The Projective Fundamental Group of a $\mathbf{Z}^2$-Shift

(Running title: Projective Fundamental Group of a $\mathbf{Z}^2$-Shift)
(Contact author: James Propp)

October 26, 1993

William Geller[1] and James Propp[2]

Abstract

We define a new invariant for symbolic $\mathbf{Z}^2$-actions, the *projective fundamental group*. This invariant is the limit of an inverse system of groups, each of which is the fundamental group of a space associated with the $\mathbf{Z}^2$-action. The limit group measures a kind of long-distance order that is manifested along loops in the plane, and roughly speaking bears the same relation to the mixing properties of the $\mathbf{Z}^2$-action that π_1 of a topological space bears to π_0. The projective fundamental group is invariant under topological conjugacy. We calculate this invariant for several important examples of $\mathbf{Z}^2$-actions, and use it to prove nonexistence of certain kinds of constant-to-one maps between two-dimensional subshifts. Subshifts that have the same entropy and periodic point data can have different projective fundamental groups.

Key words: fundamental group, covering space, subshift of finite type, Markov shift, tiling, Wang tile, constant-to-one.

AMS subject classification: 58F03 [symbolic dynamics]; secondary 52C20 [tilings in 2 dimensions], 57M05 [fundamental group], 57M10 [covering spaces].

This work was supported in part by grants from NSF, NSA, and MSRI.

[1]Department of Mathematics, U.C. Berkeley, Berkeley, CA 94720; geller@math.berkeley.edu.

[2]Department of Mathematics, M.I.T, Cambridge, MA 02139; propp@math.mit.edu.

1 Introduction

The dynamical systems we are interested in are essentially families of functions from $\mathbf{Z}^2$ to a set S of "symbols". The group $\mathbf{Z}^2$ acts on such functions f by translation: $(T_x f)(y) = f(x+y)$ $(x, y \in \mathbf{Z}^2)$. If a family of functions is closed under translation, it is called a $\mathbf{Z}^2$-shift. Many interesting $\mathbf{Z}^2$-actions can be put into this framework. The basic theory of symbolic $\mathbf{Z}^2$-actions is given in [Sch1], along with a catalogue of important examples.

In this paper we introduce a new invariant for these systems.

Instead of limiting ourselves to discussing families of functions from $\mathbf{Z}^2$ to S, we will work in a slightly broader setting that includes more general dynamical systems such as the set of Penrose tilings of the plane. Define an *album* over $\mathbf{R}^2$ as a family $\mathcal{F}$ of functions $f : D \to S$, where $D \subseteq \mathbf{R}^2$ and S is a discrete topological space. In the case where $\mathcal{F}$ is a $\mathbf{Z}^2$-shift, $D = \mathbf{Z}^2$ and $\mathcal{F}$ is invariant under the action of $\mathbf{Z}^2$. We say that the album $\mathcal{F}$ is of *finite type* if there exists a cover of $\mathbf{R}^2$ by countably many sets U_i of bounded diameter and families $\mathcal{F}_i$ of functions from U_i to S, such that a function $f : D \to S$ belongs to $\mathcal{F}$ if and only if for all i the restriction of f to U_i belongs to $\mathcal{F}_i$. An equivalent definition is that $\mathcal{F}$ is of finite type if there exists a bounded-diameter cover $\{U_i\}$ such that a function $f : D \to S$ belongs to $\mathcal{F}$ whenever for all i there exists a function f_i in $\mathcal{F}$ that agrees with f on U_i.

If the album $\mathcal{F}$ is a $\mathbf{Z}^2$-shift and is of finite type, we will call it a subshift of finite type. This coincides with the usual definition except that we are allowing the alphabet S to be infinite. Most specific albums and subshifts discussed in this article are of finite type, but our general results apply without this constraint.

Put a metric on $\mathcal{F}$ by defining

$$d(f,f') = \frac{1}{1+\inf\{||x|| : f(x) \not\equiv f'(x)\}}$$

where $||\cdot||$ is the Euclidean norm on $\mathbf{R}^2$ and where we write $f(x) \equiv f'(x)$ if $f(x) = f'(x)$ or if $f(x), f'(x)$ are undefined (that is, if $x \notin D$). This makes $\mathcal{F}$ a complete metric space in which $f_n \to f$ if and only if for every bounded set $B \subset \mathbf{R}^2$, f_n agrees with f everywhere on B for all n sufficiently large. As a topological space $\mathcal{F}$ is totally disconnected. Nevertheless, we will see that by regarding $\mathbf{R}^2 \times \mathcal{F}$ as the inverse limit of a family of spaces related to $\mathcal{F}$, we can make sense of the notion of connecting two elements f, f' of $\mathcal{F}$ by something like a path, obtaining in the process a notion of a "fundamental group" for the album $\mathcal{F}$.

Let $\mathcal{S} = \mathcal{S}(\mathcal{F})$ be the product space $\mathbf{R}^2 \times \mathcal{F}$. Given a bounded subset $B \subset \mathbf{R}^2$, define an equivalence relation $\sim_B$ on $\mathcal{S}$ by putting $(x,f) \sim_B (x',f')$ ($x, x' \in \mathbf{R}^2$, $f, f' \in \mathcal{F}$) if and only if $x = x'$ and $f(y) = f'(y)$ for all y in the translated set $B + x \subset \mathbf{R}^2$ (the "B-neighborhood of x") for which $f(y)$ is defined. Equivalently: if we define the *scene* of f at x to be the restriction of f to $B+x$ (that is, to $B+x\cap \mathrm{dom} f$), then $(x,f)\sim_B (x,f')$ if and only if f and f' have the same scene at x.

Let $\mathcal{S}_B$ be the quotient space $\mathcal{S}/{\sim_B}$ ("scene-space"), with the quotient-map $p_B:\mathcal{S}\to\mathcal{S}_B$. A subset of $\mathcal{S}_B$ is open if and only if its pre-image in $\mathcal{S}$ is open. For all B, B' with $B\subseteq B'$ we get a canonical *restriction map* $\mathcal{S}_{B'}\to\mathcal{S}_B$ (continuous). Fix $x_0\in\mathbf{R}^2$ and $f_0\in\mathcal{F}$, and let $q=p_B(x_0,f_0)$ and $q'=p_{B'}(x_0,f_0)$. Then q is the image of q' under the restriction map $\mathcal{S}_{B'}\to\mathcal{S}_B$, and we obtain a homomorphism $\pi_1(\mathcal{S}_{B'},q')\to\pi_1(\mathcal{S}_B,q)$, where $\pi_1(\mathcal{S}_B,q)$ denotes the fundamental group of $\mathcal{S}_B$ relative to the basepoint q (similarly for $q',\mathcal{S}_{B'}$). Thus the directed system of bounded sets $B\subset\mathbf{R}^2$, where arrows are inclusions, gives rise to an inverse system of groups $G_B=\pi_1(\mathcal{S}_B,p_B(x_0,f_0))$, where arrows are group homomorphisms going the other way. We define G_∞ as the inverse limit of these fundamental groups. It is not hard to show that G_∞ is independent of x_0. If the inverse system $\{\mathcal{S}_B\}$ satisfies a property we call "projective connectedness", then G_∞ turns out to be independent of f_0 as well; in this case we call it the *(projective) fundamental group* of the album $\mathcal{F}$ and denote it by $\pi_1^{\mathrm{proj}}(\mathcal{S}(\mathcal{F}))$. Projective connectedness will be defined in Section 3. For now it suffices to describe it as a novel kind of mixing property, and to remark that all the examples considered in this paper (other than in Section 4) exhibit this property. A factor map (i.e., a shift-commuting continuous surjection) from one $\mathbf{Z}^2$-subshift to another gives rise to a group homomorphism between their projective fundamental groups.

Henceforth, when there is no danger of confusion we will drop the modifier "projective".

The remainder of this article is organized as follows. First, we will lay some groundwork for the fundamental group by considering two examples in an informal way (Section 2). We then discuss inverse limits of spaces and fundamental groups (Section 3), and apply this to scene-space (Section 4); in particular we prove, as a special case of a more general result on albums, that the fundamental group of a $\mathbf{Z}^2$-subshift is invariant under topological conjugacy (Theorem 1). The two sections that follow calculate the fundamental group for two important and instructive special cases, namely, the full shift (Section 5, Theorem 2) and the square ice model (Section 6, Theorem 3). Thereafter (Section 7) we give a concrete characterization of covering factor maps in terms of (cardinality and separation) properties of their fibers, and proceed to prove that, if a factor map from a subshift $\mathcal{F}$ to another subshift $\mathcal{G}$ is d-to-1 and covering, where $\mathcal{F}$ and $\mathcal{G}$ are projectively connected and $\mathcal{G}$ is compact, then the fundamental group of $\mathcal{F}$ is isomorphic to an index-d subgroup of the fundamental group of $\mathcal{G}$. One consequence of this is that there exists no non-trivial constant-to-one covering factor map from a full shift to itself. We also show (Section 8) that there exist subshifts of finite type that have the same entropy and periodic point data but are distinguished by their fundamental groups. We prove there that every group of order d arises as the fundamental group of a d-to-1 factor of a full shift. One can view these results as giving a first step towards the classification of constant-to-one factors of full $\mathbf{Z}^2$-shifts.

2 Examples

To get an intuitive picture of scene-space, think of $\mathcal{S}$ as a totally disconnected stack of copies of $\mathbf{R}^2$ (or *sheets*), one for each function $f \in \mathcal{F}$; each sheet is decorated in accordance with the associated f, and has a projection map to the standard (undecorated) copy of $\mathbf{R}^2$. Let B be some bounded set (it is helpful to think of it as a disk centered on the origin). Then $\sim_B$ identifies points in two sheets if they lie above the same point in $\mathbf{R}^2$ and if the respective planes have identical decoration within the B-neighborhoods of the two points. $\mathcal{S}_B$ is thus a space each of whose elements is given by a B-shaped aperture in $\mathbf{R}^2$ (i.e. a translate of the bounded set B) together with a decoration of the enclosed region. We may think of points in scene-space as pairs $(x, \overline{f})$, where $\overline{f}$ is a function on $B + x$ that arises as the restriction of some function in $\mathcal{F}$. Note that the fiber of $\mathcal{S}_B$ above any point x in $\mathbf{R}^2$ is discrete. Any continuous map from $\mathbf{R}^2$ to $\mathcal{S}_B$ that is a one-sided inverse of the projection map takes $\mathbf{R}^2$ to a subset of $\mathcal{S}_B$ homeomorphic to $\mathbf{R}^2$; we call this subset a *section* of $\mathcal{S}_B$. (It would be more consonant with tradition to call the *map* the section, rather than its image, but since points in $\mathcal{S}_B$ retain their $\mathbf{R}^2$-labelling, this is merely a technical distinction.) A section is thus a way of assigning a scene to $B + x$ for each $x \in \mathbf{R}^2$, such that overlapping scenes agree. When $\mathcal{F}$ is of finite type, by taking B to be sufficiently large we can guarantee that every section of $\mathcal{S}_B$ comes from a (unique) sheet in $\mathcal{S}$ via the projection map p_B.

A path in $\mathcal{S}_B$ is a "movie" in which the aperture travels through $\mathbf{R}^2$ with scenery appearing along the way. That is, a path $((x_t, \overline{f}_t) : 0 \leq t \leq 1)$ in $\mathcal{S}_B$ is given by a path $(x_t : 0 \leq t \leq 1)$ in $\mathbf{R}^2$, together with a set of scenes $\overline{f}_t$ on the respective sets $B + x_t$; the scenes must be locally consistent with one another, in the sense that for all x, $\overline{f}_t(x)$ (viewed as a function of t) is locally constant on its domain (a subset of $[0, 1]$).

The simplest paths in $\mathcal{S}_B$ are those that come from a single function f and from a path $(x_t : 0 \leq t \leq 1)$ in $\mathbf{R}^2$: for each t, let $\overline{f}_t$ be the restriction of f to $B + x_t$ (if x_t is constant, we call the path trivial). We say that such a path lies on a single section in $\mathcal{S}_B$. Not all paths are of this type; indeed, if we let $(x_t : 0 \leq t \leq 1)$ be a path that crosses itself (say $x_{t_1} = x_{t_2} = x^*$), then one can have a path $((x_t, \overline{f}_t) : 0 \leq t \leq 1)$ in $\mathcal{S}_B$ such that $\overline{f}_{t_1}$ and $\overline{f}_{t_2}$ do not agree at x^*, let alone in a B-neighborhood of x^*.

