “zoo” and by strolling through it the reader should get a better sense of the objects under consideration. It was the confusing and mysterious variety of examples that initially sparked our interest, and we hope that they will stimulate the reader’s curiosity. In Section 3 we construct and classify the general solutions of (4) and in Section 4 we derive a necessary condition for a solution of (4) to be a $\mathbb{Z}^d$-tiling. Section 5 establishes the relation to the theory of multiwavelets and constructs a class of orthonormal bases for $L^2(\mathbb{R}^d)$ starting from a SAT.

In a sequel to this paper we will use Fourier analytic methods and the theory of the transfer operator to study SATs in a more systematic fashion.

While preparing the final version of this manuscript we became aware of an interesting preprint [6] of Flaherty and Wang titled “Haar-type multiwavelet bases and self-affine multi-tiles” which overlaps slightly with our Section 5 in results, but not in its approach.

2. EXAMPLES

In order to demonstrate the almost confusing wealth of different phenomena, the examples will be presented first, with the more involved details left to the end of the section.

In Section 3 it will be shown that the dilation equation (4) always has a unique (maximal) solution for which each of the T_i's is compact. This solution is denoted $\mathcal{Q} = \{Q_i\}$. The dilation equations have the trivial solution $T_i = \emptyset$, $i = 1, \ldots, M$, and often other solutions as well. We assume, throughout this discussion, that for some i, $T_i \neq \emptyset$.

2.1. A General Example

We begin with a construction that produces nontrivial examples in any dimension d. It is based on the existence of lattice tilings with a single tile. Let $\Gamma = \{\gamma_1, \ldots, \gamma_q\}$ be a set of coset representatives for the group $\mathbb{Z}^d / A\mathbb{Z}^d$ where A is expanding and $|\det A| = q$. By the theorem [3, 18] there is a lattice $\Lambda \subseteq \mathbb{Z}^d$ and a unique compact solution Q to the equation $AT = \Gamma + T$, so that $\Lambda + Q$ is a tiling of $\mathbb{R}^d$.

Now choose arbitrary $\alpha_i \in \Lambda$, $i = 1, \ldots, q$, set $\gamma_{ij} = \gamma_i + A\alpha_i - \alpha_j$ and $\Gamma_{ij} = \{\gamma_{ij}\}$. The dilation equations (4) then have the solution $\mathcal{Q} = \{Q_i\}$ with $Q_i = A^{-1}Q + A^{-1}\gamma_i + \alpha_i$. Since the sets $\beta + Q = \bigcup_{i=1}^{q} (\beta + A^{-1}\gamma_i + A^{-1}Q)$ are all disjoint for $\beta \in \Lambda$, so are the sets Q_i. Thus insofar as we understand Q, we can construct a self-affine Λ-tiling set $\mathcal{Q} = \{Q_i\}$.

In particular the choice $\Gamma = \{\gamma_1 = (0,0),\ \gamma_2 = (0,1)\}$, $\alpha_1 = (0,0)$, and $\alpha_2 = (1,1)$ for the dilation matrix

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

yields the equations

$$\begin{aligned} AT_1 &= T_1 \cup ((-1,-1) + T_2), \\ AT_2 &= ((0,3) + T_1) \cup ((-1,2) + T_2). \end{aligned} \tag{6}$$

They have a solution which consists of two contracted, translated copies of the well-known twin dragon, cf. Fig. 1. Since the solution of $AT = \Gamma + T$ yields a $\mathbb{Z}^2$-tiling [10], by the above argument $\mathcal{Q}$ also yields a $\mathbb{Z}^2$-tiling of the plane.

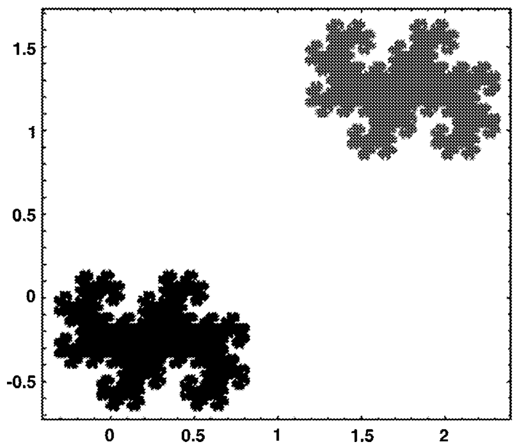

FIGURE 1

2.2. *Examples with $d = 1$ and $M > 1$*

For simplicity we mainly consider examples with $M = 2$ or $M = 3$ and $Ax = 2x$ or $Ax = 3x$ in dimension $d = 1$. Although this may seem a very special case, it already shows the puzzling variety of phenomena in the tiling problem with several tiles.

EXAMPLE 1.

$$\begin{aligned} 2T_1 &= a + T_2, \\ 2T_2 &= T_1 \cup (b + T_1). \end{aligned} \tag{7}$$

Taking measure we have $2|T_1| = |T_2|$ and $2|T_2| \leq 2|T_1|$. It follows that any solution has measure zero, and certainly these sets do not tile $\mathbb{R}$.

EXAMPLE 2.

$$\begin{aligned} 3T_1 &= T_1 \cup (1 + T_1) \cup (2 + T_1) \cup T_3, \\ 3T_2 &= (2 + T_2) \cup (3 + T_2) \cup (4 + T_2) \cup T_4, \\ 3T_3 &= (-2 + T_3) \cup (-4 + T_3) \cup (1 + T_4), \\ 3T_4 &= 2 + T_4. \end{aligned} \tag{8}$$

These equations have a number of different sets of solutions in each of which T_3 and T_4 are sets of measure zero. The simplest solutions are: $T_1 = [0, 1]$, $T_i = \emptyset$ for $i \neq 1$; $T_2 = [1, 2]$, $T_i = \emptyset$ for $i \neq 2$; and $T_1 = [0, 1]$, $T_2 = [1, 2]$, $T_i = \emptyset, i = 3, 4$. The first two solutions are $\mathbb{Z}$-tilings, whereas the third one is a $2\mathbb{Z}$-tiling set. There are three other solutions for which the T_i are either compact or empty. For the "maximal" solution with all T_i compact $Q_4 = \{1\}$, Q_3 is a Cantor set, and Q_1 and Q_2 are fractal sets containing the intervals $[0, 1]$ and $[1, 2]$, respectively. As in (7), the form of the dilation equations, without consideration of the particular digits, forces some prototiles to have measure zero. More elaborate instances of this behavior will be considered in Section 4.

EXAMPLE 3. For $q \in \mathbb{Z}$ consider the equations

$$qT_i = \bigcup_{j=1}^{q} (\alpha_{ij} + T_i) \qquad \text{for } i = 1, \ldots, M,\ \alpha_{ij} \in \mathbb{Z}. \tag{9}$$

The equations decouple and an area argument shows that each equation, and hence the whole set of them, determines a self-affine collection. Observe that some of the prototiles may be chosen to be empty. However, if two T_i's have positive measure, then $\mathcal{T} = \{T_i\}$ is not a $\mathbb{Z}$-tiling set.

The question of when a set is a Λ-tiling set for a lattice Λ is subtle. In the next example we look at a special case where no such Λ exists.

EXAMPLE 4. Assume that $a, c \equiv 1 \pmod 2$ and $b \equiv 0 \pmod 2$ and consider

$$\begin{aligned} 2T_1 &= T_2 \cup (a + T_2), \\ 2T_2 &= (b + T_1) \cup (c + T_1). \end{aligned} \tag{10}$$

In this example there is a unique compact solution which is a self-affine collection. In fact, multiplying the equations by 2 and substituting (10) for $2T_i$, we obtain the decoupled equations

$$\begin{aligned} 4T_1 &= (b+T_1)\cup(c+T_1)\cup(b+2a+T_1)\cup(c+2a+T_1), \\ 4T_2 &= (2b+T_2)\cup(2c+T_2)\cup(a+2b+T_2)\cup(a+2c+T_2). \end{aligned} \tag{11}$$

From our choice of digits a, b, c it is easily deduced that the digit sets $\Gamma_1' = \{b, c, 2a+b, 2a+c\}$ and $\Gamma_2' = \{2b, 2c, a+2b, a+2c\}$ are both congruent to $\mathbb{Z}/4\mathbb{Z}$. Therefore by the tiling theorem [9, Theorem 2.3], both Q_1 and Q_2 are lattice tiles. Consequently, $Q = Q_1 \cup Q_2$ has measure > 1 and is not a $\mathbb{Z}$-tile. Then $\{Q_i\}$ is not a self-affine $\mathbb{Z}$-tiling set either.

The simplest choice of digits is $a = 1$, $b = 0$, $c = 1$, which results in $Q_1 = Q_2 = [0, 1]$. When $a = 1$, $b = 1$, $c = 2$, we get $Q_1 = [1/3, 4/3]$ and $Q_2 = [2/3, 5/3]$. In general, one obtains tiling sets of infinite connectivity.

For the particular choice $a = 1$, $b = 2$, $c = 5$ we show that $\{Q_i\}$ is not a lattice tiling for any lattice Λ, although each Q_i is a $\mathbb{Z}$-tile by itself. In this case $\Gamma_1' = \{2, 4, 5, 7\}$ and $\Gamma_2' = \{4, 5, 10, 11\}$. Since by the digit expansion theorem [10], see also Section 3, $x \in Q_i$ if and only if $x = \sum_{j=1}^{\infty} 4^{-j}\epsilon_j$, $\epsilon_j \in \Gamma_i'$, we deduce that $Q_1 \subseteq [2/3, 7/3]$ and $Q_2 \subseteq [4/3, 11/3]$. Again by [9, Theorem 2.3], $\mathbb{Z} + Q_i \cong \mathbb{R}$ and thus $I = [4/3, 7/3] \subseteq Q_1 \cup (1 + Q_1)$ and $I \subseteq Q_2 \cup (-1+Q_2) \cup (-2+Q_2)$. It is easy to see that $|I \cap (\epsilon + Q_i)| > 0$ for $i = 1, 2$ and the corresponding translates ϵ.

On the other hand, if $Q_1 \cup Q_2$ is a Λ-tile, then necessarily $\Lambda = 2\mathbb{Z}$, since $|Q_1 \cup Q_2| \leq 2$. This implies that $I \subseteq Q_2 \cup (-2+Q_2)$ and thus $|I \cap (-1+Q_2)| = 0$ yields a contradiction.

We conjecture that in general, if Γ_i' does not consist of consecutive integers, then $\mathcal{Q}$ cannot be a lattice tiling set.

EXAMPLE 5.

$$\begin{aligned} 2T_1 &= T_2 \cup (2+T_2), \\ 2T_2 &= T_1 \cup (1+T_1) \cup (2+T_2). \end{aligned} \tag{12}$$

The unique compact solution is $Q_1 = Q_2 = [0, 2]$. The first equation expresses $2Q_1$ as a disjoint union of translates of Q_2, but in the second equation $2Q_2$ is a union of overlapping intervals. Thus (4) and (5) are only satisfied for the first equation, and $\mathcal{Q}$ is not a self-affine collection. Nonetheless they satisfy (2) and (3) with $\Lambda = 2\mathbb{Z}$.

