Journal of Statistical Physics, Vol. 74, Nos. 3/4, 1994

Trace Maps as 3D Reversible Dynamical Systems with an Invariant

John A. G. Roberts[1,2] **and Michael Baake**[1,3]

Received February 24, 1993; final September 29, 1993

One link between the theory of quasicrystals and the theory of nonlinear dynamics is provided by the study of so-called trace maps. A subclass of them are mappings on a one-parameter family of 2D surfaces that foliate $\mathbb{R}^3$ (and also $\mathbb{C}^3$). They are derived from transfer matrix approaches to properties of 1D quasicrystals. In this article, we consider various dynamical properties of trace maps. We first discuss the Fibonacci trace map and give new results concerning boundedness of orbits on certain subfamilies of its invariant 2D surfaces. We highlight a particular surface where the motion is integrable and semiconjugate to an Anosov system (i.e., the mapping acts as a pseudo-Anosov map). We identify properties of symmetry and reversibility (time-reversal symmetry) in the Fibonacci trace map dynamics and discuss the consequences for the structure of periodic orbits. We show that a conservative period-doubling sequence can be identified when moving through the one-parameter family of 2D surfaces. By using generator trace maps, in terms of which all trace maps obtained from invertible two-letter substitution rules can be expressed, we show that many features of the Fibonacci trace map hold in general. The role of the Fricke character $\hat{I}(x, y, z) = x^2 + y^2 + z^2 - 2xyz - 1$, its symmetry group, and reversibility for the Nielsen trace maps are described algebraically. Finally, we outline possible higher-dimensional generalizations.

KEY WORDS: Dynamical systems; substitution rules; trace maps; quasicrystals; pseudo-Anosov systems; reversible systems; period doubling; orbit analysis.

[1] Department of Mathematics, University of Melbourne, Parkville, Victoria 3052, Australia.
[2] Institute for Theoretical Physics, University of Amsterdam, 1018XE Amsterdam, The Netherlands.
[3] Institut für Theoretische Physik, Universität Tübingen, 72076 Tübingen, Germany. e-mail: ptiba01@mailserv.zdv.uni-tuebingen.de.

1. INTRODUCTION

There has been a growing interest in quasiperiodic structures and phenomena.[1] Likewise, dynamical systems are an area of recent rapid development. These two topics have been linked via a transfer matrix approach to the study of quasiperiodic systems, which leads to the derivation of three-dimensional (3D) mappings called *trace maps*.[2,3] A large class of these trace maps, to be considered here, possesses one integral of motion and consequently these trace maps induce motion on the foliation of $\mathbb{R}^3$ by the family of 2D level sets of the integral. In this article, we study the dynamics of such trace maps, particularly with regard to their periodic orbits, and identify various features and symmetries common to a large subset of such mappings.

The results of refs. 4–6 place the construction of trace maps on a firm mathematical footing. In the physics literature, however, trace maps first occurred in studies of systems with spatial, and later temporal, quasiperiodic structure described by the Fibonacci sequence.[2,3,7]

In the category of spatial structure lies the study of 1D Schrödinger operators (or tight-binding approximations thereof) on Fibonacci or similar chains. Here, the arrangements of potentials follow the chain. The interest is then to determine spectrum and wave functions of such operators. Also, classical as well as quantum spin systems like the Ising model are interesting systems on such aperiodic structures. The category of temporal structure includes the example of a single spin in a time-dependent magnetic field $B(t)$, or, more generally, any (nonperiodically) kicked two-level system (cf. also ref. 8 for a survey).

The above models can be investigated with a transfer matrix approach which introduces dynamical systems theory into the problem. The cell structure allows the use of recursively defined transfer matrices $\{M_n\}$ which act, say, over the f_n (nth Fibonacci number) sites of the nth Fibonacci approximant according to

$$M_{n+1} = M_n M_{n-1} \tag{1}$$

The matrix recurrence (1) is a renormalization scheme, relating the transfer matrices over successively longer approximants to the infinite chain (recall that the Fibonacci numbers f_n diverge like $f_n \sim \tau^n$, where τ is the golden mean). If one is interested in spectral properties only, much physical information can be obtained from the implication of (1) on the *traces* of the transfer matrices. They obey a decoupled recurrence relation themselves—which is the origin of trace maps. In the Fibonacci case, this is a third-order difference equation which can equivalently be regarded as a mapping of 3D space. Various dynamical features of this mapping have now been studied and related back to physical implications for the quasiperiodic phenomena. More recently, much work has been devoted to trace maps obtained from generalizations of the transfer matrix recurrence (1). Many similarities in the results have emerged, simultaneously being linked to the underlying algebraic structure and to certain properties of the corresponding dynamical system.