A loop in $\mathcal{S}_B$ is a path whose endpoints coincide. For an example of a loop, let $\mathcal{F}$ be the full (two-dimensional) 2-shift $\{0, 1\}^{\mathbf{Z}^2}$, viewed as a set of functions from $\mathbf{Z}^2 \subset \mathbf{R}^2$ to $\{0, 1\}$, and let B be the closed 1-by-1 square centered on $(0, 0)$. Consider the loop L in $\mathcal{S}_B$ in which the center x_t of the 1-by-1 square aperture travels from $(0, 0)$ to $(6, 0)$ to $(6, 6)$ to $(0, 6)$ to $(0, 0)$, with the "all-0's" scenery shown in Figure 1. (As in many of our pictures, we indicate the location of the origin by underlining the associated symbol.) Let $\overline{L}$ denote the projection of L into $\mathbf{R}^2$. Then any homotopy of $\overline{L}$ in $\mathbf{R}^2$ lifts to a homotopy of L in $\mathcal{S}_B$. (To see this, return to our original, unquotiented picture: every homotopy in $\mathbf{R}^2$ lifts to a homotopy in the all-0's sheet σ of $\mathcal{S} = \mathbf{R}^2 \times \mathcal{F}$, which maps down to a homotopy in $\mathcal{S}_B$ under the projection p_B.) In particular, if we homotopically transform the loop in $\mathbf{R}^2$ into a trivial loop, we will do the same for the loop in $\mathcal{S}_B$. That is, the loop L is contractible. More generally, any loop that lives on a single section of $\mathcal{S}_B$ is contractible.

For a more interesting example, consider the loop L in $\mathcal{S}_B$ in which the center x_t of the 1-by-1 square aperture travels from $(0,0)$ to $(6,0)$ to $(6,6)$ to $(0,6)$ to $(0,0)$ to $(0,-6)$ to $(-6,-6)$ to $(-6,0)$ to $(0,0)$, with the scenery shown in Figure 2. (Part (a) shows the scenery for $0 \le t \le 3/8$; part (b), for $3/8 \le t \le 5/8$; and part (c), for $5/8 \le t \le 1$.) With the exception of the symbol "1" seen at $(0,0)$ at time $t = 1/2 \pm \epsilon$, our movie shows the symbol "0" everywhere. This loop does not live on a single section, as in the preceding example; nevertheless L is contractible. For, let σ_0 be the all-0's sheet in $\mathcal{S}$, and let σ_1 be the sheet with a 1 at the origin and 0's everywhere else, with images $\overline{\sigma}_0$, $\overline{\sigma}_1$ in $\mathcal{S}_B$; note that $\overline{\sigma_0}$ and $\overline{\sigma_1}$ are pinched together everywhere except in a 1-by-1 square centered on the origin. The portion of the loop L that goes through the origin at time $t = 1/2 \pm \epsilon$ can be homotopically "pushed out of the way" without leaving the section $\overline{\sigma}_1$. This new loop skirts the origin at time $t = 1/2$ so that the symbol 1 is never seen, which insures that the loop can now be shrunk down on the section $\overline{\sigma}_0$.

It will be shown later that in the case of the full 2-shift, if B is an open $2m$-by-$2m$ square then every loop in $\mathcal{S}_B$ is contractible – that is, the group $\pi_1(\mathcal{S}_B)$ is trivial. This is not true if B is a general bounded set in $\mathbf{R}^2$ (such as an annulus), but this fact will not trouble us; since every bounded set lies in an m-by-m square for m sufficiently large, the inverse limit of the fundamental groups of the spaces $\mathcal{S}_B$ will be trivial.

The same analysis that works for the full 2-shift (or indeed the full n-shift for any positive integer n) can also be used to show that the fundamental group of the "two-dimensional golden mean subshift" [Sch1] is trivial. On the other hand, for the square ice model [Lieb], the fundamental group is $\mathbf{Z}$. Recall that a state for the square ice model is an orientation of the edges of the infinite square grid satisfying the *divergence condition*: each vertex has indegree 2 and outdegree 2 (see Figure 3(a)). It is more convenient to work with the dual model, in which vertices become squares and vice versa, with each oriented edge getting rotated 90 degrees clockwise about its midpoint (see Figure 3(b)). States of the dual model must satisfy the *curl condition*: each 1-by-1 square cell of the grid must have two of its edges oriented clockwise and the other two oriented counterclockwise. It follows that around *any* simple closed lattice path, there must be equal numbers of clockwise and counterclockwise edges. Now observe that in Figure 3(c), there is no way of filling the interior of the square so as to obtain a dual ice-configuration, for along the loop L (drawn in bold) there are 12 counterclockwise edges and only 4 clockwise edges. One can show that for sufficiently large B the curl around a loop is invariant under homotopy of the loop in $\mathcal{S}_B$. This implies that for such B, $\mathcal{S}_B$ has fundamental group $G_B \cong \mathbf{Z}$, where the obstruction to contractibility comes from the non-trivial curl along loops. This phenomenon does not disappear if we make B larger, or if we pass to the inverse limit with respect to B; that is, $G_\infty \cong \mathbf{Z}$.

The dimer model on a square grid [Kast] can be analyzed by much the same method as the square ice model. We remind the reader that the dimer model is the $\mathbf{Z}^2$-shift whose points are the ways of choosing certain edges of the infinite square grid to serve as "bonds," in such a way that each vertex of the grid lies on exactly one bond. The partial configuration shown in Figure 4, with bonds represented by bold line segments, cannot be extended to the full interior of the annular region, since this region has 1 more black vertex than white vertex under an obvious alternating coloring. Once again, the fundamental group is $\mathbf{Z}$, except that now the obstruction to contractibility is the imbalance between the number of white vertices and black vertices enclosed. This group $\mathbf{Z}$ is arrived at in a different way by Thurston [Thur], in his exposition of the Conway-Lagarias theory of "boundary invariants"; he uses a dual picture, in which a dimer-cover of an infinite square grid becomes a domino-tiling of the plane.

3 Topological Preliminaries

If $P : [0,1] \to X$ is a path in a topological space X, we denote by $[P]$ the path-class of P, i.e. the set of paths that have the same initial point $P(0)$ and terminal point $P(1)$ as P and are homotopic to P. If the terminal point of $[P]$ coincides with the initial point of $[Q]$, then the composition $[Q]\circ[P] = [Q\circ P]$ is well-defined; the set of path classes under this form of restricted composition forms the fundamental groupoid of X, which we denote by $\pi_{\mathbf{path}}(X)$.

We denote by $\pi_1(X, x)$ the fundamental group of X based at $x \in X$. If X is path-connected, so that $\pi_1(X, x)$ is independent of x, we simply write $\pi_1(X)$.

If X, Y are topological spaces with $\psi : X \to Y$ continuous, we let ψ_* denote the induced homomorphism from the fundamental group(oid) of X to the fundamental group(oid) of Y.

Suppose that $\psi : X \to Y$ is a covering map. Given a point $x \in X$ and a path P in Y with $P(0) = y = \psi(x)$, we let $\widehat{P}$ denote the lift of P to X with initial point x. When P is a closed loop, we refer to the self-map of the fiber over $P(0)$ determined by P as the monodromy of the loop.

Let $(\mathcal{A}, \preceq)$ be a directed set, indexing a system of topological spaces X_α ($\alpha \in \mathcal{A}$) equipped with commuting continuous maps $\psi^{\beta\leftarrow\alpha} : X_\beta \leftarrow X_\alpha$ ($\alpha, \beta \in \mathcal{A}$, $\alpha \succeq \beta$). The inverse (or projective) limit of the inverse system $\{X_\alpha, \psi^{\beta\leftarrow\alpha}\}$ is the set of $x \in \prod_{\alpha\in\mathcal{A}} X_\alpha$ having the property that $\psi^{\beta\leftarrow\alpha}(x_\alpha) = x_\beta$ for all $\alpha, \beta \in \mathcal{A}$ with $\alpha \succeq \beta$; we write $x = \mathrm{Lim}_\alpha\, x_\alpha$ and we denote the inverse limit by $\mathrm{Lim}_{\alpha\in\mathcal{A}} X_\alpha$ or just X_∞. We define projection maps $p_\alpha : X_\infty \to X_\alpha$, $x \mapsto x_\alpha$. We take the topology on X_∞ that is generated by pre-images of open sets under the maps p_α. If $\mathcal{B}_1, \mathcal{B}_2$ are subsets of $\mathcal{A}$ having the property that for all $\beta_1 \in \mathcal{B}_1$ there exists $\beta_2 \in \mathcal{B}_2$ with $\beta_2 \succeq \beta_1$, then the maps $\psi^{\beta_1 \leftarrow \beta_2}$ induce a canonical map from $\mathrm{Lim}_{\beta\in\mathcal{B}_2} X_\beta$ to $\mathrm{Lim}_{\beta\in\mathcal{B}_1} X_\beta$; if moreover it happens that for all $\beta_2 \in \mathcal{B}_2$ there exists $\beta_1 \in \mathcal{B}_1$ with $\beta_1 \succeq \beta_2$ ("mutual domination"), this map is bijection, and the two inverse limits coincide (i.e. are homeomorphic as spaces).

Similarly, if $\{G_\alpha, \psi^{\beta\leftarrow\alpha}\}$ is an inverse system of group(oid)s G_α ($\alpha \in \mathcal{A}$) with commuting group(oid) homomorphisms $\psi^{\beta\leftarrow\alpha} : G_\beta \leftarrow G_\alpha$ ($\alpha, \beta \in \mathcal{A}$, $\alpha \succeq \beta$), then the inverse limit is the set G_∞ of $g \in \prod_{\alpha\in\mathcal{A}} G_\alpha$ satisfying $\psi^{\beta\leftarrow\alpha}(g_\alpha) = g_\beta$; it becomes a group(oid) if one puts $(g \circ g')_\alpha = g_\alpha \circ g'_\alpha$. As before, $\mathrm{Lim}_{\beta\in\mathcal{B}_1} G_\beta \cong \mathrm{Lim}_{\beta\in\mathcal{B}_2} G_\beta$ provided that every element of $\mathcal{B}_1$ (resp. $\mathcal{B}_2$) is dominated by some element of $\mathcal{B}_2$ (resp. $\mathcal{B}_1$).

If $\{X_\alpha, \psi^{\beta\leftarrow\alpha}\}$ is an inverse system of spaces, and x is in $\mathrm{Lim}_\alpha X_\alpha$, then we obtain an inverse system $\{\pi_1(X_\alpha, x_\alpha), \psi_*^{\beta\leftarrow\alpha}\}$ of fundamental groups; its projective limit is the *projective fundamental group* of $\{X_\alpha\}$ at the point x, and its elements are *projective loop-classes*. For $x \in X_\infty$, we let $\pi_1^{\mathrm{proj}}(X_\infty, x)$ denote the projective fundamental group of $\{X_\alpha\}$ at x. This inverse limit of fundamental groups is not to be confused with $\pi_1(X_\infty, x)$, the fundamental group of an inverse limit of spaces; however, we retain the X_∞ (instead of the more accurate $\{X_\alpha\}$) for notational convenience, and to remind ourselves that an element of $\pi_1^{\text{proj}}(X_\infty, x)$ is a projective loop-class that is naturally seen as having a "basepoint" in the space X_∞.

In an analogous way, we can define the *projective fundamental groupoid* $\pi_{\text{path}}^{\text{proj}}(X_\infty)$ of an inverse system of spaces. An element of this groupoid, called a *projective path-class*, is a mapping that assigns to each $\alpha \in \mathcal{A}$ a path-class $[P_\alpha]$ in X_α, subject to the consistency condition $\psi_*^{\beta\leftarrow\alpha}[P_\alpha] = [P_\beta]$; such a path-class has well-defined "endpoints" $\text{Lim}_\alpha\, P_\alpha(0)$ and $\text{Lim}_\alpha\, P_\alpha(1)$ in X_∞.