EXAMPLE 6.

$$\begin{aligned} 2T_1 &= T_1 \cup (-1+T_3), \\ 2T_2 &= (2+T_1) \cup (1+T_3), \\ 2T_3 &= (4+T_1) \cup (4+T_2). \end{aligned} \tag{13}$$

The unique compact solution is $Q_1 = [0, 1]$, $Q_2 = [1, 2]$, and $Q_3 = [2, 3]$. Thus $\mathcal{Q}$ is a self-affine $3\mathbb{Z}$-tiling set. This example will come back to haunt us.

EXAMPLE 7.

$$\begin{aligned} 2T_1 &= T_1 \cup T_2, \\ 2T_2 &= (1+T_1) \cup (3+T_2). \end{aligned} \tag{14}$$

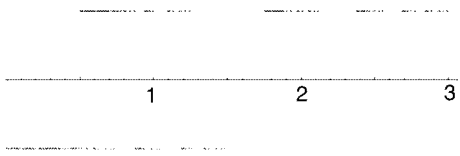

FIGURE 2

This is the most interesting and complicated of the one-dimensional examples we consider. We show that the compact solution $\mathcal{Q}$ is a self-affine $\mathbb{Z}$-tiling set at the end of this section. See Fig. 2.

The more general case

$$2T_1 = T_1 \cup (a + T_2),$$
$$2T_2 = (b + T_1) \cup (c + T_2)$$

is already beyond the scope of this paper and will be considered in the second part. It is worth noting that when $a = 0$, $b = 1$, $c = 1$, the solution is $Q_1 = [0, 1/2]$, $Q_2 = [1/2, 1]$, a self-affine $\mathbb{Z}$-tiling set. On the other hand, for $a = 1$, $b = 1$, $c = 0$, yields $Q_1 = [0, 1]$, $Q_2 = [0, 1]$, which does not satisfy (1), although it is a self-affine collection. While the digit sets are very similar, the second example results in an obvious redundancy.

2.3. *Examples with $d = 2$ and $M = 2$*

EXAMPLE 8. Let

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

and consider

$$\begin{aligned} AT_1 &= T_1 \cup ((1,0) + T_1) \cup ((1,1) + T_1) \cup ((1,0) + T_2), \\ AT_2 &= T_2 \cup ((0,1) + T_2) \cup ((1,1) + T_2) \cup ((0,1) + T_1). \end{aligned} \tag{15}$$

These equations yield triangles as a solution; Q_1 having vertices $(0,0), (1,0), (1,1)$, and Q_2 having vertices $(0,0), (0,1), (1,1)$. $\mathcal{Q}$ is a $\mathbb{Z}^2$-tiling set related to the very basic square tiling of the plane.

Another example which is illustrated in Fig. 3 with

$$A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

is

$$AT_1 = T_1 \cup T_2,$$
$$AT_2 = ((1,0) + T_2) \cup T_3,$$
$$AT_3 = ((1,0) + T_1) \cup ((1,0) + T_3).$$

Since $T = T_1 \cup T_2 \cup T_3$ satisfies $AT = T \cup ((1,0) + T)$, the equation for the "twin dragon," $\mathcal{Q}$ is a $\mathbb{Z}^2$-tiling set.

More generally one can choose the digits so that for each j, $\bigcup_{i=1}^{M} \Gamma_{ij} = \Gamma$ is a fixed set congruent to $\mathbb{Z}^2/A\mathbb{Z}^2$. Computing gives

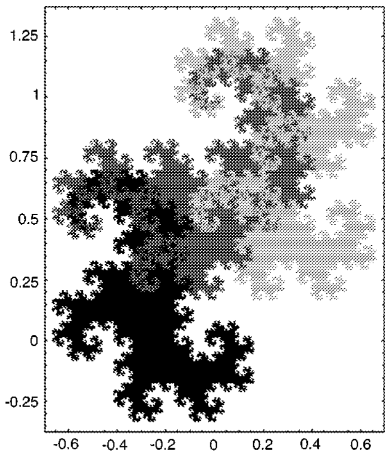

FIGURE 3

$$A\left(\bigcup_{i=1}^{M} T_i\right) = \bigcup_{i=1}^{M}\left(\bigcup_{j=1}^{M} \Gamma_{ij} + T_j\right) = \bigcup_{j=1}^{M}\left(\bigcup_{i=1}^{M} \Gamma_{ij} + T_j\right)$$
$$= \bigcup_{j=1}^{M} (\Gamma + T_j) = \Gamma + \left(\bigcup_{j=1}^{M} T_j\right).$$

With $T = \bigcup_{i=1}^{M} T_i$ we have $AT = \Gamma + T$ and the tiling theorem in [3] implies that the solution T is a self-affine tile for $\mathbb{R}^d$. Still we do not know whether (1) or (5) are satisfied. In fact, there are examples, cf. Examples 3 and 7 with $a = 1$, $b = 1$, $c = 0$, for which they are not. It would be interesting to know additional conditions necessary to guarantee that these examples work.

EXAMPLE 9. Again

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

and consider

$$\begin{aligned} AT_1 &= T_1 \cup ((1,0) + T_1) \cup ((1,1) + T_1) \cup ((0,-1) + T_2), \\ AT_2 &= ((2,3) + T_1) \cup ((1,1) + T_2) \cup ((2,2) + T_2) \cup ((1,2) + T_2). \end{aligned} \tag{16}$$

Q_1 and Q_2 are triangles with vertices $(0,0)$, $(1,0)$, $(1,1)$ and $(1,1)$, $(1,2)$, $(2,2)$, respectively, and $\mathcal{Q}$ is a self-affine $\mathbb{Z}^2$-tiling set. This example can be derived from (15) in much the same way as the examples in (9) are constructed from lattice tilings with one tile, i.e., the prototiles have been translated by elements of the lattice Λ. We leave the formalities to the reader.

EXAMPLE 10. Let

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

and

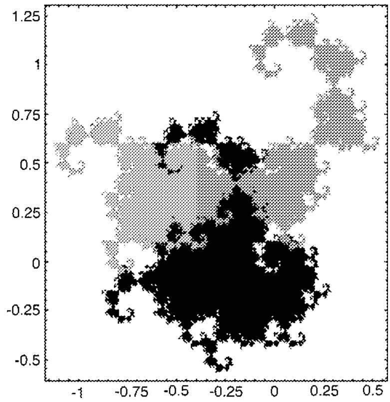

FIGURE 4

$$\begin{aligned} AT_1 &= T_1 \cup T_2, \\ AT_2 &= ((0,1) + T_1) \cup ((1,0) + T_2). \end{aligned} \tag{17}$$

The compact solution is a self-affine $\mathbb{Z}^2$-tiling set, as will be discussed in the next section. This tiling resembles the well-known twin dragon but cannot, to our knowledge, be obtained from it by any simple maneuver. See Fig. 4.

EXAMPLE 11. Figure 5 shows the maximal solution of

$$\begin{aligned} AT_1 &= T_1 \cup ((0,1) + T_1) \cup T_2, \\ AT_2 &= ((1,0) + T_1) \cup ((1,0) + T_2) \cup ((0,1) + T_2) \end{aligned}$$

with respect to the dilation matrix

$$A = \begin{pmatrix} 1 & 1 \\ -2 & 1 \end{pmatrix}.$$

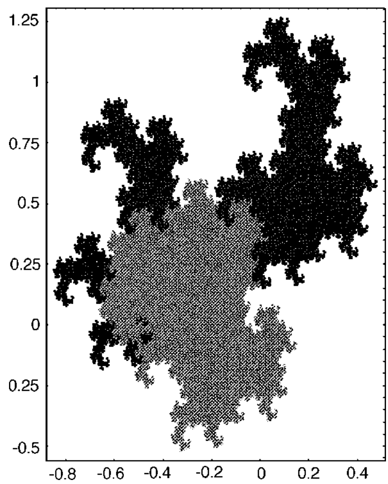

FIGURE 5

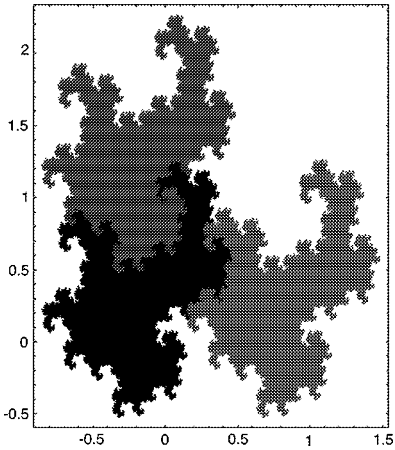

FIGURE 6

Note that one tile Q_1 is connected, whereas the other tile Q_2 is not. The tile $Q = Q_1 \cup Q_2$ is a solution of $AQ = Q \cup ((1,0) + Q) \cup ((0,1) + Q)$. Q and its translates by $(0,1)$ and $(1,0)$ are depicted in Fig. 6.

EXAMPLE 12. Figure 7 shows a bizarre solution of

$$AT_1 = T_1 \cup ((3,0) + T_1) \cup ((1,1) + T_1) \cup ((0,-1) + T_2),$$
$$AT_2 = ((2,3) + T_1) \cup ((1,1) + T_2) \cup ((2,2) + T_2) \cup ((1,2) + T_2)$$

with

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

It seems that $\mathcal{Q}$ tiles, but we do not have a formal proof for this.

2.4. *Details*

In this section we prove that the self-affine collections in Examples 7 and 10 are $\mathbb{Z}^d$-tiling sets. We shall use a direct elementary method that is based on the geometric interpretation

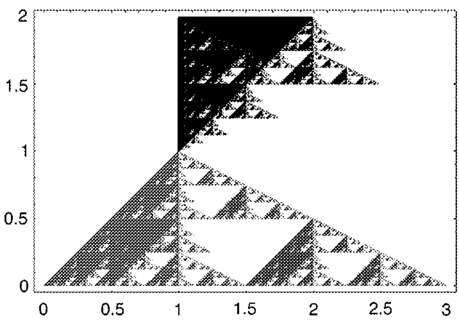

FIGURE 7

of the transfer operator in [12]. We hope that in view of these calculations the reader will appreciate the more systematic approach by means of Fourier analysis and the transfer operator which is the subject of Part II.