From a dynamical point of view, trace maps are interesting for various reasons (compare also ref. 9). First, they can be derived as discrete dynamical systems from a continuous one without any approximation—they are not idealized or approximate discretizations of a continuous dynamical system.[10] Second, the 3D trace maps considered here possess an invariant quantity, which means that their motion is confined to the 2D level sets of this invariant quantity.[5,11] On one of these (compact) level sets, the action of the mapping is related to a hyperbolic toral automorphism, and thus to an Anosov system. Motion on nearby level sets can be studied to see the approach to the chaotic dynamics on this particular surface. Third, as will be shown here, many trace maps provide nice models of *reversible* dynamical systems on 2D manifolds that are firmly rooted in a physical problem (reversible dynamical systems are those with a generalized time-reversal symmetry[12]). Note that reversibility is a much stronger property than just (time) invertibility of a dynamical system.

On the other hand, common features are often the result of a systematic algebraic structure, and trace maps are no exception. In fact, their reformulation by means of two-letter substitution rules provides the right basis to exploit this. It will turn out that Nielsen's work on free groups and their automorphisms is central here,[13] while Nielsen's work on manifolds[14] shows up in the dynamical aspects—all together a somewhat surprising connection.

The structure of this paper is as follows. In Section 2, we set up notation and some preliminaries on substitution rules, in particular *invertible* ones, and their trace maps. In Section 3, we present the well-known Fibonacci trace map as our first example and discuss its dynamical structure in some detail. In Section 4, we briefly describe the most important properties of reversibility in our context and discuss a period-doubling bifurcation cascade found for the Fibonacci trace map.

Section 5 deals with a family of generalizations of the Fibonacci trace map that show structurally similar behavior, including the existence of an invariant and reversibility. In Section 6 we present a more unified picture to explain the prevalence of these properties using a generator approach to volume-preserving trace maps. In particular, the role of the invariant becomes obvious and reversibility also follows for a significant subclass of mappings. Our main results are summarized in Propositions 17–23 of this Section.

Finally, in Section 7, we turn to a particular class of generalizations in higher dimensions, i.e., to n-parameter families of diffeomorphisms of $\mathbb{R}^n$ or $\mathbb{C}^n$ that are reversible and possess an invariant (though they are, in general, not trace maps). We briefly describe this class of mappings, which is followed by some concluding remarks in Section 8, while the Appendix recalls some properties of dynamical systems on 2D manifolds reformulated for our present needs.

2. PRELIMINARIES: SUBSTITUTION RULES, SUBSTITUTION MATRICES, TRACE MAPS

This section provides a brief summary, for the unfamiliar reader, of the concepts of substitution rules and substitution matrices, and discusses how to derive the associated trace maps.

Part of the study of quasiperiodic structures has been devoted to 1D nonperiodic tilings that possess a deflation/inflation symmetry (see refs. 1 and 15 and references therein). Those with two tiles, a and b, can be built from a two-letter replacement or *substitution rule* of the form

$$\rho\colon \quad \begin{array}{l} a \to \rho(a) = w_a(a, b) \\ b \to \rho(b) = w_b(a, b) \end{array} \tag{2}$$

In (2), $w_a(a, b)$ and $w_b(a, b)$ are *words* or strings built from the two-letter alphabet $\{a, b\}$, where a, b then stand for the two different tiles. However, for many purposes, it is advantageous to include the inverses $\{a^{-1}, b^{-1}\}$ formally defined by $a^{-1}a = aa^{-1} = e$, where e is the empty word and multiplication of words is defined by concatenation.[16] That is, $w_a(a, b)$ and $w_b(a, b)$ take the form $x_1 x_2 \cdots x_r$, where x_i is a, b, a^{-1}, or b^{-1}. The set of all finite words then constitutes the free group $\mathcal{F}_2$ generated by the two-letter alphabet $\{a, b\}$—more will be said about this in Section 6 below. In order to be able to calculate the image $\rho(w)$ of a given word w under the rule ρ, one should only consider ρ's that are homomorphisms on the group of all words, i.e.,

$$\rho(w_1 w_2) = \rho(w_1)\, \rho(w_2) \tag{3}$$

for any two words w_1 and w_2, where again the 'products' on the right-hand side correspond to word juxtaposition. Because of the property (3), $\rho(w)$ is completely described by the images $\rho(a)$ and $\rho(b)$, and therefore (2) is sufficient to specify ρ.