We say that the inverse system $\{X_\alpha\}$ is *projectively connected* if for all $x, x' \in X_\infty$ there is a projective path-class with endpoints x and x'. (The phrase "projectively homotopically path-connected" is more descriptive but unwieldy.) In order for X_∞ to be projectively connected, it is necessary that all the spaces X_α be path-connected, but this condition alone is not sufficient; consider for instance the solenoid obtained as the inverse limit of $S^1 \leftarrow S^1 \leftarrow S^1 \leftarrow ...$ under the doubling map. On the other hand, in order for X_∞ to be projectively connected it is not necessary for X_∞ itself to be path-connected; indeed, for all the examples we will study X_∞ is a totally disconnected set.

It is not hard to show that if $\{X_\alpha\}$ is projectively connected, then the group $\pi_1^{\text{proj}}(X_\infty, x)$ is independent of the basepoint $x \in X_\infty$. In this case, we write the projective fundamental group as simply $\pi_1^{\text{proj}}(X_\infty)$.

If we have two inverse systems $\{X_\alpha\}$, $\{Y_\alpha\}$ along with maps $X_\alpha \to Y_\alpha$ that commute with all the maps within each inverse system, there is an induced map ψ from $\mathrm{Lim}_\alpha X_\alpha$ to $\mathrm{Lim}_\alpha Y_\alpha$, which in turn induces a group homomorphism from $\pi_1^{\mathrm{proj}}(X, x)$ to $\pi_1^{\mathrm{proj}}(Y, \psi(x))$. Moreover, if $\{X_\alpha\}$ is projectively connected, then so is $\{Y_\alpha\}$. In any case, if the maps $X_\alpha \to Y_\alpha$ are covering maps, then every projective path-class in Y_∞ lifts to a projective path-class in X_∞.

4 Subshifts and Albums

The scene-spaces $\mathcal{S}_B$ introduced in Section 1, along with the restriction maps $\mathcal{S}_{B'} \to \mathcal{S}_B$ $(B' \supseteq B)$, form an inverse system $\{\mathcal{S}_B\}$ of topological spaces. It is easy to check that the inverse limit of this system is just $\mathcal{S}$, endowed with its original topology. (One can think of $\mathcal{S}$ as "projective scene-space".) The spaces $\mathcal{S}_B$ may be non-Hausdorff, but we can still define paths, loops, etc. in all these spaces, and thereby define, for every $(x, f) \in \mathbf{R}^2 \times \mathcal{F} = \mathcal{S} = \mathrm{Lim}_B \mathcal{S}_B$, the fundamental group

$$\pi_1^{\mathrm{proj}}(\mathcal{S}, (x, f)) = \mathrm{Lim}_B\, \pi_1(\mathcal{S}_B, p_B(x, f))$$

(where p_B is the canonical projection $\mathcal{S} \to \mathcal{S}_B$). To avoid congestion of notation, we will often write $\pi_1(\mathcal{S}_B, (x, f))$ instead of $\pi_1(\mathcal{S}_B, p_B(x, f))$.

Given a subshift $\mathcal{F}$, viewed as a set of maps from $\mathbf{Z}^2$ to S, we can obtain other subshifts by pre-composing a map from $\mathbf{Z}^2$ to S with an automorphism of the group $\mathbf{Z}^2$. This is called a reparametrization of the $\mathbf{Z}^2$-shift. The reparametrization changes the geometry of the scene-spaces but not their topology nor the way they map to each other, and so has no effect on π_1^{proj}.

That is, the projective fundamental group is invariant under reparametrization. Also note that if we replace $\mathbf{Z}^2$ by a finite-index subgroup (also isomorphic to $\mathbf{Z}^2$), the resulting $\mathbf{Z}^2$-action must have the same fundamental group as the original.

We now consider invariance of π_1^{proj} under topological conjugacy. Take sets $D, D' \subseteq \mathbf{R}^2$, and let $\mathcal{F}, \mathcal{F}'$ be families of functions having these respective domains. A mapping $\psi : \mathcal{F} \to \mathcal{F}'$ is uniformly continuous (with respect to the metric defined in Section 1) if and only if there exists a bounded set $A \subset \mathbf{R}^2$ (the *aperture of continuity*) such that as f varies over $\mathcal{F}$, the value of $\psi(f) \in \mathcal{F}'$ at a point $x \in D'$ is determined by the values taken by f on the set $(A + x) \cap D$. If ψ is uniformly continuous and surjective we say that $\mathcal{F}'$ is a *factor* of $\mathcal{F}$.

Factor Lemma: *A factor of a projectively connected album is projectively connected.*

Proof: Suppose $\psi : \mathcal{F} \to \mathcal{F}'$ is a uniformly continuous surjection with aperture of continuity A. Then ψ induces a continuous map $\mathcal{S}_{A+B}(\mathcal{F}) \to \mathcal{S}_B(\mathcal{F}')$ (for all bounded sets B) and thus gives a homomorphism from $G_{A+B} = \pi_1(\mathcal{S}_{A+B}(\mathcal{F}))$ to $G'_B = \pi_1(\mathcal{S}_B(\mathcal{F}'))$ (with tacit basepoints $p_{A+B}(x, f)$ and $p'_B(x, \psi(f))$). If $\mathcal{F}$ is projectively connected, then the results at the end of Section 3 guarantee that $\mathcal{F}'$ must be projectively connected as well. $\square$

If the uniformly continuous surjection $\psi : \mathcal{F} \to \mathcal{F}'$ is invertible and its inverse ψ^{-1} is uniformly continuous, we call ψ an *isomorphism* of albums, and say that $\mathcal{F}'$ is isomorphic to $\mathcal{F}$. It is easily checked that the property of being of finite type is preserved under isomorphism.

Theorem 1: *The projective fundamental group of a* $\mathbf{Z}^2$*-shift is invariant under topological conjugacy.*

Proof: Since a conjugacy between $\mathbf{Z}^2$-shifts yields an isomorphism between the associated albums, it will suffice to prove a more general result, namely, that the projective fundamental group of an album is invariant under album isomorphism.

Consider albums $\mathcal{F}$, $\mathcal{F}'$, isomorphic via ψ. By the Factor Lemma, if either of $\mathcal{F}$, $\mathcal{F}'$ is projectively connected then so is the other. Suppose ψ has aperture of continuity A and ψ^{-1} has aperture of continuity A'. We get a homomorphism from G_{A+B} to G'_B and from $G'_{A'+B}$ to G_B for all bounded sets B. Consider the composite inverse system given by the groups G_B and G'_B (with B varying over the bounded subsets of $\mathbf{R}^2$), together with all the maps $G_{B'} \to G_B$, $G'_{B'} \to G'_B$ ($B' \supseteq B$) and all the maps $G_{A+B} \to G'_B$, $G'_{A'+B} \to G_B$. Note that the subsystems $\{G_B\}$ and $\{G'_B\}$ satisfy the mutual domination condition discussed in Section 3, so the respective inverse limits coincide. Hence, $\pi_1^{\text{proj}}(\mathcal{S}(\mathcal{F}))$ and $\pi_1^{\text{proj}}(\mathcal{S}(\mathcal{F}'))$ either are isomorphic to one another (the projectively connected case) or else are both undefined. $\square$

Theorem 1 assures us that the projective fundamental group of a subshift is an intrinsic topological property of the dynamical system, and does not depend on any specific symbolic encoding of the subshift. One is therefore entitled to ask for a definition of the fundamental group that does not make use of a specific album. This can easily be done, at least for compact systems. If X is a compact space on which $\mathbf{Z}^2$ acts, and $\sim$ is an equivalence relation on X whose finitely-many equivalence classes are both closed and open, then we may define $\mathcal{S}_\sim$ in an obvious way. These scene-spaces form a directed system that includes the system of $\mathcal{S}_B$'s, and we may obtain the same inverse limit as before. Verification is routine.

The proof of Theorem 1, with minor alterations, yields a lemma that is useful for computing π_1^{proj} in many concrete cases.

Sequence Lemma: *Let $B_1 \subseteq B_2 \subseteq ...$ be a cofinal sequence of nested bounded subsets of $\mathbf{R}^2$, i.e. suppose every bounded subset of $\mathbf{R}^2$ is a subset of some B_n. Then G_∞ is isomorphic to the inverse limit of $\{G_{B_n} : 1 \leq n < \infty\}$.*

Proof: By cofinality, the mutual domination condition is satisfied, whence the sub-system has the same inverse limit as the full system. □

There is another lemma that will help us to compute the projective fundamental group for specific albums $\mathcal{F}$. To state it, we need a definition. Let $n\mathbf{Z}^2$ denote the lattice $n\mathbf{Z} \times n\mathbf{Z} \subset \mathbf{R}^2$. Say that a path in $\mathbf{R}^2$ is *n-straight* if it consists of edges of length n, each of which joins two nearest neighbors in $n\mathbf{Z}^2$, and if no edge in the path occurs twice or more in immediate succession (that is, no edge is traversed in one direction and then immediately re-traversed in the other direction). Say that a path in a scene-space $\mathcal{S}_B$ is n-straight if its projection into $\mathbf{R}^2$ is n-straight.

Straightening Lemma: *Fix a basepoint $(x, f) \in \mathcal{S}$, let g be any element of $G_\infty = \pi_1^{\text{proj}}(\mathcal{S}, (x, f))$, let B be a bounded subset of $\mathbf{R}^2$, and let n be a fixed integer. Then g_B (an element of $\pi_1(\mathcal{S}_B, (x, f))$) has a representative (that is, a loop in $\mathcal{S}_B$) that is n-straight.*

Proof: Let B' be a square in $\mathbf{R}^2$ containing B, such that the exterior of B' in $\mathbf{R}^2$ is at (L^1) distance n from B. Since g_B lifts to G_∞, it lifts to $G_{B'}$; let L be a loop in $\mathcal{S}_{B'}$ whose projection down to $\mathcal{S}_B$ is in the homotopy class g_B, with projection $(x_t : 0 \leq t \leq 1)$ in $\mathbf{R}^2$. For all t, let y_t be the point in $n\mathbf{Z}^2$ closest to x_t. Then one can use the points y_t to construct a loop $(z_t : 0 \leq t \leq 1)$ in $\mathbf{R}^2$ that is n-straight and stays within distance n of x_t (so that $z_t + B$ is a subset of $x_t + B'$). For each t, let $\overline{f}_t$ be the scene that is inside $z_t + B$ (according to the scenery associated with the loop L); then $((z_t, \overline{f}_t) : 0 \leq t \leq 1)$ is an n-straight representative of g_B. □

It can be useful to have a concrete picture of a scene-space $\mathcal{S}_B$ as a 2-complex. This we now describe.

Suppose that the domain D of an album $\mathcal{F}$ of finite type is $\mathbf{Z}^2$. Then by a suitable block-encoding (as in [MaPa]), we can construct an album isomorphic to $\mathcal{F}$ in which the sets U_i in the definition of finite type are the integer translates of the horizontal segment joining $(0,0)$ and $(1,0)$ together with the integer translates of the vertical segment joining $(0,0)$ and $(0,1)$. Without loss of generality, we may assume we are dealing with such a recoded $\mathcal{F}$. We also wish to impose an "extension property," which may entail some loss of generality. Specifically, we wish to assume that for all s,t,u,v in the symbol-set S and all i,j in $\mathbf{Z}$, if there exist f_1,f_2,f_3,f_4 in $\mathcal{F}$ such that

$$\begin{aligned} f_1(i,j) &= s, & f_1(i+1,j) &= t,\\ f_2(i,j+1) &= u, & f_2(i+1,j+1) &= v,\\ f_3(i,j) &= s, & f_3(i,j+1) &= u,\\ f_4(i+1,j) &= t, & f_4(i+1,j+1) &= v, \end{aligned}$$

then there exists $f \in \mathcal{F}$ such that

$$f(i,j)=s,\ f(i+1,j)=t,\ f(i,j+1)=u,\ f(i+1,j+1)=v.$$

All of the examples we study in this article have the extension property. We will assume throughout the rest of this section that the property holds.