Suppose that $\mathcal{T}$ is an (A, Γ)-self-affine collection, $|\det A| = q$, and that $\mathbb{Z}^d + \bigcup_{i=1}^{M} T_i \equiv \mathbb{R}^d$. Define the non-negative numbers

$$a_{ij}(k) = |T_i \cap (k + T_j)| \qquad \text{for } i = 1, \dots, M, \text{ and } k \in \mathbb{Z}^d. \tag{18}$$

To prove that the self-affine collection $\mathcal{T}$ is a $\mathbb{Z}^d$-tiling, it is sufficient to show that

$$a_{ij}(k) = 0 \qquad \text{if } i \neq j \text{ or } k \neq 0. \tag{19}$$

The following two identities which follow from the self-similarity (4) will prove useful,

$$a_{ij}(k) = a_{ji}(-k) \tag{20}$$

and

$$q a_{ij}(k) = \sum_{l=1}^{M} \sum_{m=1}^{M} \sum_{\alpha \in \Gamma_{il}} \sum_{\beta \in \Gamma_{jm}} a_{lm}(\beta - \alpha + Ak), \tag{21}$$

where a sum over an empty set equals 0 by definition. Equation (20) is obvious, (21) follows by computation from

$$\begin{aligned} q|T_i \cap (k + T_j)| &= |AT_i \cap (Ak + AT_j)| \\ &= \left| \left(\bigcup_{l=1}^{M} \Gamma_{il} + T_l \right) \cap \left(\bigcup_{m=1}^{M} Ak + \Gamma_{jm} + T_m \right) \right| \\ &= \sum_{l=1}^{M} \sum_{m=1}^{M} \sum_{\alpha \in \Gamma_{il}} \sum_{\beta \in \Gamma_{jm}} |T_l \cap (\beta - \alpha + Ak + T_m)| \\ &= \sum_{l=1}^{M} \sum_{m=1}^{M} \sum_{\alpha \in \Gamma_{il}} \sum_{\beta \in \Gamma_{jm}} a_{lm}(\beta - \alpha + Ak). \end{aligned}$$

In Example 7, $2T_1 = T_1 \cup T_2$, $2T_2 = (1 + T_1) \cup (3 + T_2)$; (21) takes the form of the following four equations:

$$2a_{11}(k) = a_{11}(2k) + a_{22}(2k) + a_{12}(2k) + a_{21}(2k), \tag{22}$$

$$2a_{22}(k) = a_{11}(2k) + a_{22}(2k) + a_{12}(2k + 2) + a_{21}(2k - 2), \tag{23}$$

$$2a_{12}(k) = a_{11}(2k + 1) + a_{22}(2k + 3) + a_{12}(2k + 3) + a_{21}(2k + 1), \tag{24}$$

$$2a_{21}(k) = a_{11}(2k - 1) + a_{22}(2k - 3) + a_{12}(2k - 1) + a_{21}(2k - 3). \tag{25}$$

Referring to Theorem 2, at least one, and therefore both prototiles Q_i of the maximal compact solution have positive measure and cover $\mathbb{R}^d$. In our notation that is $a_{11}(0) \neq 0$ and $a_{22}(0) \neq 0$. Employing (20) for $k = 0$, the first two equations become $a_{11}(0) = a_{22}(0) + 2a_{12}(0)$ and $a_{22}(0) = a_{11}(0) + 2a_{12}(2)$. From this we conclude that $a_{12}(0) = a_{21}(0) = 0$ and $a_{11}(0) = a_{22}(0)$.

Since all entries are nonnegative, we see from (24) with $k=0$ that $a_{11}(1)=a_{22}(3)=a_{12}(3)=a_{21}(1)=0$.

In (22) $k=1$ gives $a_{11}(2)=a_{22}(2)=a_{12}(2)=a_{21}(2)=0$. Also, $k=1$ in (25) yields $a_{22}(1)=a_{12}(1)=0$. So far we have $a_{ij}(k)=0$ for $i,j=1,2$ and $k=\pm 1,\pm 2$.

To deal with $|k|\geq 3$, we check the size of the prototiles. Set $\alpha=\min T_1\cup T_2$ and $\beta=\max T_1\cup T_2$. Then from (14) we conclude that $\alpha\geq 0$ and $\beta\leq 3$. This implies that $a_{ij}(k)=0$ for $|k|\geq 3$ and by (19) $\mathcal{Q}$ is a self-affine $\mathbb{Z}$-tiling set.

In Example 10, where

$$A=\begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix},\qquad AT_1=T_1\cup T_2,\qquad \text{and}\qquad AT_2=((0,1)+T_1)\cup((1,0)+T_2),$$

(21) takes the form of the following four equations:

$$\begin{aligned}
2a_{11}(k)&=a_{11}(Ak)+a_{22}(Ak)+a_{12}(Ak)+a_{21}(Ak),\\
2a_{22}(k)&=a_{11}(Ak)+a_{22}(Ak)+a_{12}(Ak+(1,-1))+a_{21}(Ak+(-1,1)),\\
2a_{12}(k)&=a_{11}(Ak+(0,1))+a_{22}(Ak+(1,0))+a_{12}(Ak+(1,0))+a_{21}(Ak+(0,1)),\\
2a_{21}(k)&=a_{11}(Ak-(0,1))+a_{22}(Ak-(1,0))+a_{12}(Ak-(1,0))+a_{21}(Ak-(0,1)).
\end{aligned}$$

Let $\delta_i=\max\{\|x\|: x\in T_i\}$ be the (Euclidean) extension of T_i, then (17) implies $\sqrt{2}\delta_1=\max(\delta_1,\delta_2)$ and $\sqrt{2}\delta_2\leq\max(\delta_1+1,\delta_2+2)$, from which $\delta_1\leq 1+2^{-1/2}$ and $\delta_2\leq\sqrt{2}+1$. Therefore $T_1\cup T_2$ is contained in a disk of radius $\sqrt{2}+1$ and consequently $a_{ij}(k)=0$ for $|k|\geq 5$.

In order to show that $T_1\cup T_2$ $\mathbb{Z}^2$-tiles $\mathbb{R}^2$, it is sufficient to show (19) for $|k|\leq 4$. This is possible and an exercise in patience, but we shall take a more experimental approach and use the evidence from Fig. 4 that the maximum extension of both tiles in the x- and y-direction is 2. Using this geometric fact we have to verify (19) for $k=(0,0),(\pm 1,0),(0,\pm 1),\pm(1,1),\pm(1,-1)$.

As above, by recourse to Theorem 2 both prototiles Q_i of the maximal compact solution have positive measure, that is $a_{11}(0,0)\neq 0$ and $a_{22}(0,0)\neq 0$. Employing (20) for $k=(0,0)$, the first two equations become $a_{11}(0,0)=a_{22}(0,0)+2a_{12}(0,0)$ and $a_{22}(0,0)=a_{11}(0,0)+2a_{12}(0,0)$. From this we conclude that $a_{12}(0,0)=a_{21}(0,0)=0$ and $|T_1|=a_{11}(0,0)=a_{22}(0,0)=|T_2|$.

Since all entries are nonnegative, we see from the equation for $2a_{12}(0)$ that $a_{11}(0,1)=a_{22}(1,0)=a_{12}(1,0)=a_{21}(0,1)=0$ and by symmetry (20) $a_{11}(0,-1)=a_{22}(-1,0)=a_{21}(-1,0)=a_{12}(0,-1)=0$.

Using $a_{11}(0,1)=a_{11}(1,0)=0$ and (20), we obtain $a_{ij}(1,-1)=0$ and $a_{ij}(1,1)=0$ for $i,j=1,2$. Next using $a_{12}(1,0)=0$ implies $a_{11}(1,0)=a_{21}(1,0)=0$, whereas $a_{21}(0,1)=0$ gives $a_{22}(0,1)=a_{12}(0,1)=0$. Together with the symmetry (20) this implies (19) for all $k=(k_1,k_2)$, $|k_i|\leq 1$ and that $T_1\cup T_2$ is a $\mathbb{Z}^2$-tiling.

3. EXISTENCE OF SELF-AFFINE COLLECTIONS

Fix a dilation matrix A, a multiplicity M, and arbitrary finite digit sets $\Gamma_{ij}\subseteq\mathbb{Z}^d$, $i,j=1,\ldots,M$. Set $\bigcup_{i,j=1}^{M}\Gamma_{ij}=\Gamma$ and write $\{1,\ldots,M\}=\mathbf{S}$. In this section we determine all solutions to the dilation equation (4). As in the one tile case, the solutions are described by digit expansions, the digits of which are taken from Γ. Unlike the one tile case, not all sequences of digits may occur as digit expansions. The difference is somewhat like that between the full sequence space on **S** and a subspace determined by a subshift of finite type [14].

To motivate the following definitions, observe that the dilation equations can be used to rewrite each prototile of a solution $\{T_i\}$ in the form

$$T_i = A^{-1}\left(\bigcup_{j=1}^{M} \Gamma_{ij} + T_j\right).$$

Applying this operation again to each prototile T_j gives

$$T_i = \bigcup_{j=1}^{M} A^{-1}\Gamma_{ij} + A^{-1}\left(\bigcup_{k=1}^{M} A^{-1}\Gamma_{jk} + A^{-1}T_k\right) = \bigcup_{j=1}^{M}\bigcup_{k=1}^{M} A^{-1}\Gamma_{ij} + A^{-2}\Gamma_{jk} + A^{-2}T_k.$$

Iterating this procedure n times shows that $x \in T_i$ if and only if for some $k \in \mathbf{S}$

$$x \in \sum_{j=1}^{n} A^{-j}\epsilon_j + A^{-n}T_k, \tag{26}$$

where $\epsilon_j \in \Gamma_{\rho_j\rho_{j+1}}$ for $j = 1, \ldots, n$ and $\rho_1 = i$. Since A^{-1} is contractive, all of the sets $A^{-n}T_k$ lie in a small ball for n large, and consequently x is essentially determined by the ϵ_j. An elaboration of this argument yields

PROPOSITION 1. *Set*

$$Q_i = \left\{x \in \mathbb{R}^d : x = \sum_{k=1}^{\infty} A^{-k}\epsilon_k,\ \epsilon_k \in \Gamma_{\rho_k\rho_{k+1}} \neq \emptyset \textit{ for some } \rho_k \in S \textit{ and } \rho_1 = i\right\}.$$

$\mathcal{Q} = \{Q_i\}_{i=1}^{M}$ *is the unique solution to* (4) *for which all prototiles are nonempty and compact.*

As was observed in Examples 2 and 3, there are solutions for which some, but not all, of the sets are empty. Let $\mathcal{T} = \{T_i\}$ be any solution of (4) with T_i empty or compact. Define $N = \{i \in \{1, \ldots, M\} : T_i \neq \emptyset\}$. Removing the sets $\{T_i : i \notin N\}$ in (4), we obtain another dilation equation for which the unique solution with all prototiles being nonempty and compact is defined as $\{Q_i, i \in N\}$ instead of $i \in \mathbf{S}$.

This remark leads to the following.

DEFINITION. A nonempty subset $N \subseteq \mathbf{S}$ is said to be (A, Γ)-*closed*, if $j \in N$ and $i \notin N$ imply $\Gamma_{ij} = \emptyset$.

The definition is similar to the one from the theory of Markoff chains and will appear less coincidental in the next section. In this context closed sets are used to catalogue the nonempty prototiles in a solution to (4).