If ρ and σ are two substitution rules, then their product $\sigma\rho$ is defined, following ref. 17, to be the substitution rule obtained by applying σ followed by ρ,

$$\sigma\rho := \rho \circ \sigma \tag{4}$$

More explicitly, if ρ is defined by (2) and σ by $\sigma: a \to \sigma(a) = W_a(a, b)$, $b \to \sigma(b) = W_b(a, b)$, then $\sigma\rho$ is given by the rule $a \to \rho \circ \sigma(a) = W_a(w_a(a, b), w_b(a, b))$, $b \to \rho \circ \sigma(b) = W_b(w_a(a, b), w_b(a, b))$. A substitution rule ρ is called *invertible* if there exists another substitution rule, denoted ρ^{-1}, such that

$$\rho\rho^{-1} = \rho^{-1}\rho = id \tag{5}$$

where id: $a \to a$, $b \to b$. The invertible substitution rules form a group, Φ_2, known as the automorphism group of the free group $\mathscr{F}_2$.[13]

Using composition, we can repeatedly apply the same rule ρ to a given word w_0. This leads to a sequence of new words $w_1 = \rho(w_0)$, $w_2 = \rho^2(w_0)$, $w_3 = \rho^3(w_0)$, etc. For example, applying the rule twice to $w_0 = a$ gives $a \to \rho(a) = w_a(a, b) \to \rho^2(a) = w_a(w_a(a, b), w_b(a, b))$.

One (abbreviated) way to encode the statistical properties of a substitution rule ρ is via the *substitution matrix* R_ρ defined by

$$R_\rho = \begin{pmatrix} \#_a(w_a) & \#_b(w_a) \\ \#_a(w_b) & \#_b(w_b) \end{pmatrix} \tag{6}$$

In (6), $\#_a(w_a)$ is the number of a's in $w_a(a, b)$, (adding $+1$ for every occurrence of a, and -1 for every occurrence of a^{-1}) and $\#_b(w_a)$ is the number of b's etc. See refs. 13 and 10 for details of the underlying Abelianization process. Evidently R_ρ and its powers are integer matrices. If the integer row vector $(\#_a(w_0), \#_b(w_0))$ gives the number of a's and b's in a word w_0, then multiplication of this vector from the right byR_ρ^n gives a new row vector $(\#_a(w_n), \#_b(w_n))$ whose components are the numbers of a's and b's in $w_n = \rho^n(w_0)$. If the integer matrix R_ρ is hyperbolic with real eigenvalues λ and μ satisfying $|\lambda| > 1 > |\mu|$, then the relative frequencies of the two letters as $n \to \infty$ are given by the entries in the left eigenvector of R_ρ corresponding to λ. Hence the ratio of b's to a's in the infinite word is equal to

$$(\lambda - \#_a(w_a))/\#_a(w_b) = \#_b(w_a)/(\lambda - \#_b(w_b))$$

This is of course a special case of the Perron–Frobenius eigenvalue and eigenvector of R_ρ if the entries are nonnegative only.

One canonical example of a 1D quasiperiodic structure that is obtained from a two-letter substitution rule is the famous 1D Fibonacci chain. The rule in this case is