Now suppose that the aperture B is a square of the form $(-m,m) \times (-m,m)$. By once again passing to a higher block presentation of the album, we may without loss of generality restrict ourselves to the case $m=1$. In this situation, the scene-space $\mathcal{S}_B$ admits a simple description as an infinite collections of squares with some identifications along their boundaries, in the spirit of [Thur]. More precisely, $\mathcal{S}_B$ is an infinite 2-complex with vertex set

$\mathbf{Z}^2 \times S$, with an edge $E_{i,j}^{s,t}$ joining $((i,j),s)$ to $((i+1,j),t)$ when there exists an $f \in \mathcal{F}$ satisfying

$$f(i,j) = s,\ f(i+1,j) = t,$$

with an edge ${E'}_{i,j}^{s,t}$ joining $((i,j),s)$ to $((i,j+1),t)$ when there exists an $f \in \mathcal{F}$ satisfying

$$f(i,j) = s,\ f(i,j+1) = t,$$

and with a face $F_{i,j}^{s,t,u,v}$ spanning $E_{i,j}^{s,t}$, ${E'}_{i+1,j}^{t,v}$, $E_{i,j+1}^{u,v}$, ${E'}_{i,j}^{s,u}$ when there exists an $f \in \mathcal{F}$ satisfying

$$f(i,j) = s,\ f(i+1,j) = t,\ f(i,j+1) = u,\ f(i+1,j+1) = v.$$

(If this is confusing, it may be helpful to consider the analogous but simpler state of affairs that prevails in the context of $\mathbf{Z}$-actions; there the scene-space is a graph rather than a 2-complex.)

Choose a point in $\mathbf{Z}^2 \times S$, and let L be a loop in $\mathcal{S}_B$ based at the corresponding point in $\mathcal{S}_B$; the loop may be perturbed so that it only travels along the 1-skeleton, for wherever the loop travels through the interior of one of the squares, one can perturb it so that it hugs the boundary. Indeed, further perturbations insure that the loop is composed of edges (that is, no edges need be partially traversed and then retraced backwards); and we may insist that no edges may be fully traversed and then immediately retraced backwards, since such "trivial excursions" can be homotopically shrunk away. Hence, every loop in $\mathcal{S}_B$ is homotopic to one that is 1-straight. (*Note:* Unlike the Straightening Lemma, this fact holds true whether or not the loop lifts to the inverse limit.) A homotopy between 1-straight loops is generated by "elementary homotopies," wherein a portion of a loop that traverses part of the boundary of a square is "pulled through the square" so that it traverses the complementary portion of the boundary, and any resulting trivial excursions are excised. In many cases, this point of view furnishes a fully combinatorial definition of $\pi_1(\mathcal{S}_B)$ that is amenable to analysis with the tools of combinatorial group theory; see [CoLa] and [Thur].

If instead of taking the window B to be open, as above, we take it to be closed, the scene-space $\mathcal{S}_B$ is not a Hausdorff space. It is nevertheless in some cases a better model to use than the 2-complex constructed above; see the proof of Theorem 4.

We end this section with some remarks on the relation between connectedness properties of scene-spaces (and of inverse limits) and mixing properties of dynamical systems.

Suppose that the album $\mathcal{F}$ has domain $D = \mathbf{Z}^2$ and is closed under integer translations. We claim that if the $\mathbf{Z}^2$-action on $\mathcal{F}$ is topologically mixing, then $\mathcal{S}_B$ is connected. For, fix $x_1, x_2 \in \mathbf{R}^2$, $f_1, f_2 \in \mathcal{F}$. We wish to find a path in $\mathcal{S}_B$ from $p_B(x_1, f_1)$ to $p_B(x_2, f_2)$. Let f^* be an arbitrary function in $\mathcal{F}$. Since $\mathcal{F}$ is mixing, there exists a function $f_1' \in \mathcal{F}$ that agrees with f_1 on $B + x_1$ and with $T_{x^*} f^*$ on $B + x^*$, provided only that x^* is sufficiently far away from x_1. Similarly, there exists a function $f_2' \in \mathcal{F}$ that agrees with f_2 on $B + x_2$ and with $T_{x^*} f^*$ on $B + x^*$, provided only that x^* is sufficiently far away from x_2. Hence, taking x^* to be suitably far from both x_1 and x_2, we may find f_1' and f_2' as above. But then there is a path from $p_B(x_1, f_1) = p_B(x_1, f_1')$ to $p_B(x^*, f_1') = p_B(x^*, T_{x^*} f^*)$ on the f_1'-section of $\mathcal{S}_B$, and a path from $p_B(x^*, T_{x^*} f^*) = p_B(x^*, f_2')$ to $p_B(x_2, f_2') = p_B(x_2, f_2)$ on the f_2'-section of $\mathcal{S}_B$.

On the other hand, suppose the action of $\mathbf{Z}^2$ on $\mathcal{F}$ is imprimitive with respect to some proper sub-lattice Λ of $\mathbf{Z}^2$ (that is, suppose the action of $\mathbf{Z}^2$ on $\mathcal{F}$ admits as a factor the action of $\mathbf{Z}^2$ on Λ; this is the analogue of periodicity for $\mathbf{Z}^2$-actions). For instance, $\mathcal{F}$ could be the set of functions $\phi : \mathbf{Z}^2 \to \mathbf{Z}$ subject to the constraint that $\phi(u) - \phi(v) = \pm 1$ when u, v are adjacent in $\mathbf{Z}^2$, with Λ the tilted sublattice of index 2. (These functions will play a key role in the analysis of the square ice model.) $\mathcal{F}$ splits into two components, according to whether the function ϕ takes an even value or an odd value at the origin, and $\mathcal{S}_B$ splits up into two path-components accordingly. In general, $\mathcal{S}_B$ splits up into many path-components, one (or more) for each element of $\mathbf{Z}^2/\Lambda$. This will be true for all (sufficiently large) B, so that we obtain, in the inverse limit, a decomposition of $\mathcal{S}$ into projective path-class components.

5 The Full Shift

Theorem 2: *The full n-shift has trivial fundamental group.*

Proof: Let $\mathcal{F}$ be the album corresponding to a full shift on an alphabet of size n, one of whose symbols we call 0. To prove that π_1^{proj} of $\mathcal{F}$ is trivial, we must first show that $\mathcal{S}$ is projectively connected. This accomplished, it suffices to prove that for $B = B_m = (-m, m) \times (-m, m)$, an arbitrary free loop L in $\mathcal{S}_B$ can be shrunk down to a point. For this implies that $\pi_1(\mathcal{S}_B)$ (the homotopy group of paths with a fixed basepoint) is trivial for all such B's, and by the Sequence Lemma, it follows that the inverse-limit group $G_\infty = \pi_1^{\text{proj}}$ is also trivial.

To show that the inverse system $\{\mathcal{S}_B\}$ is projectively connected, let f_0 be the constant function sending everything in $\mathbf{Z}^2$ to 0, let f be some arbitrary function in $\mathcal{F}$, and let f_m be the function that agrees with f on the square block $B_m \cap \mathbf{Z}^2 = \{-m+1, ..., m-1\} \times \{-m+1, ..., m-1\}$ and has 0's everywhere else. (In Figure 5, we consider a specific example with $m = 2$ and $f(x, y) = 1 + |x| + |y|$ for $-2 \leq x, y \leq 2$.) Then we can define a path in $\mathcal{S}_B$ that moves from $(0, 0)$ to $(3m, 0)$ along the f_m-section (see Figure 5(a)) and moves from $(3m, 0)$ back to $(0, 0)$ along the f_0-section (see Figure 5(b)). Call the combined path P_m. We need to check that as m varies, these paths are homotopically consistent with each other. That is, if $m' > m$, we need to check that the path $P_{m'}$ in $\mathcal{S}_{B_{m'}}$ projects to a path in $\mathcal{S}_{B_m}$ that is homotopic to P_m (Figure 5(c) shows $P_{m'}$ in the case $m' = 3$). To this end, let $f_{m',m}$ be the function that agrees with f on $\{-m'+1, ..., m'-1\} \times \{-m+1, ..., m-1\}$ and has 0's everywhere else. Then the image of $P_{m'}$ in $\mathcal{S}_{B_m}$ can be seen as taking its outbound leg (shown in Figure 5(d)) on the $f_{m',m}$-section rather than the $f_{m'}$-section. We can perturb this outbound path so that its projection in $\mathbf{R}^2$, instead of going directly from $(0,0)$ to $(3m',0)$, goes via $(0,3m)$ and $(3m',3m)$, in the obvious rectilinear fashion, as shown in Figure 5(e). Now the outbound path lives on the f_m-section (as well as the $f_{m',m}$-section), and we can undo the detour, obtaining a direct path from $(0,0)$ to $(3m',0)$, as shown in Figure 5(f). The resulting closed path is homotopic to P_m, since we can shrink away the portion of the path that goes from $(3m,0)$ to $(3m',0)$ and back. Hence $P_{m'}$ does indeed project to P_m (up to homotopy). In the inverse limit, we obtain a projective path-class joining $((0,0),f_0)$ and $((0,0),f)$ in $\mathcal{S}$. Since f was arbitrary, projective connectivity follows. Thus the fundamental group is well-defined (i.e., does not depend on choice of basepoint).

Next fix $B = (-m,m) \times (-m,m)$; we wish to show that an arbitrary free loop L in $\mathcal{S}_B$ can be contracted. By the remarks made near the end of Section 4, we may suppose that our loop L is 1-straight, so that its projection into $\mathbf{R}^2$ is a rectilinear path $x_0, x_1, \cdots, x_n = x_0$, where the x_i's are points in $\mathbf{Z}^2$ such that $||x_i - x_{i+1}|| = 1$ for all i. Let $\overline{f}_i$ denote the scene in $x_i + B$. For each i, we choose a point y_i in $\mathbf{Z}^2$, such that $x_i + B$ is disjoint from $y_i + B$, and such that the line joining x_i to y_i is either horizontal or vertical and has only the point x_i in common with the unit segment joining x_{i-1} to x_i and the unit segment joining x_i to x_{i+1}. See Figure 6, in which the straight undashed segments form an excerpt of the loop L.

We now define a new loop L' homotopic to L, obtained by adding contractible loops based at each point $(x_i, \overline{f}_i)$ along the loop. For each i, let f_i be the function in $\mathcal{F}$ whose restriction to $x_i + B$ is $\overline{f}_i$ and whose values outside of $x_i + B$ are all 0's. Then there is a unique path on the f_i sheet of $\mathcal{S}$ whose projection down to $\mathbf{R}^2$ goes from x_i to y_i via a straight line and then returns to x_i along the same route. The projection of this contractible loop into $\mathcal{S}_B$ is also contractible, so we can insert it into L at the point $(x_i, \overline{f}_i)$ without changing the homotopy class of L. The undashed segments shown in Figure 6, both straight and curved, illustrate the resulting loop L'. (Note that the paths from x_i to y_i and back are drawn as arcs, even though they are in fact straight lines; were we to draw them in a more literal-minded way, the two paths would coincide.)

Consider now the part of L' whose projection down in $\mathbf{R}^2$ goes from y_i to x_i to x_{i+1} to y_{i+1}. It lifts to a path in $\mathcal{S}$, where it lives on a sheet that shows the scene $\overline{f}_i$ on $x_i + B$, the scene $\overline{f}_{i+1}$ on $x_{i+1} + B$, and 0's everywhere else. This path can be perturbed so that it skirts $(x_i + B) \cup (x_{i+1} + B)$, and thus shows only 0's in its scenery. This homotopic perturbation projects down to $\mathcal{S}_B$. If we do this for all i, we get a new loop L'', depicted by dashed edges in Figure 6. Observe that the loop L'' lives on the all-0's section of $\mathcal{S}_B$, and thus is contractible. $\square$

It's worthwhile to point out that if $\mathcal{S}_B$ is not a square block, $\mathcal{S}_B$ can fail to be simply connected. If for instance

$$B = ([-3/2, 3/2] \times [-3/2, 3/2]) \setminus ([-1/2, 1/2] \times [-1/2, 1/2])$$

(a 3-by-3 square with a hole in the middle), then $\pi_1(\mathcal{S}_B)$ is not even finitely generated.

To see why this makes intuitive sense, first consider a path $((x_t, f_t) : 0 \leq t \leq 1)$, in which x_t travels at uniform speed from (0,0) to (4,0) to (4,4) to (0,4) to (0,0), and $f_t(y) = 0$ for all t between 0 and 1 for all $y \in \mathbf{Z}^2 \cap (B + x_t)$, *except* that $f_t((2,0)) = 1$ for $5/32 \leq t \leq 7/32$ (which is "inconsistent" with having $f_t((2,0)) = 0$ for $1/32 \leq t \leq 3/32$). That is, the aperture slides four units to the right, such that the scenery at the point (2,0) is initially invisible, then a 0, then invisible, then a 1, then invisible. The usual tricks for shrinking a loop do not work here; in fact, it can be shown that there is no way to get rid of the local "kink" at the point (2,0). We can play this game with any finite set of points simultaneously, and devise a path which is tied to all of them. In this way we see that $\pi_1(\mathcal{S}_B)$ (for this particular B) is quite complicated.