For this we define the following sets of sequences in $\mathbf{S}^{\mathbb{N}}$ and $\Gamma^{\mathbb{N}}$. Given $N \subseteq \mathbf{S}$, let

$$R_i^N = \left\{(\rho_k)_{k=1}^{\infty} \in \mathbf{S}^{\mathbb{N}} \mid \rho_1 = i,\ \rho_k \in N,\ \Gamma_{\rho_k\rho_{k+1}} \neq \emptyset,\ \forall k\right\}$$

and let

$$\Omega_i^N = \left\{(\epsilon_k)_{k=1}^{\infty} \in \Gamma^{\mathbb{N}} \mid \epsilon_k \in \Gamma_{\rho_k \rho_{k+1}} \text{ for some } \rho \in R_i^N\right\}$$

be the set of all "paths" in N starting at i. Then we define the sets

$$Q_i^N = \left\{ x \mid x = \sum_{j=1}^{\infty} A^{-j} \epsilon_j,\ (\epsilon_j) \in \Omega_i^N \right\}, \tag{27}$$

if $i \in N$ and $Q_i^N = \emptyset$ when $i \notin N$. Set $\mathcal{Q}^N = \{Q_i^N\}$. Obviously $\mathbf{S}$ is closed, $Q_i^{\mathbf{S}} = Q_i$, and $\mathcal{Q}^{\mathbf{S}} = \mathcal{Q}$. It follows easily from the definition that if N is (A,Γ)-closed then $Q_i^N \neq \emptyset$ if and only if $i \in N$.

Since the $\{Q_i\}$ may overlap, these are not always self-affine collections. See Example 5.

THEOREM 1. (a) *For any (A,Γ)-closed set N, $\mathcal{Q}^N$ satisfies the dilation equation* (4). *The set Q_i^N is compact if $i \in N$ and empty otherwise.*

(b) *If $\mathcal{T} = \{T_i\}$ is a collection of empty or compact sets that satisfies* (4), *then $\mathcal{T} = \mathcal{Q}^N$ for some (A,Γ)-closed set N.*

(c) *If N_1 and N_2 are (A,Γ)-closed sets and $\mathcal{Q}^{N_1} = \mathcal{Q}^{N_2}$, then $N_1 = N_2$. Also, if $N_1 \subseteq N_2$ then $\mathcal{Q}_i^{N_1} \subseteq \mathcal{Q}_i^{N_2}$ for $i = 1, \ldots, M$.*

In Example 2 the various (A,Γ)-closed subsets of $\mathbf{S}$ are $\{1\}, \{2\}, \{1,2\}, \{1,3\}, \{1,2,3\}$, and $\mathbf{S}$. They correspond to the solutions described in Section 2, in the respective orders.

We assume from now on that an (A,Γ)- closed set $N \subseteq \mathbf{S}$ is given.

Choose a norm on $\mathbb{R}^d$ for which A^{-1} is a contraction; that is, $\|A^{-1}x\| \leq \lambda\|x\|$ for all $x \in \mathbb{R}^d$ and for some $\lambda < 1$. Define ω: $\Gamma^{\mathbb{N}} \to \mathbb{R}^d$ by $\omega(\epsilon) = \sum_{k=1}^{\infty} A^{-k}\epsilon_k$. With the product topology on $\Gamma^{\mathbb{N}}$ the map ω is continuous. Furthermore, for $i \in N$, $\Omega_i^N \subseteq \Gamma^{\mathbb{N}}$ is closed and hence compact, and therefore the set $Q_i^N = \omega(\Omega_i^N)$ is also compact.

We introduce further definitions and prove a lemma before proceeding with the proof of Theorem 1. The Euclidean metric on $\mathbb{R}^d$ is written $d(\cdot,\cdot)$. Let $H_\emptyset(\mathbb{R}^d)$ be the set containing the compact subsets of $\mathbb{R}^d$ and the empty set with the (modified) Hausdorff metric

$$D(X,Y) = \begin{cases} \max\{\sup_{y\in Y} d(X,y), \sup_{x\in X} d(x,Y)\} & \text{for } X, Y \neq \emptyset \\ \sup_{y\in Y} d(0,y) + 10 & \text{for } X = \emptyset. \end{cases}$$

Let $H_\emptyset^M(\mathbb{R}^d)$ be the M-fold Cartesian product of $H_\emptyset(\mathbb{R}^d)$ with the product metric.

Define the functions φ_i: $H_\emptyset^M \to H_\emptyset$ by

$$\varphi_i(Z_1,\ldots,Z_M) = A^{-1}\left(\bigcup_{j=1}^{M} \Gamma_{ij} + Z_j\right) \tag{28}$$

and let φ: $H_\emptyset^M \to H_\emptyset^M$ be the product $\varphi = (\varphi_1,\ldots,\varphi_M)$. φ^n will denote the n-fold iterate of φ and we write $\varphi^n(Z_1,\ldots,Z_M) = (Z_1^n,\ldots,Z_M^n) = Z^n$.

LEMMA 1. *Given any $Z = (Z_1,\ldots,Z_M) \in H_\emptyset^M$, suppose that $N = \{i \mid Z_i \neq \emptyset\}$ is (A,Γ)-closed. Then*

(a) *$\varphi^n(Z_1,\ldots,Z_M)$ converges to $(Q_1^N,\ldots,Q_M^N)$ in the Hausdorff metric.*

(b) *If $Y \in H_\emptyset^M$ satisfies $Z_i = Y_i$ whenever $Y_i \neq \emptyset$, then for each $n \in \mathbb{N}$ and $i = 1,\ldots,M$, $\varphi_i^n(Z) \supseteq \varphi_i^n(Y)$.*

Proof. (a) By (27) the sets Q_i^N can be written in a concise form as follows:

$$Q_i^N = \bigcup_{\rho \in R_i^N} \left(\sum_{k=1}^{\infty} A^{-k} \Gamma_{\rho_k \rho_{k+1}} \right). \tag{29}$$

If $i \notin N$ then $R_i^N = \emptyset$ and therefore $Q_i^N = \emptyset$ as required. Furthermore, if $i \notin N$ then it is clear from the definitions that $Z_i^n = \emptyset$ for all $n \in \mathbb{N}$ and the convergence is assured in those components. We argue by induction that for all $n \in \mathbb{N}$ and $i \in N$,

$$Z_i^n = \bigcup_{\rho \in R_i^N} \left(A^{-n} Z_{\rho_{n+1}} + \sum_{k=1}^{n} A^{-k} \Gamma_{\rho_k \rho_{k+1}} \right). \tag{30}$$

Before proceeding with the induction observe that the characterization of Z_i^n given by (30) can be applied to prove convergence. Given $\alpha > 0$ choose $n_0 > 0$ so that for $i = 1, \dots, M$, and $n > n_0$, $A^{-n} Z_i \subseteq B(0, \alpha/2)$ and $\lambda^n \|\omega(\epsilon)\| < \alpha/2$ for all $\epsilon \in \Omega_i^N$. Then for $i \in N$ and $n > n_0$

$$\begin{aligned} D\big(Z_i^n, Q_i^N\big) &= D\left(\bigcup_{\rho \in R_i^N} (A^{-n} Z_{\rho_{n+1}} + \sum_{k=1}^{n} A^{-k} \Gamma_{\rho_k \rho_{k+1}}), \sum_{k=1}^{\infty} A^{-k} \Gamma_{\rho_k \rho_{k+1}} \right) \\ &\le \sup_{\rho \in R_i^N} D\left(A^{-n} Z_{\rho_{n+1}}, \sum_{k=n+1}^{\infty} A^{-k} \Gamma_{\rho_k \rho_{k+1}} \right) < \alpha. \end{aligned} \tag{31}$$

Therefore $\{Z^n\}$ converges to $(Q_1^N, \dots, Q_M^N)$ as claimed.

We now turn to the proof of (30). It is obvious for Z_i^0. If it holds for each Z_i^n, then by definition

$$\begin{aligned} Z_i^{n+1} &= A^{-1} \left(\bigcup_{j=1}^{M} \Gamma_{ij} + Z_j^n \right) \\ &= \bigcup_{j=1}^{M} \left(\bigcup_{\rho \in R_j^N} A^{-(n+1)} Z_{\rho_{n+1}} + \sum_{k=1}^{n} A^{-(k+1)} \Gamma_{\rho_k \rho_{k+1}} + A^{-1} \Gamma_{ij} \right). \end{aligned} \tag{32}$$

Given $\rho \in R_j^N$ with $\Gamma_{ij} \neq \emptyset$ define $\beta = \{\beta_k\}_{k=1}^{\infty}$ by $\beta_{k+1} = \rho_k$ for $k \in \mathbb{N}$ and $\beta_1 = i$. Then by definition $i \in N$ and $\beta \in R_i^N$. Conversely, if $\beta \in R_i^N$ and ρ is defined as above, then $\rho \in R_j^N$ with $j = \beta_2$. Thus (32) can be continued

$$\begin{aligned} Z_i^{n+1} &= \bigcup_{\{j \mid \Gamma_{ij} \neq \emptyset\}} \left(\bigcup_{\rho \in R_j^N} (A^{-(n+1)} Z_{\rho_{n+1}} + \sum_{k=1}^{n} A^{-(k+1)} \Gamma_{\rho_k \rho_{k+1}} + A^{-1} \Gamma_{ij} \right) \\ &= \bigcup_{\beta \in R_i^N} \left(A^{-(n+1)} Z_{\beta_{n+2}} + \sum_{k=1}^{n+1} A^{-k} \Gamma_{\beta_k \beta_{k+1}} \right) \end{aligned}$$

which completes the induction.

(b) follows easily from the definition:

$$\begin{aligned}\varphi_i(Y_1,\ldots,Y_M) &= A^{-1}\left(\bigcup_{j=1}^{M}\Gamma_{ij}+Y_j\right) \subseteq A^{-1}\left(\bigcup_{j=1}^{M}\Gamma_{ij}+Z_j\right)\\ &= \varphi_i(Z_1,\ldots,Z_M).\end{aligned}$$

■

Proof of Theorem 1. We have already seen that the sets Q_i^N are either compact or empty. Since

$$\begin{aligned}AQ_i^N &= \bigcup_{\rho\in R_i^N}\left(\sum_{k=1}^{\infty}A^{-k+1}\Gamma_{\rho_k\rho_{k+1}}\right) = \bigcup_{\rho\in R_i^N}\left(\Gamma_{\rho_1\rho_2}+\sum_{k=1}^{\infty}A^{-k+1}\Gamma_{\rho_{k+1}\rho_{k+2}}\right)\\ &= \bigcup_{\substack{1\le j\le M\\ j\in N}}\left(\Gamma_{ij}+\bigcup_{\rho\in R_j^N}\left(\sum_{k=1}^{\infty}A^{-k+1}\Gamma_{\rho_k\rho_{k+1}}\right)\right)\\ &= \bigcup_{\substack{1\le j\le M\\ j\in N}}\left(\Gamma_{ij}+Q_j^N\right) = \bigcup_{j=1}^{M}\left(\Gamma_{ij}+Q_j^N\right)\end{aligned}$$

they satisfy the dilation equations.

Let $\mathcal{T}$ be a collection of compact sets that satisfy (4). Define $N=\{i \mid T_i\neq\emptyset\}$ and suppose that $i\in N$ and $\Gamma_{ni}\neq\emptyset$. Since $T_i\neq\emptyset$

$$T_n = A^{-1}\left(\bigcup_{j=1}^{M}(\Gamma_{nj}+T_j)\right)\neq\emptyset \tag{33}$$

and therefore N is (A,Γ)-closed. From the previous lemma $\varphi^n(T_1,\ldots,T_M)$ converges to $(Q_1^N,\ldots,Q_M^N)$ in the Hausdorff metric, but since $(T_1,\ldots,T_M)$ is a fixed point of φ we have $\mathcal{T}=\mathcal{Q}^N$.

If N_1 and N_2 are distinct (A,Γ)-closed sets then without loss of generality there is an $n\in N_1\setminus N_2$. It is immediate from the definition that $Q_n^{N_1}\neq\emptyset$ while $Q_n^{N_2}=\emptyset$. The inclusion $Q_i^{N_1}\subseteq Q_i^{N_2}$ for $N_1\subseteq N_2$ follows directly from the definition (27). ■

Remark 1. Theorem 1 can be seen as giving a classification of the fixed points of the operator φ defined in (28). In contrast to the single tile case, in which φ is an iterated function system with a unique attracting fixed point [1], we may have a finite number of fixed points and periodic cycles, each having an associated basin of attraction.

In Example 4 from Section 2 the set $\{(Q_1,\emptyset),(\emptyset,Q_2)\}$ is a period two cycle and any Z of the form $(Z_1,\emptyset)$ or $(\emptyset,Z_2)$ will accumulate at the cycle. The system also has the fixed point (Q_1,Q_2) which is the limit of any Z with both components nonempty. Along with the trivial fixed point $(\emptyset,\emptyset)$ the above completely describes the limiting behavior of iterating φ in this example. We will be better equipped to describe this phenomenon later in the section.

Following [16] we call $\{\Gamma_{ij},\ i,j=1,\ldots,M\}=\Gamma$ a *standard digit set* if for each $j=1,\ldots,M$, $\Gamma_j=\bigcup_{i=1}^{M}\Gamma_{ij}$ is a complete set of coset representatives for the group $\mathbb{Z}^d/A\mathbb{Z}^d$. As we will see, it is precisely the standard digit sets that have relevance to wavelet theory.

THEOREM 2. *Suppose that* Γ *is a standard digit set and* N *is an* (A, Γ)*-closed set. Then* $\mathcal{Q}^N$ *is a self-affine collection and* $\mathbb{Z}^d + (\bigcup_{i=1}^M \mathcal{Q}_i^N) = \mathbb{R}^d$.

Observe that these conditions are not necessary. In Example 6 Γ is not a standard digit set, but $\mathcal{Q}$ is nevertheless a self-affine collection. It will follow from Theorem 3, that for such examples $\mathcal{Q}$ is not a $\mathbb{Z}^d$-tiling set.

If N_1 and N_2 are two nonempty, disjoint (A, Γ)-closed sets, then we infer from the theorem that for $i = 1, 2$, the $\mathbb{Z}^d$-translates of the set $\bigcup_{j=1}^M \mathcal{Q}_j^{N_i}$ cover $\mathbb{R}^d$. Consequently we have:

COROLLARY 1. *Suppose that* Γ *is a standard digit set and* N_1 *and* N_2 *are disjoint* (A, Γ)*-closed sets. Then* $\mathcal{Q}$ *is not a* $\mathbb{Z}^d$*-tiling set.*

However, as is shown in Example 6, $\mathcal{Q}$ might tile with a coarser lattice.

It is necessary to introduce some new ideas in the proof of Theorem 2. Let C be the $M \times M$ matrix, called the *counting matrix*, with entries $c_{ij} = \#\Gamma_{ij}$ and let $|\det A| = q$. The importance of C lies in the following lemma.

LEMMA 2. (a) *For any dilation* A*, digit set* Γ*, and solution* $\mathcal{T}$ *to* (4)

$$q|T_i| \le \sum_{j=1}^M c_{ij} |T_j| \qquad \text{for } i = 1, \ldots, M. \tag{34}$$

Equality holds for all i*, if and only if the disjointness property* (5) *is also satisfied and* $\mathcal{T}$ *is a self-affine collection. In particular for any self-affine collection* $\mathcal{T}$ *the column vector* $(|T_1|, \ldots, |T_M|)^t$ *is an eigenvector of* C *to the eigenvalue* q.

(b) *If* Γ *is a standard digit set, then for each* $j = 1, \ldots, M$,

$$\sum_{i=1}^M c_{ij} = q; \tag{35}$$

in other words, $(1/q)C^t$ *is a stochastic matrix.*

Moreover, any solution $\mathcal{T}$ *to* (4) *is automatically a self-affine collection.*

Proof. (a) For each $i = 1, \ldots, M$,

$$q|T_i| = |AT_i| = \left| \bigcup_{j=1}^M \Gamma_{ij} + T_j \right| \le \sum_{j=1}^M c_{ij} |T_j|. \tag{36}$$

Equality holds in the ith equation if and only if the sets $\gamma + T_j$ are essentially disjoint for $\gamma \in \bigcup_{j=1}^M \Gamma_{ij}$. Thus, equality holds for all i if and only if $\mathcal{T}$ is a self-affine collection.

(b) The sum $\sum_{i=1}^M c_{ij}$ represents the number of elements in the digit set $\Gamma_j = \bigcup_{i=1}^M \Gamma_{ij}$ and thus it equals $\#\Gamma_j = \#\mathbb{Z}^d / A\mathbb{Z}^d = |\det A|$.

This gives

$$\sum_{i=1}^M \sum_{j=1}^M c_{ij} |T_j| = \sum_{j=1}^M q |T_j|. \tag{37}$$

This means that in (36) equality must hold for all i and thus $\mathcal{T}$ is a self-affine collection by (a). ■

Proof of Theorem 2. It remains to be proven that

$$\mathbb{Z}^d + \left(\bigcup_{i=1}^{M} Q_i^N \right) = \mathbb{R}^d. \tag{38}$$

We may choose compact sets Z_i, $i = 1, \ldots, M$, in $\mathbb{R}^d$ with the following properties: $Z_i = \emptyset$ if and only if $i \notin N$, $Z_i \cap Z_j \cong \emptyset$ for $i \neq j$, and $\bigcup_{i=1}^{M} Z_i = [0, 1]^d$.

Recall that we write $\varphi^n(Z_1, \ldots, Z_M) = (Z_1^n, \ldots, Z_M^n)$, where φ was defined following (28). We will show by induction that the following statements are true for $n \geq 0$:

$$Z_i^n = \emptyset \qquad \text{if and only if} \qquad i \notin N; \tag{39}$$

$$Z_i^n \cap Z_j^n \cong \emptyset \quad \text{for } i \neq j; \tag{40}$$

$$\bigcup_{i=1}^{M} Z_i^n \qquad \text{is a } \mathbb{Z}^d\text{-tile.} \tag{41}$$

Since by Lemma 1 Z_i^n converges to Q_i^N in the Hausdorff metric, (38) follows from (41) as in [10].

The claim is argued by induction. It is certainly true for $Z^0 = Z$. Suppose (39)–(41) hold for Z^n. Combining (40) and (41) we infer that for $i \neq j$ or $l \neq k$, $(k + Z_i^n) \cap (l + Z_j^n) \cong \emptyset$. Since each Γ_j is a set of coset representatives, it follows that for $i \neq j$ or $l \neq k$, $\gamma_i \in \Gamma_i$ and $\gamma_j \in \Gamma_j$

$$(\gamma_i + Ak + Z_i^n) \cap (\gamma_j + Al + Z_j^n) \cong \emptyset$$

and

$$\bigcup_{j=1}^{M} \bigcup_{\gamma \in \Gamma_j} (\gamma + Z_j^n) = \bigcup_{i,j=1}^{M} \left(\bigcup_{\gamma \in \Gamma_{ij}} (\gamma + Z_j^n) \right)$$

is an $A\mathbb{Z}^d$-tile. Consequently, the sets $A^{-1}(\gamma + Z_j^n)$ are essentially disjoint for $\gamma \in \Gamma_{ij}$ and $1 \leq i, j \leq M$, and their union is a $\mathbb{Z}^d$-tile. Finally, since

$$Z_i^{n+1} = A^{-1} \left(\bigcup_{j=1}^{M} \left(\bigcup_{\gamma \in \Gamma_{ij}} (\gamma + Z_j^n) \right) \right)$$

we conclude that $Z_i^{n+1} \cap Z_j^{n+1} \cong \emptyset$ for $i \neq j$ and $\bigcup_{i=1}^{M} Z_i^{n+1}$ is a $\mathbb{Z}^d$-tile. ■

A different proof of Theorem 2 based on Fourier analytic methods will be given in Part II.

4. NECESSARY CONDITIONS AND MARKOFF CHAINS

We shall now discuss a number of necessary conditions for a dilation A and a digit set Γ to determine a self-affine tiling $\mathcal{Q}$. Where not stated otherwise we take $\mathcal{Q}$ to be a self-affine $\mathbb{Z}^d$-tiling set. The setting will usually be further simplified by the assumption that for each $i = 1, \ldots, M$, the set Q_i has nonzero Lebesgue measure.

The importance of standard digit sets becomes apparent in the following theorem, which generalizes a similar result in the one-tile case [10].

THEOREM 3. *If $\mathcal{Q}$ is a $\mathbb{Z}^d$-tiling set and if Q_i has positive Lebesgue measure for each $i = 1, \ldots, M$, then Γ is a standard digit set.*

Proof. If Γ is not a standard digit set then, for some j, Γ_j is not a set of distinct coset representatives for $\mathbb{Z}^d/A\mathbb{Z}^d$. Then either there are $k_1 \in \Gamma_{i_1 j}$ and $k_2 = k_1 - A\ell \in \Gamma_{i_2 j}$, $\ell \in \mathbb{Z}^d$, representing the same coset, or some coset is not represented in $\Gamma_j = \bigcup_{i=1}^M \Gamma_{ij}$. In the former case by the self-similarity, $k_1 + T_j \subseteq AT_{i_1}$ and $k_2 + T_j \subseteq AT_{i_2}$, which implies $k_1 + T_j \subseteq A(T_{i_2} + \ell)$. Since by assumption all T_j have positive measure, the inequality

$$|T_{i_1} \cap (\ell + T_{i_2})| \geq |A^{-1}(k_1 + T_j)| > 0$$

furnishes a contradiction to the tiling property (1) and (3). Thus Γ_j consists of distinct representatives of $\mathbb{Z}^d/A\mathbb{Z}^d$. In particular, $\sum_{i=1}^M c_{ij} \leq q$.