Of course, none of these bad paths lift to paths in $\mathcal{S}_{B'}$, where B' is the 3-by-3 square (that is, B with its missing middle restored), since we know that $\pi_1(\mathcal{S}_{B'})$ is trivial. Here we see how taking the inverse limit eliminates unwanted complexities that can arise from individual scene-spaces.

The method introduced in this section can be used to show that many two-dimensional shifts of finite type have trivial fundamental group. For instance, consider the set of proper 4-colorings of $\mathbf{Z}^2$ as a $\mathbf{Z}^2$-shift, where a proper 4-coloring of $\mathbf{Z}^2$ is a map from the grid-graph $\mathbf{Z}^2$ to the color-set $\{0, 1, 2, 3\}$ with the property that no two adjacent vertices of the graph are assigned the same color. Let f_0 be either of the two proper 4-colorings of $\mathbf{Z}^2$ that use only the colors 0 and 1, in alternating chessboard fashion. Then it is not hard to use f_0 to mimic the proof of Theorem 2 by demonstrating both the fact that the system is projectively connected and the fact that an arbitrary free loop in the scene-space is contractible. For instance, the latter fact is proved in much the same manner as in Theorem 2, by replacing each scene $(x_i, \overline{f}_i)$ on the path L by a suitable (nearby and compatible) scene $(y_i, \overline{f}_0)$, and "pushing" the scenes on the path onto the f_0-sheet. The needed trick, due to Klaus Schmidt [Sch2], is the observation that, given a proper 4-coloring of a rectangular excerpt of the graph, we can extend the coloring along one edge of the rectangle in such a way that each vertex along the new frontier is given a color that is congruent mod 2 to its f_0-color; this is because each new vertex has only one old vertex constraining it, and the two new vertices that neighbor it do not constrain it once we have committed ourselves to giving those vertices colors of opposite parity. This new frontier can in turn be extended to give a rectangle whose leading edge is colored precisely as in f_0. (Figure 7 shows the scheme, using the same layout as Figure 6.)

This argument works equally well for k-colorings for any $k \geq 4$, but we will see below (see Section 7) that the situation is very different for $k = 3$.

6 The Square Ice Model

Theorem 3: *The square ice model has fundamental group* $\cong \mathbf{Z}$.

Proof: Since the fundamental group is invariant under album isomorphism, we may replace square ice with the dual square ice model introduced in Section 2. Let $\mathcal{F}$ be this dual model on the square grid G (with vertex set $\mathbf{Z}^2$), encoded in some fashion as a set of maps with domain $D \subseteq \mathbf{R}^2$.

Let $\hat{\mathcal{F}}$ be the set of functions ϕ from $\mathbf{Z}^2$ to $\mathbf{Z}$, subject to the constraints that (a) $\phi(i,j) \equiv i + j \pmod 2$ for all $(i,j) \in \mathbf{Z}^2$, and (b) $\phi(u) - \phi(v) = \pm 1$ when u, v are adjacent in $\mathbf{Z}^2$. Given $\phi \in \hat{\mathcal{F}}$, define $\partial\phi$ to be the orientation of the grid G in which the edge uv is oriented from u to v if and only if $\phi(v) - \phi(u) = +1$. It is easily shown that ∂ sends $\hat{\mathcal{F}}$ to $\mathcal{F}$. For, let u, v, w, x be vertices of a 1-by-1 square in G, taken in cyclic order; since $[\phi(u) - \phi(v)] + [\phi(v) - \phi(w)] + [\phi(w) - \phi(x)] + [\phi(x) - \phi(u)] = 0$, two of the bracketed expressions will be $+1$'s and two will be -1's, so that two of the edges of the square acquire clockwise orientation and the other two acquire counterclockwise orientation. Conversely, every square ice configuration lifts to a function $\phi \in \hat{\mathcal{F}}$. Two functions $\phi, \phi' \in \hat{\mathcal{F}}$ satisfy $\partial\phi = \partial\phi'$ if and only if $\phi - \phi'$ is a constant c (necessarily even).

The map $\partial : \hat{\mathcal{F}} \to \mathcal{F}$ is continuous; moreover, it is a covering map, in the sense defined in Section 3. Our determination of $\pi_1^{\text{proj}}(\mathcal{S}(\mathcal{F}))$ will hinge on the determination of $\pi_1^{\text{proj}}(\mathcal{S}(\widehat{\mathcal{F}}))$.

To show that $\mathcal{F}$ is projectively connected, it suffices to show that $\widehat{\mathcal{F}}$ is projectively connected (see the remarks made in the last paragraph of Section 3). Our proof is modeled on that of the last section, but we need one extra idea. Given a partial function ϕ whose domain is a rectangle in $\mathbf{Z}^2$, whose range is in $\mathbf{Z}$, and which satisfies conditions (a) and (b) above, we can in an iterative fashion define an extension of ϕ to all of $\mathbf{Z}^2$ that continues to satisfy (a) and (b). If $\phi(i,j)$ is defined for $(i,j) \in [A,B] \times [C,D]$, one can extend to $[A,B] \times [C,D+1]$, by putting

$$\phi(i,D+1) = \begin{cases} \phi(i,D)-1 & \text{if } \phi(i,D) \geq 1, \\ \phi(i,D)+1 & \text{if } \phi(i,D) \leq 0 \end{cases}$$

for all $i \in [A,B]$. One can apply the same trick three more times to get ϕ defined on $[A-1,B+1] \times [C-1,D+1]$. (See Figure 8.) Repeating this ad infinitum, we get ϕ defined on all of $\mathbf{Z}^2$. Moreover, this ϕ takes only the values 0 and 1 outside of some finite region. We call this the natural extension of our partial function. (Incidentally, this idea can be applied to a partial function whose domain is a union of two rectangles, provided that the separation between the rectangles is large compared to the absolute values of the integers appearing in those rectangles; this observation permits one to prove that the action of $\mathbf{Z}^2$ on $\widehat{\mathcal{F}}$ is mixing.)

Let ϕ_0 be the function sending $(i,j) \in \mathbf{Z}^2$ to 0 or 1 according to whether $i+j$ is even or odd, and let ϕ be some other function in $\mathcal{F}$. For all m, let ϕ_m be the natural extension of the restriction of ϕ to the square block $B_m \cap \mathbf{Z}^2 = \{-m+1, ..., m-1\} \times \{-m+1, ..., m-1\}$; and let P_m be a path in $\mathcal{S}_B$ that moves away from $(0,0)$ on the ϕ_m-section (in the direction of $(\infty, 0)$) and returns on the ϕ_0-section. As in the proof of Theorem 2, one can prove that if $m' > m$, the projection of $P_{m'}$ into $\mathcal{S}_{B_m}$ is homotopic with P_m.

To prove that the projective fundamental group of $\mathcal{F}$ is $\mathbf{Z}$, we proceed in two stages. First we show that the projective fundamental group of $\widehat{\mathcal{F}}$ is trivial. Then we show that the monodromy of a loop in $\mathcal{F}$, when lifted to $\widehat{\mathcal{F}}$, acts on the fiber of the basepoint ($\cong 2\mathbf{Z}$) by simple addition of an even integer, and that every even integer arises from the monodromy of some loop — implying that the projective fundamental group of $\mathcal{F}$ is $2\mathbf{Z} \cong \mathbf{Z}$. (Note the reliance on covering-space ideas; these are generalized in Section 7.)

Consider an element of $\pi_1^{\text{proj}}(\mathcal{S}(\widehat{\mathcal{F}}))$, based at $((0,0), \phi_0)$; for each square block B, it projects to a loop-class in $\pi_1(\mathcal{S}_B(\widehat{\mathcal{F}}))$. Let L be a representative of this loop-class. The Straightening Lemma tells us that without loss of generality, we may assume that L is n-straight, where n is much larger than the diameter of the block B. We will show that L is contractible as a free loop. This implies that $[L]$ (the homotopy class of L in $\pi_1(\mathcal{S}_B(\widehat{\mathcal{F}}))$) is trivial. Since B is arbitrary, our (arbitrary) element of $\pi_1^{\text{proj}}(\mathcal{S}(\widehat{\mathcal{F}}))$ is the identity, so that the projective fundamental group of $\widehat{\mathcal{F}}$ is trivial.

Let σ_0 be the ϕ_0-sheet in $\mathcal{S}(\widehat{\mathcal{F}})$, and let $\overline{\sigma}_0$ be its image in $\mathcal{S}_B(\widehat{\mathcal{F}})$. We will show that L can be continuously deformed into a loop that lives in $\overline{\sigma}_0$.

It suffices to consider loops in $\mathcal{S}_B(\widehat{\mathcal{F}})$ such that for all $(x_t, \overline{\phi}_t)$ on the path, $\overline{\phi}_t$ takes on only non-negative values in $x_t + B$. Let M_t be the maximum value taken by $\overline{\phi}_t$ on $x_t + B$, and let M be the maximum of the M_t's. Let $r_{i,j}$ be the number of distinct times t for which $\overline{\phi}_t(i,j) = M$ (where two times t_1, t_2 count as "the same" if $(i,j) \in x_t + B$ for all t between t_1 and t_2), and let $r = \sum_{(i,j)\in \mathbf{Z}^2} r_{i,j} \geq 1$. We will prove our claim about deformation of loops by iteratively reducing M and r.

If $M = 1$, then the loop already lives on $\overline{\sigma}_0$, and we are done. If $M \geq 2$, then we can reduce r (or, if r is already 1, reduce M) in the following way. Find i^*, j^*, t^* with $\phi_{t^*}(i^*, j^*) = M$.

Case I: $(x_{t^*} + B) \cap n\mathbf{Z}^2 = \emptyset$. Since the loop is n-straight, there exist t_1, t_2 with $t_1 < t^* < t_2$ such that x_{t_1}, x_{t_2} are adjacent point in the grid $n\mathbf{Z}^2$ and are consecutive on the loop L, with x_{t^*} lying somewhere between them. For definiteness, suppose $x_{t_1} = (0,0)$, $x_{t_2} = (n, 0)$. The scenes $\overline{\phi}_t$ $(t_1 < t < t_2)$ are all consistent with one another, and thus form a legal rectangular block. Let $\tilde{\phi}$ be the natural extension of this block to a full function from $\mathbf{Z}^2$ to $\mathbf{Z}$. Let σ be the corresponding sheet in $\mathcal{S}(\widehat{\mathcal{F}})$, and $\overline{\sigma}$ the corresponding section in $\mathcal{S}_B(\widehat{\mathcal{F}})$. Let $\phi' \in \widehat{\mathcal{F}}$ be a modified version of $\tilde{\phi}$, with

$$\phi'(i,j) = \begin{cases} \tilde{\phi}(i,j) - 2 = M - 2 & \text{if } (i,j) = (i^*, j^*) \\ \tilde{\phi}(i,j) & \text{otherwise.} \end{cases}$$

(This is legal, since $\tilde{\phi}(v) = M - 1$ for all four neighbors v of (i^*, j^*).) Let σ', $\overline{\sigma}'$ be the corresponding sheet and section in $\mathcal{S}(\widehat{\mathcal{F}})$ and $\mathcal{S}_B(\widehat{\mathcal{F}})$, respectively. Then the segment of the loop between time t_1 and time t_2 lives on $\overline{\sigma}$ and may be perturbed so that its projection into $\mathbf{R}^2$ goes from $(0,0)$ to $(0,n)$ to (n,n) to $(n,0)$. But this path lives on $\overline{\sigma}'$ as well as $\overline{\sigma}$, and thus may be perturbed so that its projection into $\mathbf{R}^2$ once more goes directly from $(0,0)$ to $(n,0)$. The new loop is like the old, except that the value M achieved by $\phi_t(i^*,j^*)$ for t near t^* has been replaced by $M-2$. Thus, in the new path either the maximal M has decreased or the decrease in r_{i^*,j^*} by 1 has led r to decrease by 1.

Case II: $(x_{t^*}+B)\cap n\mathbf{Z}^2 \neq \emptyset$. This is slightly more complicated than Case I; it becomes necessary to consider three consecutive grid-points in nZ^2 along the path. What is more, there are two geometrically distinct sub-cases to consider, according to whether the n-straight path makes a $90°$ turn at the middle grid-point or keeps going straight. The key observation is that we can perturb L on the ϕ-section so that it stays away from the point (i^*,j^*) and then perturb it back on the ϕ'-section. We leave details of the argument to the reader.