Since the disjoint union $T = \bigcup_{i=1}^M T_i$ yields a $\mathbb{Z}^d$-tiling, T has measure 1 and Lemma 2 implies

$$\sum_{i=1} \sum_{j=1} c_{ij} |T_j| = q \sum_{i=1}^M |T_i| = q.$$

If $\sum_{i=1}^M c_{ij} < q$, then $\sum_{j=1} \sum_{i=1} c_{ij} |T_j| < q \sum |T_j| = q$ provides a contradiction. This means that $\bigcup_{i=1}^M \Gamma_{ij}$ is a complete set of representatives of $\mathbb{Z}^d/A\mathbb{Z}^d$. ■

At this point, it is necessary to introduce concepts from the theory of Markoff chains which can be found in the standard texts, e.g., [5]. The results to which we make explicit reference appear as Proposition 2. Let P denote the stochastic matrix $(1/q)C^t$. Then P is the matrix of transition probabilities for a Markoff chain with state space $\mathbf{S}$. An *invariant probability* ν on $\mathbf{S}$ is a right eigenvector $(\nu(1), \ldots, \nu(M))^T$ of P of eigenvalue 1 with $\sum_{i=1}^M \nu(i) = 1$.

A set $N \subseteq \mathbf{S}$ is closed, if $p_{jk} = 0$, whenever $j \in N$ and $k \notin N$ (see [5, remark on p. 384]). Since $p_{jk} \neq 0$, if and only if $\Gamma_{kj} \neq \emptyset$, closed sets in the sense of Markoff chains coincide with the (A, Γ)-closed sets defined in Section 3.

An *irreducible* set is a closed set $N \subseteq \mathbf{S}$ that contains no proper closed subsets. Let R denote the union of all irreducible subsets of $\mathbf{S}$ and call a state $x \in R$ *recurrent* (or *persistent*). The complement of R is the set I of *transient* states. Let $p_{ij}^{(n)}$ denote the ijth entry of the matrix P^n. A state j is *periodic* of period τ if $p_{jj}^{(n)} = 0$ unless $n = m\tau$ for $m \in \mathbb{N}$, and τ is the smallest such integer. A Markoff chain is called *aperiodic* if $\mathbf{S}$ contains no periodic states and *irreducible* if $\mathbf{S}$ is irreducible.

PROPOSITION 2. (a) *There exist disjoint irreducible sets $R_1, \ldots, R_k$, with $k > 0$ so that $\mathbf{S}$ can be partitioned as*

$$\mathbf{S} = R_1 \cup \cdots \cup R_k \cup I. \tag{42}$$

(b) *For each i, $1 \leq i \leq k$, there is a unique invariant probability ν_i with the property that $\nu_i(x) = 0$ for $x \notin R_i$, and $\nu_i(x) > 0$ for $x \in R_i$. Every invariant probability is a linear combination of the ν_i's.*

(c) *Suppose the Markoff chain is irreducible and some state x is periodic of period τ. Then every state is periodic of period τ and for the Markoff chain with the matrix of transition probabilities P^τ, $\mathbf{S}$ is partitioned into τ nonempty irreducible subsets.*

(d) *If a Markoff chain is irreducible and aperiodic then for all $i, j \in \mathbf{S}$*

$$\lim_{n\to\infty} p_{ij}^{(n)} = \nu(\{j\}) > 0. \tag{43}$$

where ν is the unique invariant probability on $S = R_1$.

The proposition has the following consequence.

THEOREM 4. *If Γ is a standard digit set and $\mathcal{Q}$ is a $\mathbb{Z}^d$-tiling set with $|Q_i| > 0$ for $1 \le i \le M$, then the associated Markoff chain is irreducible and aperiodic.*

Proof. By Corollary 1 only one irreducible set can appear in the decomposition (42). Set $p = \sum_{i=1}^M |Q_i|$. It follows from Lemma 2 that the column vector ν with $\nu(i) = p^{-1}|Q_i|$ is a probability. As a result of the hypothesis that all prototiles have positive measure, $\nu(i) > 0$ for all $i \in \mathbf{S}$. In light of Proposition 2(b) the set of transient states I is empty and therefore in (42) $\mathbf{S}$ consists of a single recurrent set.

Suppose that the chain is periodic. As in Example 4 it is possible to define the set of dilation equations

$$A^2 Q_i = \bigcup_{j=1}^M A\Gamma_{ij} + AQ_j = \bigcup_{j=1}^M A\Gamma_{ij} + \left(\bigcup_{k=1}^M \Gamma_{kj} + Q_k\right) = \bigcup_{k=1}^M \Gamma_{ik}^2 + Q_k \tag{44}$$

for some digit set Γ_{ik}^2. It is then possible to recursively define the dilation equations $A^n Q_i = \bigcup_{k=1}^M \Gamma_{ik}^n + Q_k$, which we call eqn($n$). For all $n \in \mathbb{N}$ the maximal solution $\mathcal{Q}^S$ of eqns(n) is then equal to the original solution $\mathcal{Q}$ of (4).

Let $C^{(n)}$ denote the counting matrix of the dilation equations eqn(n). Using the fact that Γ is a standard digit set, a computation shows that $C^{(n)} = C^n$. Consequently, the equations determine a Markoff chain with transition matrix $q^{-n}(C^{(n)})^t = P^n$, where P is the original transition matrix. By assumption there is a number $\tau \in \mathbb{N}$ so that for the Markoff chain with transition matrix P^τ, $\mathbf{S}$ contains τ nonempty irreducible subsets (Proposition 2(c)). Corollary 1 implies that $\mathcal{Q}^{\mathbf{S}} = \mathcal{Q}$ is not a $\mathbb{Z}^d$-tiling set. ∎

A nonnegative matrix B is called *primitive* if, for some $\tau \in \mathbb{N}$, B^τ has all positive entries. It follows from Proposition 2(d) that, under the hypothesis of the previous theorem, P must be primitive. Then certainly the original counting matrix C is also primitive. We have proved:

COROLLARY 2. *If Γ is a standard digit set and $\mathcal{Q}$ is a $\mathbb{Z}^d$-tiling set with $|Q_i| > 0$, $1 \le i \le M$, then C is primitive.*

Using arguments from Markoff chains we can now further clarify the nature of the solutions of (4). The next theorem shows that for standard digit sets the tiles in $\mathcal{Q}^N$ are — up to null sets — either equal to the corresponding tiles in the maximal solution $\mathcal{Q}$ or equal to sets of measure zero.

THEOREM 5. *If Γ is a standard digit set then for any (A, Γ)-closed $N \subseteq \mathbf{S}$, $Q_i^N \cong Q_i$, if $i \in N$, and $Q_i^N = \emptyset$, if $i \notin N$.*

Proof. If $i \in N$ is recurrent, it belongs to a unique irreducible set $L = R_{j_0} \subseteq N$ appearing in the decomposition (42). Since, by Theorem 1 $Q_i^L \subseteq Q_i^N \subseteq Q_i$, it will suffice to show that $|Q_i^L| = |Q_i|$.

According to Lemma 2, the vector with entries $|Q_i^N|$ is an eigenvector of P. Then Proposition 2(b) implies that $|Q_i^N| = 0$ for all transient states $i \in \mathbf{S}$.

Let $U \supseteq L$ be the smallest set with the property that if $j \in U$ and $k \notin U$ then $\Gamma_{jk} = \emptyset$. If $j \in U \setminus L$, then j is transient. This follows since, $j \in U \setminus L$ if and only if there is a $l \in L$ and a sequence $j = \rho_1, \ldots, \rho_m = l$ so that for each $k = 1, \ldots, m-1$, $\Gamma_{\rho_k \rho_{k+1}} \neq \emptyset$. Then, by the definition of irreducible, ρ_{m-1} is a transient state. Furthermore, by the definition of recurrent, ρ_{m-2} is transient and then, by induction, j is also transient. We conclude that for $j \in U \setminus L$, $|Q_j| = 0$.

Note that the equations of (4) express AQ_i for $i \in U$ as a union of translates of Q_j where the j also belong to U. Then the maximal solution $\hat{\mathcal{Q}} = \{\hat{Q}_i\}_{i \in U}$ of the equations

$$AT_i = \bigcup_{j \in U} (\Gamma_{ij} + T_j) = \bigcup_{j=1}^{M} (\Gamma_{ij} + T_j)$$

satisfies $\hat{Q}_i = Q_i$.

In terms of the representation (26), for each $i \in U$, $Q_i = \hat{Q}_i$ can be split into a disjoint union of sets, where the first contains all expansions with $\rho_k \in L$. Then Q_i can be written as

$$Q_i = Q_i^L \cup \bigcup_{n=2}^{\infty} E_{i,n}, \tag{45}$$

where

$$E_{i,n} = \left\{ \bigcup_{j \in U \setminus L} \left(\sum_{k=1}^{n-1} A^{-k} \Gamma_{\rho_k, \rho_{k+1}} + A^{-n} Q_j \right) \mid \rho_1 = i,\ \rho_n = j,\ \rho_k \in L,\ \forall 1 \leq k \leq n-1 \right\}.$$

Since $|Q_j| = 0$, each $E_{i,n}$ and therefore $\bigcup_n E_{i,n}$ is a countable union of sets of measure 0. It follows that $|Q_i^L| = |Q_i|$. ■

The next elementary theorem explains the problem that arises in cases like Example 7 with $a = 1$, $b = 1$, and $c = 0$. Note that no assumptions are made about the structure of the digit set.

THEOREM 6. *Let A and Γ be arbitrary and assume that for each $i \in \mathbf{S}$, Q_i has positive measure. Suppose also that there is a nontrivial permutation σ of $\mathbf{S}$ for which the set of dilation equations*

$$AT_{\sigma(i)} = \bigcup_{j=1}^{M} \Gamma_{ij} + T_{\sigma(j)} \tag{46}$$

with indices permuted is identical to the unpermuted set of dilation equations. Then the solution $\mathcal{Q}$ is not a lattice tiling set for any lattice Λ.

Proof. Choose $i \in \mathbf{S}$ for which $\sigma(i) \neq i$. Then since they are defined by the same digit expansions, $Q_i = Q_{\sigma(i)}$. By hypothesis Q_i has positive Lebesgue measure and therefore

$|Q_i \cap Q_{\sigma(i)}| > 0$. Thus $\mathcal{Q}$ fails to satisfy the intersection property (2) and can never be a tiling set. ■

It is possible to use the above information to give a complete analysis of the limiting behavior of points under iteration of the map φ defined in (28). Basically, looking at the decomposition of **S** in Proposition 2(a), aperiodic irreducible sets in the decomposition determine unique attracting fixed points, each with a well-defined basin of attraction. The irreducible sets that are not aperiodic fall under Proposition 2(c) and determine various attracting cycles, one corresponding to each of the the subgroups of the cyclic group of order τ. Each cycle has its own basin of attraction. We saw an example of this in the previous remark. Entries corresponding to transitive states will always converge to the empty set. The various behaviors combine to describe the general case.

5. WAVELET THEORY AND LATTICE TILINGS

There exists an interesting and at first glance surprising connection between wavelet theory and certain SATs of multiplicity 1 [10]. It is not surprising that similar results can be established for SATs with several tiles. This link between the geometric object of SATs and the analytic object of wavelet theory is important for several reasons:

(a) It singles out — among all SATs which seem too complicated to be classified completely — a special class of SATs which are more accessible to a detailed analysis;
(b) we will be able to apply methods from Fourier analysis and the theory of multiwavelets [7, 8, 13] to study SATs;
(c) SATs will furnish a new class of examples of multiwavelet bases, of which only few concrete examples and constructions are known so far.