In either case, the ordered pair (M,r) has been lexicographically reduced. Performing this operation sufficiently many times, we eventually achieve $M=1$. The loop now lives in $\overline{\sigma}_0$, and hence is contractible.

This completes the proof that $\pi_1^{\mathrm{proj}}(\mathcal{S}(\widehat{\mathcal{F}}))$ is trivial.

Since $\mathcal{S}(\widehat{\mathcal{F}})$ is projectively connected, so is $\mathcal{S}(\mathcal{F})$, and it remains only to find the projective fundamental group of $\mathcal{F}$, which we will do by determining the monodromy on a fiber. Consider an element of the projective fundamental group of $\mathcal{F}$, and fix a block B; we get a loop-class in $\mathcal{S}_B(\mathcal{F})$ from which we may choose some representative path L. Since ∂ is a covering map, we can lift L to a path $\widehat{P}$ in $\mathcal{S}_B(\widehat{\mathcal{F}})$ whose endpoints both lie in some fiber over $\mathcal{S}_B(\mathcal{F})$. Say that $\widehat{P}$ goes from $(x_0, \overline{\phi}_0)$ to $(x_0,\ \overline{\phi}_1 = \overline{\phi}_0 + c)$ with $c \in 2\mathbf{Z}$. Now take some other point $(x_0, \overline{\phi}_0 + d)$ in that fiber, where d is any even constant we choose, and let $\widehat{P}'$ be the lift of L to $\mathcal{S}_B(\widehat{\mathcal{F}})$ with initial point $(x_0, \overline{\phi}_0 + d)$; to verify that the monodromy group is $2\mathbf{Z}$ ($\cong \mathbf{Z}$), we need to know that the endpoint of $\widehat{P}'$ in $\mathcal{S}_B(\widehat{\mathcal{F}})$ is $(x_0, \overline{\phi}_0 + c + d)$. But this is easily shown. For, the path $((x_t, \overline{\phi}_t + d) : 0 \leq t \leq 1)$ projects under ∂ to the same path in $\mathcal{S}_B(\mathcal{F})$ as $\widehat{P} = ((x_t, \overline{\phi}_t) : 0 \leq t \leq 1)$ does, namely P; since $((x_t, \overline{\phi}_t + d) : 0 \leq t \leq 1)$ moreover has initial point $(x_0, \overline{\phi}_0 + d) : 0 \leq t \leq 1)$, it must be the desired path $\widehat{P}'$ and its endpoint is $(x_1, \overline{\phi}_1 + d) = (x_1, \overline{\phi}_0 + c + d)$. $\square$

Interestingly, determining the respective fundamental groups of the individual spaces $\mathcal{S}_B$ seems to be harder than determining the inverse limit of these groups; the Straightening Lemma played a key role in the proof of Theorem 3. In fact, we do not know if any or all of the spaces $\mathcal{S}_B$ with B a square block are simply connected.

7 Covering Maps

As an example of a covering map between two-dimensional subshifts of finite type, let $\mathcal{F} = \mathcal{G} =$ the full 1-dimensional 2-shift with alphabet $\mathbf{Z}/2\mathbf{Z}$, and define $\psi(f) = g$ with $g(i) = f(i) + f(i-1) \pmod 2$. Note that the map ψ is everywhere 2-to-1. To see that it is a covering map, note that for any $g \in \mathcal{G}$ the neighborhood $V = \{g' \in \mathcal{G} : g'(0) = g(0)\}$ of g has the property that its pre-image can be written as the disjoint union of two sets each of which is mapped homeomorphically to V by ψ.

One example of a covering map that comes from statistical mechanics is the two-dimensional Ising model [Baxt]. A state of this model is given by an array of arrows, each of which points either up or down; all configurations are allowed, so we are dealing with a full 2-shift. A sample configuration is shown in Figure 9. We represent an up-arrow by putting a 1 in the associated face of the infinite square grid, and a down-arrow by putting a 0 in that face. The energy of the configuration is the sum, over all pairs of adjacent arrows, of an interaction energy between the two, which takes one value if the two arrows point in the same direction and another value if they point in opposite directions. Thus, physicists often depict such a configuration by drawing edges to mark the boundary between a region of contiguous up-arrows and a region of contiguous down-arrows; the energy of a configuration can thus be determined by counting the marked edges. It is not hard to see that the graphs that arise in this way are precisely those subgraphs of the infinite square grid in which each vertex has degree 0, 2, or 4. The set of such subgraphs forms a $\mathbf{Z}^2$-subshift of finite type, and it is easy to see that the map from the full 2-shift to the new shift is 2-to-1.

A more interesting example, also drawn from statistical mechanics, is given by the set of proper three-colorings of the square grid [Baxt]. This is the set of mappings from the vertices of the infinite square grid to the color-set $\{0, 1, 2\}$ having the property that no two adjacent vertices are assigned the same color. See Figure 10. We can turn this into an edge-marked configuration by giving each edge of the square grid an orientation according to the rule that an edge that joins a vertex marked i to a vertex marked j should be oriented from i to j if and only if $j - i \cong 1 \pmod 3$. It is not hard to check that the orientations of the square grid that arise in this way are precisely those that satisfy the curl-constraint of Section 6. Each such orientation has a pre-image of size 3 in the set of proper three-colorings.

In the Ising model example, the two pre-images of a point in the "downstairs" album are obtained from one another by turning 0's into 1's and vice versa. In the three-coloring example, the three pre-images of a point in the downstairs album are obtained from one another by cyclic shifts of the symbols 0, 1, 2. In each case, two points that are in the pre-image of a single point downstairs have the property of exhibiting different symbols everywhere. This makes it easy to verify that the mapping between the two albums is indeed a covering map – simply take V_y to be the set of all points that agree with y on some arbitrary non-empty subset of $\mathbf{Z}^2$.

For the rest of this section, we will restrict ourselves to the study of $\mathbf{Z}^2$-shifts, rather than general albums. We shall use the term "covering map" as a shorthand for "covering factor map". We will also, for the most part, restrict ourselves to the study of covering maps $\psi : \mathcal{F} \to \mathcal{G}$ for which the $\mathbf{Z}^2$-shift $\mathcal{G}$ is compact. Finally, note that by passing to a higher block presentation of $\mathcal{F}$ if necessary, we can always take ψ to be a 1-block map.

We say that a factor map $\psi : \mathcal{F} \to \mathcal{G}$ between $\mathbf{Z}^2$-shifts is *totally separated* if whenever f and f' belong to the same fiber of ψ we have that $f(i,j) \neq f'(i,j)$ for all $i, j \in \mathbf{Z}$. More generally, if $\psi : X \to Y$ is any map between metric spaces, we call ψ *uniformly separated* if there is some $\delta > 0$ such that whenever x and x' belong to the same fiber of ψ, we have that the distance between x and x' is at least δ. It is clear that a totally separated factor map is uniformly separated. Conversely, if a factor map $\psi : \mathcal{F} \to \mathcal{G}$ between $\mathbf{Z}^2$-shifts is uniformly separated, then for a suitable higher block presentation $\mathcal{F}'$ of $\mathcal{F}$ the corresponding factor map $\psi' : \mathcal{F}' \to \mathcal{G}$ is totally separated.

Proposition 1: *Suppose $\mathcal{G}$ is compact (as a topological space) and topologically transitive (as a dynamical system). Then every covering map $\psi : \mathcal{F} \to \mathcal{G}$ is constant-to-one and uniformly separated.*

Proof: Let d be the smallest cardinality for which the set $G = \{g \in \mathcal{G} : |\psi^{-1}g| = d\}$ is non-empty, so that its complement is $G^c = \{g \in \mathcal{G} : |\psi^{-1}g| > d\}$. Since ψ is a covering map, G^c is open, and since ψ intertwines with the actions of $\mathbf{Z}^2$ on $\mathcal{F}$ and $\mathcal{G}$, G^c is invariant under $\mathbf{Z}^2$. If G^c were non-empty, it would have to comprise all of $\mathcal{G}$ (by topological transitivity), contradicting the fact that G is non-empty. Hence G^c is empty, and every $g \in \mathcal{G}$ has exactly d pre-images in $\mathcal{F}$. To prove uniform separation, give each g a cylinder-set neighborhood that is evenly covered under ψ and pass to a finite subcover (using compactness). For each open set O in the finite cover, $\psi^{-1}(O)$ is a union of d disjoint cylinder sets which are at positive distance from one another. Letting O vary and taking the minimum of these distances, we get a uniform lower bound on the distance between pre-images of points in $\mathcal{G}$. $\square$

As a partial converse, we have:

Proposition 2: *Suppose $\mathcal{F}$ and $\mathcal{G}$ are both compact, with d some positive integer. Then every d-to-1 continuous map $\psi : \mathcal{F} \to \mathcal{G}$ that is uniformly separated is a covering map.*

Proof: We first prove that ψ must be open. For, if ψ were not open, then there would be an open set U in $\mathcal{F}$, a point $f \in U$, and a sequence of points $\{g_n\}$ in $\mathcal{G}$ converging to $g = \psi(f)$ such that for each n, $\psi^{-1}(g_n) = \{f_n^1, \ldots, f_n^d\}$ does not meet U. Since $\mathcal{F}$ is compact, we may pass to a subsequence of n's, which we again index by n, so that $\{f_n^1\}$ converges, say to f^1. Dropping again to a subsequence of these, we may assume that $\{f_n^2\}$ converges to some f^2. In this way we continue refining the the sequence of indices until we obtain that $\{f_n^d\}$ converges to f^d. Then $f^1, \ldots, f^d$ are distinct since $f_n^1, \ldots, f_n^d$ are δ-separated. Also, no f^i is in U since the f_n^i are not in U and U is open. By continuity, we have that $\psi(f^i) = g$. But then $\{f, f^1, \ldots, f^d\} \subset \psi^{-1}(g)$ so $|\psi^{-1}(g)| > d$, a contradiction.

Now, for $g \in \mathcal{G}$, let $\psi^{-1}(g) = \{f_1, \ldots, f_d\}$ and take disjoint open sets $U_1, \ldots, U_d$ in $\mathcal{F}$ with $f_i \in U_i$, using the fact that $\mathcal{F}$ is Hausdorff. Then

$V_i = \psi(U_i)$ is open, and g is in each V_i. Let $V = V_1 \cap \ldots \cap V_d$, so that V is an open neighborhood of y. If we define $W_i = U_i \cap \psi^{-1}(V)$, then the W_i are disjoint open sets in $\mathcal{F}$. Since ψ is (only) d-to-1, $\psi^{-1}(V)$ is just the union of the W_i's. The continuous map ψ, restricted to a map from W_i to V, is bijective and open, hence a homeomorphism. $\square$

It is worth noting that for transitive $\mathbf{Z}$-shifts of finite type, a constant-to-one factor map is necessarily uniformly separated (and so totally separated up to a conjugacy); see [Nasu]. We do not know if this it true in the $\mathbf{Z}^2$ case.

Let ψ be a d-to-1 covering map from $\mathcal{F}$ to $\mathcal{G}$, with $\mathcal{G}$ topologically transitive and compact. ψ induces a map from $\mathcal{S}_B(\mathcal{F})$ to $\mathcal{S}_B(\mathcal{G})$, but this map is, unfortunately, not generally a covering map. To circumvent this problem, we introduce a stand-in for $\mathcal{S}_B(\mathcal{F})$, the "quasi-scene-space" $\mathcal{S}_B^{\psi}(\mathcal{F})$, which will turn out to be a covering space of $\mathcal{S}_B(\mathcal{G})$. Taking these covering maps to the inverse limit will allow us to prove Theorem 4.

Since $\mathcal{G}$ is compact, there is a finite cover of $\mathcal{G}$ by open sets $O_1, ..., O_r$ that are evenly covered by their pre-images in $\mathcal{F}$. Let $\psi_{i,j}$ denote the restriction of ψ to the jth component of the pre-image of O_i. Each $\psi_{i,j}$ is a homeomorphism, so its inverse is a block code relative to some bounded subset $B_{i,j}$ of $\mathbf{R}^2$. Since there are only finitely many $B_{i,j}$'s, there exists a square block B^* of the form $[-m, m] \times [-m, m]$ such that any translate of B^* contains all of the $B_{i,j}$'s.

Note that every translate of B^* must intersect $\mathbf{Z}^2$. By making B^* even larger, if necessary, we can ensure that for every $x \in \mathbf{R}^2$, every cylinder set associated with a particular $(B^* + x)$-scene is evenly covered (use the open cover $O_1, ..., O_r$).

Given a bounded set B in $\mathbf{R}^2$, define B^- as the set of all points x in $\mathbf{R}^2$ for which the translate $B^* + x$ is a subset of B. We assume henceforth that B is sufficiently large that B^- contains B^*, and that B is closed, so that B^- is closed as well. Define the quasi-scene-space $\mathcal{S}_B^\psi(\mathcal{F})$ as the set of pairs (x, f) under the equivalence relation that puts $(x, f) \sim (x, f')$ if and only if f and f' agree on B^- and ψf and $\psi f'$ agree on B.

Note that there are "restriction maps" from $\mathcal{S}_B$ to $\mathcal{S}_B^\psi$, and from $\mathcal{S}_B^\psi$ to $\mathcal{S}_{B^-}$. Also note that if the sequence of B_n's is cofinal then the sequence of B_n^-'s is cofinal as well. Hence, to specify an element of the inverse limit of $\mathcal{S}_B$ or of $\pi_1(\mathcal{S}_B)$ (as B increases up to $\mathbf{R}^2$), it suffices to specify an element of the inverse limit of $\mathcal{S}_B^\psi$ or of $\pi_1(\mathcal{S}_B^\psi)$.

Since each B-quasi-scene in $\mathcal{F}$ determines a B-scene in $\mathcal{G}$, there is a map ψ_B from $\mathcal{S}_B^\psi(\mathcal{F})$ to $\mathcal{S}_B(\mathcal{G})$. The maps ψ_B commute with all relevant restriction maps, and hence induce a map ψ_* from projective path classes in $\mathcal{S}(\mathcal{F})$ to projective path classes in $\mathcal{S}(\mathcal{G})$. An easy diagram-chase tells us that ψ_* of a non-trivial projective path class in $\mathcal{S}(\mathcal{F})$ is a non-trivial projective path class in $\mathcal{S}(\mathcal{G})$; hence, ψ_* is injective.

Proposition 3: *If ψ is a d-to-1 totally separated covering map, then ψ_B is a d-to-1 covering map.*

Proof: First we will show that ψ_B is d-to-1. Fix $(x, \overline{g})$ in $\mathcal{S}_B(\mathcal{G})$, where $\overline{g}$ denotes the restriction of some particular function $g \in \mathcal{G}$ to $B + x$. Since ψ is d-to-1, there are exactly d functions $f_1, ..., f_d$ that ψ maps to g, and since ψ is totally separated, no two of these functions f_i have the same symbols anywhere. It follows that the functions f_i determine d distinct pre-images of the scene $(x, \overline{g})$ in the quasi-scene-space $\mathcal{S}_B^\psi(\mathcal{F})$. To show that there are no other pre-images, note that since $B \supset B^*$, the cylinder set V consisting of those elements of $\mathcal{G}$ that agree with g on $B + x$ is a subset of the cylinder set associated with a $B^* + x$-scene, and hence is evenly covered under ψ; that is, $\psi^{-1}(V)$ consists of d open sets U_i each of which is mapped onto V homeomorphically by ψ, and each of which contains a unique f_i. We have d local inverses of ψ, taking V to the respective U_i's; write the ith local inverse as ψ_i^{-1}. Since all the points in V agree with g on B, and since each ψ_i^{-1} is a B^*-block map, every point in $U_i = \psi_i^{-1} V$ agrees with f_i on B^-. Hence, the quasi-scenes that comes from the f_i's are the only ones that are in the pre-image of $(x, \overline{g})$, and so ψ_B is d-to-1.

We now consider scenes in $\mathcal{S}_B(\mathcal{G})$ that are close to $(x, \overline{g})$. Let $\overline{g}_{x'}$ denote the restriction of g to $(B + x') \cap \mathbf{Z}^2$, where x' is a point in $\mathbf{R}^2$ near x. Choose $\epsilon > 0$ such that for all x' within distance ϵ of x, $B + x'$ contains $B^- + x$ but does not contain any points of $\mathbf{Z}^2$ not present in $B + x$ (recall that B is closed). Since B^- contains B^*, the first condition implies that the set of functions in $\mathcal{G}$ that agree with g on $B + x'$ is evenly covered under ψ, and the local inverses ψ_i^{-1} introduced above apply to the perturbed scenes $(x, \overline{g}_{x'})$ as well. The second condition implies that the scene $\overline{g}_{x'}$ reveals no symbols not shown by the scene $\overline{g}_x$ (though it may obscure some of the symbols that were shown by $\overline{g}_x$). Hence, the pre-image of $(x', \overline{g}_{x'})$ under ψ_B consists of d quasi-scenes associated with the respective f_i's. (Total separation tells us that these quasi-scenes are distinct.) The set $\{(x', \overline{g}_{x'}) \ : \ |x' - x| < \epsilon\}$ is therefore a neighborhood of $(x, \overline{g}_x)$ in $\mathcal{S}_B(\mathcal{G})$ that is evenly covered. $\square$

Proposition 4: *Suppose $f \in \mathcal{F}$, $g \in \mathcal{G}$ are such that $\psi f = g$. Fix $x \in \mathbf{R}^2$. Then every projective path class in $\mathcal{S}(\mathcal{G})$ with initial point (x, g) lifts to a unique projective path class in $\mathcal{S}(\mathcal{F})$ with initial point (x, f).*

Proof: We first verify uniqueness. A projective path class in $\mathcal{S}(\mathcal{G})$ with initial point (x, g) determines a path class $[Q_B]$ in the scene-space $\mathcal{S}_B(\mathcal{G})$ whose initial point is the image of (x, g) in $\mathcal{S}_B(\mathcal{G})$; such a $[Q_B]$, by virtue of Proposition 3, must lift to the unique path class $[P_B]$ in the quasi-scene-space $\mathcal{S}_B^{\psi}(\mathcal{F})$ whose initial point is the image of (x, g) in $\mathcal{S}_B^{\psi}(\mathcal{F})$. Since the diagram

$$\begin{array}{ccc} \pi_{\text{path}}(\mathcal{S}_B^{\psi}(\mathcal{F})) & \leftarrow & \pi_{\text{path}}^{\text{proj}}(\mathcal{S}(\mathcal{F})) \\ \downarrow & & \downarrow \\ \pi_{\text{path}}(\mathcal{S}_B(\mathcal{G})) & \leftarrow & \pi_{\text{path}}^{\text{proj}}(\mathcal{S}(\mathcal{G})) \end{array}$$

commutes, any putative lift of the original projective path class in $\pi_{\text{path}}^{\text{proj}}(\mathcal{S}(\mathcal{G}))$ up to $\pi_{\text{path}}^{\text{proj}}(\mathcal{S}(\mathcal{F}))$ must map to this particular $[P_B]$ in $\pi_{\text{path}}(\mathcal{S}_B^{\psi}(\mathcal{F}))$. This holds for all B, so uniqueness is guaranteed.

To settle the issue of existence, we need to check that the path classes determined in the above fashion for the different $\mathcal{S}_B^{\psi}(\mathcal{F})$'s (as B varies) are consistent with one another. That is, if $B' \supset B$, $[P_B]$ should be the image in $\mathcal{S}^{\psi}_{B}(\mathcal{F})$ of the path class $[P_{B'}]$ living in $\mathcal{S}^{\psi}_{B'}(\mathcal{F})$. Diagrammatically:

$$\begin{array}{ccc} [P_B] & \stackrel{?}{\leftarrow} & [P_{B'}] \\ \downarrow & & \downarrow \\ [Q_B] & \leftarrow & [Q_{B'}] \end{array}$$

Let $[\overline{P_B}]$ be the image of $[P_{B'}]$ in $\pi_{\text{path}}(\mathcal{S}^{\psi}_{B}(\mathcal{F}))$. Since the diagram

$$\begin{array}{ccc} \pi_{\text{path}}(\mathcal{S}^{\psi}_{B}(\mathcal{F})) & \leftarrow & \pi_{\text{path}}(\mathcal{S}^{\psi}_{B'}(\mathcal{F})) \\ \downarrow & & \downarrow \\ \pi_{\text{path}}(\mathcal{S}_{B}(\mathcal{G})) & \leftarrow & \pi_{\text{path}}(\mathcal{S}_{B'}(\mathcal{G})) \end{array}$$

commutes, $[\overline{P_B}]$ must be a lift of $[Q_B]$. But $[P_B]$ is also a lift of $[Q_B]$, and it has the same initial point, namely, the image of (x, f) in $\mathcal{S}^{\psi}_{B}(\mathcal{F})$. Hence $[\overline{P_B}] = [P_B]$, as required. $\square$

Proposition 5: *Let f_i $(0 \leq i < d)$ denote the ith pre-image of g under ψ, and let C_i be some projective path class in $\mathcal{S}_B(\mathcal{F})$ with initial point (x, f_0) and terminal point (x, f_i). We insist that C_0 should be the trivial projective path class, but we make no constraints on the other C_i's. Then every element of $\pi_1^{\text{proj}}(\mathcal{S}(\mathcal{G}), (x, g))$, when lifted to a projective path class in $\mathcal{S}(\mathcal{F})$ with initial point (x, f_0), can be written in a unique way as the composition of an element of $\pi_1^{\text{proj}}(\mathcal{S}(\mathcal{F}), (x, f_0))$ and one of the C_i's.*

Proof: We choose the C_i whose terminal point is the same as the terminal point of the lifted projective path class in $\mathcal{S}(\mathcal{F})$. Verification is routine. $\square$

Theorem 4: *Assume that $\mathcal{F}$ and $\mathcal{G}$ are $\mathbf{Z}^2$-shifts, with $\psi : \mathcal{F} \to \mathcal{G}$ a d-to-1 covering map. Assume that $\mathcal{F}$ (and hence $\mathcal{G}$) is projectively connected, with $\mathcal{G}$ compact. Then $\pi_1^{\text{proj}}(\mathcal{F})$ is abstractly an index-d subgroup of $\pi_1^{\text{proj}}(\mathcal{G})$.*

Proof: Take x, g, and f_i's as above. Proposition 5 implies that every element of $\pi_1^{\mathrm{proj}}(\mathcal{S}(\mathcal{G}),(x,g))$ can be written in a unique way as a composition of an element of $\psi_*(\pi_1^{\mathrm{proj}}(\mathcal{S}(\mathcal{F}),(x,f_0)))$ and some $\psi_*\beta_i$. (Here we make use of the injectivity of ψ_*.) Also, C_0 gives us a map from $\pi_1^{\mathrm{proj}}(\mathcal{S}(\mathcal{G}),(x,g))$ to $\pi_1^{\mathrm{proj}}(\mathcal{S}(\mathcal{F}),(x,f_0))$ that respects composition of loop classes. Therefore $\psi_*(\pi_1^{\mathrm{proj}}(\mathcal{S}(\mathcal{F}),(x,f_0)))$ is a subgroup of $\pi_1^{\mathrm{proj}}(\mathcal{S}(\mathcal{G}),(x,g))$. What is more, $\{\psi_*\beta_i\}$ forms a system of coset representatives, so that $\psi_*(\pi_1^{\mathrm{proj}}(\mathcal{S}(\mathcal{F}),(x,f_0)))$ is an index-d subgroup of $\pi_1^{\mathrm{proj}}(\mathcal{S}(\mathcal{G}),(x,g))$. By projective connectedness (and the fact that ψ_* is injective), these two groups are isomorphic to $\pi_1^{\mathrm{proj}}(\mathcal{F})$ and $\pi_1^{\mathrm{proj}}(\mathcal{G})$, respectively. □

Combining Theorems 2 and 4, we obtain:

Corollary: *Every $\mathbf{Z}^2$-shift that admits a d-to-1 covering map from a full shift has a fundamental group of order d.* □

In particular, for $d > 1$, there can be no d-to-1 covering map from a full shift to itself. More generally, no $\mathbf{Z}^2$-shift with finite π_1^{proj} admits a non-trivial covering map onto itself.