We first recall that in wavelet theory one studies general approximation schemes, so-called MRAs. A *multiresolution analysis* $\mathcal{V}$ with respect to a dilation matrix A is a bi-infinite sequence of closed subspaces $V_j,\ j \in \mathbb{Z}$, of $L^2(\mathbb{R}^d)$ with the following properties:

- $V_j \subseteq V_{j+1}$ for all $j \in \mathbb{Z}$.
- $f(x) \in V_0$ if and only if $f(x-k) \in V_0$ for all $k \in \mathbb{Z}^d$.
- $f(x) \in V_0$ if and only if $f(A^j x) \in V_j$ for $j \in \mathbb{Z}$.
- V_0 possesses an orthonormal basis of the form $\{\phi_i(x-k),\ k \in \mathbb{Z}^d,\ i = 1, \ldots, M\}$.

We refer to [4, 20] for background and construction procedures for MRAs. The number M of basis functions is called the *multiplicity* of the MRA. The ϕ_i's uniquely determine the MRA $\mathcal{V}$ and are said to generate the MRA. MRAs with multiplicity > 1 recently have become the object of intensive studies [7, 8, 13]. While most general results carry over from dimension 1 and multiplicity 1 to $\mathbb{R}^d$ and $M > 1$ without significant modifications, concrete examples are sparse. No generic example of an MRA with arbitrary multiplicity M is known for general dilation matrices. In this sense SATs contribute an interesting facet to wavelet theory. The following theorem is the counterpart of Theorem 1 in [10].

THEOREM 7. (A) *Suppose that $\mathcal{T} = \{T_i,\ i = 1, \ldots, M\}$ is a self-affine $\mathbb{Z}^d$-tiling set with all prototiles of positive measure. Then the characteristic functions $\chi_{T_i}, i = 1, \ldots, M$, generate a multiresolution analysis for $L^2(\mathbb{R}^d)$.*

(B) *Conversely, if a multiresolution analysis is generated by characteristic functions* χ_{T_i}, $i = 1, \ldots, M$, *then the* $\mathcal{T}$ *is a self-affine* $\mathbb{Z}^d$ *tiling set. Moreover, the corresponding digits are a standard digit set.*

For the proof of the theorem we need a simple geometrical lemma first.

LEMMA 3. *Let* T *be a compact set in* $\mathbb{R}^d$ *whose* $\mathbb{Z}^d$*-translates are essentially disjoint. Then for any parallelepiped* $B = \prod_{i=1}^d [a_i, b_i] \subseteq \mathbb{R}^d$ *we have*

$$\lim_{j\to\infty} q^{-j} \sum_{k\in\mathbb{Z}^d} |(k+T)\cap A^j B|^2 = |T|^2|B|. \tag{47}$$

Proof. Let $C_\rho(B) = \{x \in \mathbb{R}^d : \|x-u\| \le \rho \text{ for } u \in \partial B\}$ be the "collar" of thickness 2ρ around the boundary of B. Using $\|A^{-1}x\| \le \lambda\|x\|$ with $\lambda < 1$, we see that

$$A^{-j}C_\rho(A^j B) \subseteq C_{\lambda^j\rho}(B)$$

and thus for any $\rho > 0$

$$\lim_{j\to\infty} q^{-j}|C_\rho(A^j B)| = \lim_{j\to\infty} |A^{-j}C_\rho(A^j B)| = 0. \tag{48}$$

Equipped with this observation we partition the index set $\mathbb{Z}^d$ in (47) into three subsets

$$\begin{aligned} A_j &= \mathbb{Z}^d \cap (A^j B \setminus C_\rho(A^j B)),\\ B_j &= \mathbb{Z}^d \cap C_\rho(A^j B),\\ C_j &= \mathbb{Z}^d \setminus (A_j \cup B_j). \end{aligned}$$

If we choose $\rho = \max\{\|x\| : x \in T \cup [0,1]^d\}$, then $k \in A_j$ implies that $k + T \subseteq A^j B$ and thus

$$q^{-j}\sum_{k\in A_j} |(k+T)\cap A^j B|^2 = |T|^2 q^{-j}\#A_j.$$

Similarly, $q^{-j}\sum_{k\in B_j} |(k+T)\cap A^j B|^2 \le |T|^2 q^{-j}\#B_j$. Finally, $k \in C_j$ means $|(k+T)\cap A^j B| = 0$ and the sum over C_j equals 0.

To estimate $\#A_j$, we observe that, by the choice of ρ,

$$A^j B \setminus C_{2\rho}(A^j B) \subseteq A_j + [0,1]^d \subseteq A^j B.$$

We combine the estimate

$$q^{-j}\left(|A^j B| - |C_{2\rho}(A^j B)|\right) \le q^{-j}\#A_j \le q^{-j}|A^j B| = |B|$$

with (48) and obtain $\lim_{j\to\infty} q^{-j}\#A_j = |B|$.

Similarly, $q^{-j}\#B_j \le q^{-j}|C_{2\rho}(A^j B)| \to 0$.

We conclude that

$$\lim_{j\to\infty} q^{-j}\sum_{k\in\mathbb{Z}^d} |(k+T)\cap A^j B|^2 = \lim_{j\to\infty} |T|^2 q^{-j}\#A_j = |T|^2|B|$$

as claimed. ∎

Proof of Theorem 7. (A) Suppose that $\mathcal{T}$ is a self-affine $\mathbb{Z}^d$-tiling set. Then we define $\psi_{ijk}(x) = q^{j/2}|T_i|^{-1/2}\chi_{T_i}(A^j x - k)$ and

$$V_j = \left\{ f \in L^2(\mathbb{R}^d) \mid f(x) = \sum_{i=1}^{M} \sum_{k \in \mathbb{Z}^d} a_k^{(i)} \psi_{ijk}(x) \text{ with } a^{(i)} \in \ell^2(\mathbb{Z}^d) \right\}. \tag{49}$$

It is now easy to verify that $\mathcal{V} = (V_j)_{j\in\mathbb{Z}}$ is an MRA of multiplicity M. The inclusions $V_j \subseteq V_{j+1}$ are an immediate consequence of the self-similarity (4), which amounts to the scaling relations

$$\chi_{T_i}(x) = \sum_{j=1}^{M} \sum_{k\in\Gamma_{ij}} \chi_{T_j}(Ax - k). \tag{50}$$

The existence of the orthonormal basis $\{|T_i|^{-1/2}\chi_{T_i}(x-k),\ k \in \mathbb{Z}^d,\ i = 1, \ldots, M\}$ for V_0 and the translation invariance of V_0 are guaranteed by the definition.

We only have to show that $\bigcup_{j\in\mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R}^d)$. For this we first compute the orthogonal projection $P_j f = \sum_{i=1}^{M} \sum_{k\in\mathbb{Z}^d} \langle f, \psi_{ijk}\rangle \psi_{ijk}$ from $L^2(\mathbb{R}^d)$ onto V_j for the characteristic function $f = \chi_B$ of a parallelepiped B. In this case

$$\langle \chi_B, \psi_{ijk}\rangle = q^{-j/2}|T_i|^{-1/2}|(k + T_i) \cap A^j B|$$

and thus

$$\|P_j \chi_B\|_2^2 = \sum_{i=1}^{M} q^{-j}|T_i|^{-1} \sum_{k\in\mathbb{Z}^d} |(k + T_i) \cap A^j B|^2. \tag{51}$$

Since $\mathcal{T}$ is a $\mathbb{Z}^d$-tiling, $\sum_{i=1}^{M} |T_i| = 1$. Therefore, by Lemma 3,

$$\lim_{j\to\infty} \|P_j \chi_B\|_2^2 = \sum_{i=1}^{M} |T_i||B| = \|\chi_B\|_2^2.$$

But since characteristic functions of parallelepipeds span a dense subspace of $L^2(\mathbb{R}^d)$ and since $\|f - P_j f\|_2^2 = \|f\|_2^2 - \|P_j f\|_2^2$, this suffices to show that $\bigcup V_j$ is dense in $L^2(\mathbb{R}^d)$.

We have verified that $(V_j)_{j\in\mathbb{Z}}$ is a multiresolution analysis.

(B) Now assume that the functions χ_{T_i} generate a MRA. Then the orthogonality of the basis functions gives immediately

$$|T_i|\delta_{ij}\delta_{kl} = \int_{\mathbb{R}^d} \chi_{T_i}(x-k)\chi_{T_j}(x-l)\, dx = |(k + T_i) \cap (l + T_j)|,$$

in other words, the disjointness of the tiles.

Next, since $\chi_{T_i}(A^{-1}x) \in V_{-1} \subseteq V_0$, it can be expressed in terms of the orthonormal basis of V_0 in the form of a so-called *scaling relation*:

$$\chi_{AT_i}(x) = \chi_{T_i}(A^{-1}x) = \sum_{j=1}^{M} \sum_{k\in\mathbb{Z}^d} c_{jk}\chi_{T_j}(x-k) \qquad \text{for } i = 1, \ldots, M, \tag{52}$$

where the coefficients $c_{jk} = \int \chi_{AT_i}(x)\chi_{k+T_j}(x)\,dx$ can only take the values 0 or 1, because the integer translates of the prototiles are disjoint. If for each i we denote the set of translates $k \in \mathbb{Z}^d$ for which $c_{jk} = 1$ by Γ_{ij}, then (52) can be rewritten as the following self-similarity of sets

$$AT_i \cong \bigcup_{j=1}^{M} (\Gamma_{ij} + T_j) \qquad \text{for } i = 1, \ldots, M.$$

Next we deduce that the T_i tile with lattice $\mathbb{Z}^d$.

Since $\bigcup V_j$ is dense in $L^2(\mathbb{R}^d)$, we know that $\|P_j \chi_B\|_2^2 \to \|\chi_B\|_2^2 = |B|$ for any parallelepiped B. On the other hand, from (51) we know that

$$\lim_{j\to\infty} \|P_j \chi_B\|_2^2 = |B| \sum_{i=1}^{M} |T_i|.$$

It follows that $\sum_{i=1}^{M} |T_i| = 1$.

Now consider the function

$$\Phi(x) = \sum_{i=1}^{M} \sum_{k\in\mathbb{Z}^d} \chi_{T_i}(x-k).$$

Since the prototiles are pairwise disjoint, $0 \le \Phi(x) \le 1$. Then

$$\int_{[0,1]^d} \Phi(x)\,dx = \int_{\mathbb{R}^d} \left(\sum_{i=1}^{M} \chi_{T_i}(x) \right) dx = \sum_{i=1}^{M} |T_i| = 1.$$

We see that $\Phi(x) = 1$, which is equivalent to the $\mathbb{Z}^d$-tiling property of $\mathcal{T}$.

It follows now from Theorem 3 that Γ is a standard digit set, and the theorem is proved completely. ■

As a consequence we can construct orthonormal wavelet bases with compact support, but without smoothness, starting from SATs.

THEOREM 8. *Suppose that $\mathcal{T}$ is a self-affine $\mathbb{Z}^d$-tiling set. Then there exist $(q-1)M$ functions ψ_l with compact support in $\bigcup_{j=1}^{M} T_j$, such that*

$$\left\{ q^{j/2}\psi_l(A^j x - k),\ j \in \mathbb{Z},\ k \in \mathbb{Z}^d,\ l = 1, \ldots, (q-1)M \right\} \tag{53}$$

is an orthonormal basis for $L^2(\mathbb{R}^d)$. The ψ_l can be written explicitly as linear combinations of the functions $\chi_{T_i}(Ax-k),\ k \in \bigcup \Gamma_{ij}$.

Proof. Following the standard line of arguments we have to find an orthonormal basis of the form $\{\psi_l(x-k),\ k \in \mathbb{Z}^d,\ l = 1, \ldots, (q-1)M\}$ in $W_0 := V_1 \ominus V_0$, the orthogonal complement of V_0 in V_1. Since $\bigoplus_{j\in\mathbb{Z}} W_j = \overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R}^d)$, the collection (53) is then an orthonormal basis for $L^2(\mathbb{R}^d)$. We refer to [4, 20] for the general construction of wavelet bases.

By definition (49) $f \in V_1$ if

$$f(x) = \sum_{i=1}^{M} \sum_{k \in \mathbb{Z}^d} a_{ik} q^{1/2} |T_i|^{-1/2} \chi_{T_i}(Ax - k)$$

with $(a_{ik})_{k \in \mathbb{Z}^d} \in \ell^2(\mathbb{Z}^d)$, and $f \in W_0$, if and only if $f \perp \chi_{T_j}(x - l)$. Rewriting (52) as

$$\chi_{T_j}(x - l) = \sum_{r=1}^{M} \sum_{m \in \Gamma_{jr}} \left(\frac{q}{|T_r|}\right)^{-1/2} \left(\frac{q}{|T_r|}\right)^{1/2} \chi_{T_r}(A(x - l) - m),$$

we calculate for $f \in W_0$

$$\begin{aligned} 0 = \langle f, \chi_{T_j}(. - l)\rangle &= \sum_{i=1}^{M} \sum_{r=1}^{M} \sum_{m \in \Gamma_{jr}} \sum_{k \in \mathbb{Z}^d} \left(\frac{|T_r|}{q}\right)^{1/2} \delta_{ir} \delta_{k, Al+m} a_{ik} \\ &= \sum_{r=1}^{M} \sum_{m \in \Gamma_{jr}} \left(\frac{|T_r|}{q}\right)^{1/2} a_{r, Al+m}. \end{aligned} \tag{54}$$

Now consider the M linear equations

$$\sum_{r=1}^{M} \sum_{m \in \Gamma_{jr}} |T_r|^{1/2} a_{rm} = 0 \tag{55}$$

in the $\sum_{j=1}^{M} \sum_{r=1}^{M} \#\Gamma_{jr} = qM$ variables a_{rm}. It is not hard to check that these equations are linearly independent. Consequently the null space has dimension $qM - M$.

Choose an orthonormal basis $(u_{rm}^{(s)})$, $s = 1, \ldots, qM - M$, for the null space and define the functions

$$\psi_s(x) = \sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{m \in \Gamma_{ji}} u_{im}^{(s)} \left(\frac{q}{|T_i|}\right)^{1/2} \chi_{T_i}(Ax - m). \tag{56}$$

Then supp $\psi_s \subseteq \bigcup_{j=1}^{M} T_j$. From these support properties the orthogonality relations

$$\langle \psi_s, \chi_{T_j}(. - l)\rangle = 0 = \langle \psi_s, \psi_{s'}(. - l)\rangle$$

for $l \neq 0$ or $s \neq s'$ are clear.

Since by (54)

$$\langle \psi_s, \chi_{T_j}\rangle = \sum_{r=1}^{M} \sum_{m \in \Gamma_{jr}} u_{rm}^{(s)} (|T_r|/q)^{1/2} = 0,$$

we see that $\psi_s \in W_0$. Since

$$\langle \psi_s, \psi_{s'}\rangle = \sum_{r=1}^{M} \sum_{m \in \Gamma_{jr}} u_{rm}^{(s)} \overline{u_{rm}^{(s')}} = \delta_{ss'},$$

the functions $\{\psi_s(x - k),\ k \in \mathbb{Z}^d,\ s = 1, \ldots, qM - M\}$ form an orthonormal system in W_0.

We finally show that this orthonormal system is complete in W_0. Suppose that $f \in W_0$ is orthogonal to this basis. Then the coefficients of f satisfy (54) and for all l and s

$$\langle f, \psi_s(. - l)\rangle = \sum_{r=1}^{M} \sum_{m \in \Gamma_{jr}} a_{k,Al+m} \overline{u_{rm}^{(s)}} = 0.$$

The vectors $u^{(s)}$ are by definition an ONB for the null space in (54) and thus for all $l \in \mathbb{Z}^d$ we have $a_{k,Al+m} = 0$. In other words, $f = 0$ and the orthonormal system is complete in W_0. ∎

A more explicit solution can be obtained as follows: First find an orthonormal basis of vectors $u = (u_{rm})$, satisfying $\sum_{r=1}^{M} \sum_{m \in \Gamma_{jr}} u_{rm} = 0$ for a fixed j and define for each such u the function

$$\psi_u(x) = \sum_{r=1}^{M} \sum_{m \in \Gamma_{jr}} u_{rm} (q/|T_r|)^{1/2} \chi_{T_r}(Ax - m).$$

Then supp $\psi_u \subseteq T_j$. Counting dimensions, for each j there are exactly $\sum_{r=1}^{M} c_{jr} - 1$ such functions. Doing this for each j, we get a collection of $\sum_{j=1}^{M} \sum_{r=1}^{M} c_{rm} - M = qM - M$ functions ψ_u. As above they form an orthonormal basis for W_0.

In the context of multiwavelet theory it is therefore of interest to know when the construction of an SAT, starting from a standard digit set yields a $\mathbb{Z}^d$-tiling of $\mathbb{R}^d$. The examples indicate that two phenomena may contribute to a failure.

(a) The prototiles T_i tile with a coarser lattice than $\mathbb{Z}^d$. See Example 7 with $a = b = c = 2$. If the multiplicity is one, this is the only obstruction [3, 18].

(b) *The redundant case:* $\mathbb{R}^d$ can already be tiled by $\mathbb{Z}^d$ with a smaller number $N < M$ of tiles. See Examples 3 and 7 with $a = b = 1$, $c = 0$. This is a genuinely new phenomenon in the case of higher multiplicity.

We conjecture that a combination of these two cases is all that can go wrong. The fundamental question in this context is to decide which choices of standard digit sets generate $\mathbb{Z}^d$-tilings. This is a difficult and subtle question even in the case of only one tile; see [2, 3, 9, 16, 17, 18]. We shall return to this question in Part II.

ACKNOWLEDGEMENT

This research began while the third author visited the Department of Mathematics at the University of Connecticut. Its support and hospitality are gratefully acknowledged. The second author thanks the University of Washington for providing access to their research facilities while this paper was in preparation.

REFERENCES

1. M. Barnsley, “Fractals Everywhere,” Academic Press, New York, 1988.
2. D. Cerveau, J.-P. Conze, and A. Raugi, Fonctions harmoniques pour un opérateur de transition en dimension > 1, *Bol. Soc. Bras. Mat.* **27** (1996), 1–26.
3. J.-P. Conze, L. Hervé, and A. Raugi, Pavages auto-similaires, opérateurs de transfert et critères de réseau, *Bol. Soc. Bras. Mat.* **28** (1997), 1–42.

4. I. Daubechies, "Ten Lectures on Wavelets," CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 61, SIAM, Philadelphia, 1992.
5. W. Feller, "An Introduction to Probability Theory and its Applications," Vol. I, Wiley, New York, 1968.
6. T. Flaherty and Y. Wang, Haar-type multiwavelet bases and self-affine multi-tiles, preprint, March 1997.
7. J. S. Geronimo, D. P. Hardin, and P. R. Massopust, Fractal functions and wavelet expansions based on several scaling functions, *J. Approx. Theory* **78** (1994), 373–401.
8. T. Goodman and S. L. Lee, Wavelets of multiplicity r, *Trans. Amer. Math. Soc.* **342** (1994), 307–324.
9. K. Gröchenig and A. Haas, Self-similar lattice tilings, *J. Fourier Anal. Appl.* **1** (1994), 131–170.
10. K. Gröchenig and W. Madych, Multiresolution analysis, Haar bases, and self-similar tilings, *IEEE Trans. Inform. Theory* **38**, Part 2 (1992), 556–568.
11. B. Grünbaum and G. C. Shephard, "Tilings and Patterns," Freeman, New York, 1987.
12. D. Hacon, N. C. Saldanha, and J. P. Veerman, Remarks on self-affine tilings, *Experiment. Math.* **3** (1994), 317–326.
13. Q. Jiang, Multivariate matrix refinable functions with matrix dilations, preprint, 1997.
14. M. Keane, Ergodic theory and subshifts of finite type, *in* "Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces" (T. Bedford, M. Keane, and C. Series, Eds.), Oxford Univ. Press, Oxford, 1991.
15. R. Kenyon, Self-replicating tilings, *in* "Symbolic Dynamics and Applications" (P. Walters, Ed.), *Contemp. Math.* **135** (1992), 239–264.
16. J. Lagarias and Y. Wang, Integral self-affine tilings in $\mathbb{R}^n$, I, *J. London Math. Soc.* **53** (1996), 161–179.
17. J. Lagarias and Y. Wang, Haar type wavelet bases in $\mathbb{R}^n$ and algebraic number theory, *J. Number Theory* **57** (1996), 181–197.
18. J. C. Lagarias and Y. Wang, Integral self-affine tiles in $\mathbb{R}^n$. II. Lattice tilings, *J. Fourier Anal. Appl.* **3** (1997), 83–102.
19. B. Mandelbrot, "The Fractal Geometry of Nature," Freeman, New York, 1983.
20. Y. Meyer, "Ondelettes et Opérateurs I, II," Hermann, Paris, 1990.
21. R. Penrose, The role of aesthetics in pure and applied mathematical research, *Bull. Inst. Math. Appl.* **10** (1974), 266–271.
22. C. Radin, Space tilings and substitutions, *Geom. Dedicata* **55** (1995), 257–264.
23. R. S. Strichartz, Wavelets and self-affine tilings, *Constructive Approx.* **9** (1993), 283–297.
24. W. P. Thurston, Groups, tilings, and finite state automata, *in* "Lecture Notes," Summer Meeting of the Amer. Math. Soc., Boulder, 1989.
25. A. Vince, Replicating tesselations, *SIAM J. Discrete Math.* **6** (1993), 501–521.