Another consequence of Theorems 2 and 4 is that no projectively connected $\mathbf{Z}^2$-shift admits a d-to-1 covering map onto a full shift, with $d > 1$. This is analogous to Kammeyer's result [Kamm] that there are no d-point mixing cocycle extensions of the full-shift ($d > 1$). Kammeyer's result does not immediately imply ours, since not every covering-extension of a $\mathbf{Z}^2$-shift is a cocycle extension. On the other hand, we do not know if our result implies Kammeyer's, because we do not know whether every cocyle-extension of a $\mathbf{Z}^2$-shift is automatically covering.

Theorem 4 applies directly to the examples described at the beginning of this section. It does not, however, apply to the ice-system $\mathcal{F}$ and lifted ice-system $\widehat{\mathcal{F}}$, since the map between them is countable-to-one. Fortunately, the problem is remediable. Everything in Propositions 1, 3, 4, and 5 that concerns d-to-1 maps applies equally well to countable-to-one maps, except for the construction of the set B^*. We used the finiteness of d to guarantee the existence of a bounded set B^* that contains all the $B_{i,j}$'s associated with the local inverses $\psi_{i,j}$. In the case of ice, this is easy, since the local inverses are 1-block maps. Hence this section corroborates the previous section by showing that $\pi_1^{\mathrm{proj}}(\widehat{\mathcal{F}})$ is a subgroup of $\pi_1^{\mathrm{proj}}(\mathcal{F})$ of infinite index. Also, note that the set of proper three-colorings of the square grid, being a three-point extension of square ice, has a projective fundamental group that is naturally identified with the index-3 subgroup of $\pi_1^{\mathrm{proj}}(\mathcal{F})$.

More generally, to handle the case in which $\mathcal{G}$ is not compact, let us say that a covering map $\psi : \mathcal{F} \to \mathcal{G}$ is "uniformly covering" if there exists a bounded set B^* such that the B^*-cylinder sets are evenly covered under ψ and such that the local inverse maps are block maps relative to B^*. Then the proof of Theorem 4 goes through much as before.

8 Further Examples

Until now we have mostly used the framework of Markov shifts in studying two-dimensional symbolic dynamics, but as our study of the square ice model might lead one to guess, it is in some ways more convenient to use an enriched Wang tile framework (see [GrSh]). A Wang tile is normally defined as a 1-by-1 square tile with each of its four edges marked with some color drawn from a finite set of colors. Two such tiles may be placed next to one another (that is, fully sharing an edge) if they assign the same color to the shared edge. We alter this definition slightly by allowing the vertices of the tile to be assigned colors as well (distinct from the edge colors); tiles may share a vertex if they assign the same color to it.

A tile set is a set of Wang tiles, which may be translated (but not rotated) at will in the plane. The "decision problem" asks whether translates of a given set of Wang tiles can be used to tile the entire plane. This is known to be an undecidable problem [Berg]. Nonetheless, the set of such tilings makes a perfectly nice dynamical system under the action of $\mathbf{Z}^2$ (though it may be empty!).

Every two-dimensional Markov shift can easily be encoded in a Wang tile system (of our kind); one simply creates a set of vertex-marked Wang tiles, one for each allowed m-by-m block of symbols, with m suitably large. In the other direction, every Wang tile system can be encoded by a matrix subshift. For, we may without loss of generality assume that our Wang tile system involves no vertex-marking (since we can enrich the edge-markings so that they tell us the vertex-markings as well). Now we may introduce a symbol for each tile, and the consistency conditions for adjacent tiles will translate straightforwardly into adjacency conditions for the symbols. Thus, every two-dimensional Markov shift is topologically conjugate to a Wang tile system, and vice versa.

The Wang tiles that will be most interesting to us are those in which the vertex-colors are elements of some set S, and the edge-colors are elements of some group G acting transitively on S. More specifically, we will require for all $s, t \in S$ that if two adjacent vertices of a tile are marked s and t (with s either below or to the left of t) then the edge between them must be marked with some $g \in G$ that sends s to t. We further require that if the right, bottom, top, and left edges of a tile are marked g_1, g_2, g_3, and g_4, then $g_1 g_2 = g_3 g_4$. We let the tile set consist of all such tiles (unless we specify otherwise).

If we take $G = S = \mathbf{Z}$ acting on itself by addition, but allow only those tiles in which edges are marked with 1's and -1's, we get the dual ice model, with fundamental group $\mathbf{Z}$.

If G acts freely on S (i.e., no element of G besides the identity element has any fixed points), then the vertex-markings determine the edge-markings, so that we may dispense with the latter without affecting the system. In particular, if G is finite and $S = G$ with G acting on itself by multiplication, then we get the full shift on the alphabet G (the "full G-shift"). This system has trivial fundamental group, as was shown in Section 5.

More generally, the set of tilings of the plane by Wang tiles in the $G : S \to S$ situation, viewed as a dynamical system, is a $|G|/|S|$-to-1 factor of the full G-shift, assuming G acts transitively on S. In fact, the projective fundamental group in this case is precisely the stabilizer group Stab s (for $s \in S$). Thus, every finite group is π_1^{proj} of some factor of a full shift. An example is gotten by letting G be the 2-element group, acting freely on a two-element set S. The resulting 2-to-1 quotient of the full 2-shift is the set of all ways of choosing a subgraph of the infinite square grid so that around each square face, an even number of grid-edges are chosen (simply choose the edges that are marked "1"). Equivalently (via duality), one can look at this as the set of all ways of choosing a subgraph of the infinite square grid so that each vertex has even degree. This is the transformed version of the 2-dimensional Ising model discussed in Section 7.

If we are willing to work with non-compact systems, we can get π_1^{proj} to be any group G whatsoever (just let S be the one-point G-set), though if we want our system to be locally compact, as in the case of square ice, then there are undoubtedly constraints on what π_1^{proj} can be.

We can use Wang tiles to give examples of systems that have the same entropy and the same behavior on periodic points (combinatorially speaking) but different fundamental groups, so that π_1^{proj} serves, in at least some instances, as the invariant of choice for distinguishing between two systems. Let G_1 and G_2 be non-isomorphic groups of order d, with each group acting on a 1-point set S. Each resulting tiling system is a d-to-1 factor of the full d-shift, and thus has entropy $\log d$. What is more, for any sub-lattice Λ of $\mathbf{Z}^2$, it can be shown in both cases that the number of Λ-periodic points in the subshift is equal to $d^{[\mathbf{Z}^2:\Lambda]+1}$. This implies that there is a bijection between the periodic points of the first system and the periodic points of the second that commutes with the dynamics. Nevertheless, the two systems have fundamental groups G_1 and G_2, respectively, and so must be non-conjugate.

Three variants of our construction deserve mention. First, suppose $\mathcal{F}$ is the set of tilings of the plane by bounded regions of various sorts. Given $x \in \mathbf{R}^2$ and $f, f' \in \mathcal{F}$, we may define $\sim$ so that $(x, f) \sim (x, f')$ if there exists a region $T \subset \mathbf{R}^2$ that contains x and occurs as a tile in both f and f'. Then the fundamental group of the quotient $(\mathbf{R}^2 \times \mathcal{F})/\!\sim$ is exactly the tile homotopy group in the sense of Thurston [Thur] in the case where all tilings in $\mathcal{F}$ live on a grid. However, $(\mathbf{R}^2 \times \mathcal{F})/\!\sim$ is perfectly well-defined in the absence of a grid; indeed, it would be interesting to compute the fundamental group of $(\mathbf{R}^2 \times \mathcal{F})/\!\sim$ in the case where $\mathcal{F}$ is the set of Penrose tilings of the plane. (Actually, the quotient space is disconnected, but it seems likely that each connected piece should have the same fundamental group as every other.)

Second, suppose that $\mathcal{F}$ is the space of C^∞ functions on $\mathbf{R}^2$, topologized so that $f_n \to f$ in $\mathcal{F}$ if and only f_n agrees with f on B for all sufficiently large bounded sets $B \subset \mathbf{R}^2$. Let B_r be the disk of radius r. Then we get scene-spaces $\mathcal{S}_r = \mathcal{S}_{B_r}(\mathcal{F})$ for all $r > 0$. If we take the *direct* limit as $r \to 0$, we get the space of germs of C^∞ functions (denote it by $\mathcal{S}_0$). $\mathcal{S}_0$ seems to be simply connected, though we have not found a proof. In any case, $\mathcal{S}_0$ is a covering space of the space of germs of C^∞ conservative vector fields on $\mathbf{R}^2$ (where a vector field is conservative if it is the gradient of some scalar field) with monodromy group $\mathbf{R}$, so if all of the above is correct, the latter space of germs is a connected space having $\mathbf{R}$ as its fundamental group.

Third, it is possible to apply the notion of scene-space to $\mathbf{Z}^d$ actions for general values of d, and this yields not only a broader definition of π_1^{proj} but also allows one to define higher projective homotopy groups. For example, with $d = 3$ it appears to be the case that the dimer model on an infinite cubical lattice has π_1^{proj} trivial and π_2^{proj} isomorphic to $\mathbf{Z}$. In the other direction, with $d = 1$ it seems that the projective fundamental group of every non-trivial subshift of finite type is infinitely generated. This is related to the fact that one-dimensional subshifts of finite type do not exhibit the "cocyle rigidity" manifested by many higher-dimensional subshifts (see [Sch2]); that is to say, one-dimensional dynamical systems typically admit an abundance of constant-to-one extensions.

To conclude, we will deliver on a promise made at the beginning of our article. Note that for all $x, x' \in \mathbf{R}^2$, (x, f) is connected to (x', f) by a projective path class. This implies that if (x_1, f_1) and (x_2, f_2) are connected by a projective path class, then so are (x'_1, f_1) and (x'_2, f_2) for all x'_1, x'_2 in $\mathbf{R}^2$. Hence, the decomposition of the product space $\mathcal{S} = \mathbf{R}^2 \times \mathcal{F}$ into projective path-class components arises from a decomposition of $\mathcal{F}$ itself. Going further in this direction, let us call a projective path-class trivial if it "lives on a single sheet" – that is, if for every B the class has a representative in the fundamental groupoid of $\mathcal{S}_B$ that lives on a single sheet. If we mod out the fundamental groupoid by the trivial projective path-classes, we get a quotient-groupoid whose elements have "endpoints" not in $\mathbf{R}^2 \times \mathcal{F}$ but in $\mathcal{F}$ itself. This is what we meant when, in Section 1, we spoke of the possibility of connecting two elements of $\mathcal{F}$ by something like a path.

The authors thank Mike Boyle, David Feldman, Bruce Kitchens, Michael Larsen, Andy Latto, Andrew Millis, Shahar Mozes, Klaus Schmidt, Bill Thurston, and Jack Wagoner for their useful comments during the preparation of this article.

References

[1] Baxter, R. J.: Exactly solved models in statistical mechanics. New York London: Academic Press 1982

[2] Berger, R.: The undecidability of the domino problem. (Memoirs Amer. Math. Soc. vol. 66) American Mathematical Society 1966.

[3] Conway, J. H. and Lagarias, J. C.: Tilings with polyominoes and combinatorial group theory. J. Combin. Theory A **53** 183-208 (1990)

[4] Grünbaum B. and Shephard G. C.: Tilings and Patterns. New York: Freeman 1987. W. H. Freeman and Company, New York, 1987

[5] Kammeyer, J.: A complete classification of two-point extensions of a multidimensional Bernoulli shift. Doctoral Dissertation, University of Maryland, 1988.

[6] Kasteleyn, P. W.: The statistics of dimers on a lattice, I. Physica **27**, 1209-1225 (1961)

[7] Lieb, E. H.: Residual entropy of square ice. Phys. Rev. **162**, 162-172 (1967)

[8] Markley, N. G. and Paul, M. E.:, Matrix subshifts for $\mathbf{Z}^\nu$ symbolic dynamics. Proc. London Math. Soc. **43**, 251-272 (1981)

[9] Nasu, M.: Constant-to-one and onto global maps of homomorphisms between strongly connected graphs. Ergod. Thy. & Dynam. Sys. **3**, 387-413 (1983)

[10] Schmidt, K.: Algebraic Ideas in Ergodic Theory. (Conference Board of Mathematical Sciences, vol. 76) American Mathematical Society 1990.

[11] Schmidt, K.: The cohomology of higher-dimensional shifts of finite type. Preprint 1992.

[12] Thurston, W.: Conway's tiling groups. American Mathematical Monthly **97**, 757-773 (1990).