$$\rho_1 : \begin{array}{l} a \to b \\ b \to ba \end{array} \tag{7}$$

The first few applications of ρ_1 with the initial condition $w_0 = a$ produce the word sequence $\rho_1(a) = b$, $\rho_1^2(a) = ba$, $\rho_1^3(a) = bab$, $\rho_1^4(a) = babba$. Calling $S_n = \rho_1^n(a)$ the nth Fibonacci approximant, the Fibonacci chain is defined as the infinite limit of the string sequence S_n, where we disregard subtleties like infinite versus half-infinite words or isomorphism concepts for now. By construction, S_n contains f_{n-1} $\{b\}$'s, f_{n-2} $\{a\}$'s, and hence f_n total symbols, where f_n is the nth Fibonacci number (defined by $f_0 = f_1 = 1$ and $f_{n+2} = f_{n+1} + f_n$). The ratio of the number of $\{b\}$'s to the number of $\{a\}$'s in the Fibonacci chain is consequently equal to $\tau = \lim_{n \to \infty}(f_{n-1}/f_{n-2}) = (1 + \sqrt{5})/2$, the famous golden mean. The fact that this number is *irrational* indicates that the Fibonacci chain is necessarily aperiodic, since a periodic 2-symbol chain is built up by repeating a finite-length unit cell and so has a rational proportion of constituent elements. Another way to see the nonperiodicity of the infinite Fibonacci chain is via its substitution matrix

$$R_1 = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \tag{8}$$

The matrix R_1 has eigenvalues $\lambda = \tau$ and $\mu = -1/\tau$, so that from above the ratio of b's to a's is again $\tau = (\lambda - \#_a(w_a))/\#_a(w_b)$. Despite the fact that this ratio is irrational, the way that the Fibonacci chain is constructed is far from random, and it can be considered to be successively better approximated by a sequence of periodic chains with unit cell given by the Fibonacci approximants S_n.

Note that the substitution matrix does not give a complete description of a substitution rule, because different rules can have the same substitution matrix, so that some information is lost in this description, e.g., the rule $a \to b$, $a \to ab$ leads to the same matrix R as (8). In this simple example, we would obtain an equivalent chain (which is in the same local isomorphism class[18]), but that is not generally true for other substitution matrices.[19,8]

To derive a *trace map* from a two-letter substitution rule ρ, we identify the letters $\{a, b\}$ with 2×2 matrices $\{A, B\}$ in $Sl(2, \mathbb{C})$. The substitution rule (2) then induces a matrix substitution rule where $w_A(A, B)$ and $w_B(A, B)$ now represent products of strings of the two-component matrices. It is well known[5,6,17] that for *any* such matrix substitution rule, there always exists a unique polynomial mapping $F_\rho: \mathbb{C}^3 \mapsto \mathbb{C}^3$ with integer coefficients that expresses the coordinate triple $(\mathrm{tr}(\rho(A)), \mathrm{tr}(\rho(B)), \mathrm{tr}(\rho(AB)))$ in terms of $(\mathrm{tr}(A), \mathrm{tr}(B), \mathrm{tr}(AB))$. Repeated composition of a trace map allows, for instance, the trace of the matrix word $\mathrm{tr}(\rho^n(A))$ to be expressed in terms of $\mathrm{tr}(A)$, $\mathrm{tr}(B)$, and $\mathrm{tr}(AB)$.

It can be shown[17] that the trace map and substitution matrix of a product (4) of two substitution rules are given by

$$F_{\sigma\rho} = F_\sigma \circ F_\rho, \qquad R_{\sigma\rho} = R_\sigma \cdot R_\rho \tag{9}$$

where $\circ$ is functional composition and $\cdot$ is matrix multiplication. This shows that the mappings from the substitution rules to the substitution matrices, and from the substitution rules to the trace maps, are homomorphisms, a point we will return to in Section 6. In this article, we will focus our attention on *invertible* substitution rules and their trace maps.

3. THE FIBONACCI TRACE MAP: INTEGRAL OF MOTION AND DYNAMICS ON SOME LEVEL SETS

In this Section, we begin a discussion of the dynamics of the well-known Fibonacci trace map and its associated, equally well-known, integral of motion.[5,11,2,3] We identify some of the level sets of the integral where the motion is solvable or essentially describable.

As discussed in the Introduction, we are interested in an interpretation of (7) as a renormalization approach to certain physical systems that live on such a chain—either in space or in time, and this is usually described by some sort of transfer, formulated in terms of matrices. Therefore, we rewrite ρ_1 of (7) as a matrix recursion after making the identifications with unimodular 2×2 matrices, i.e., $\rho_1^n(a) \to A_n$, $\rho_1^n(b) \to B_n$, with $A_n \in Sl(2, \mathbb{C})$, etc. We obtain

$$\begin{aligned} A_{n+1} &= B_n \\ B_{n+1} &= B_n A_n \\ C_{n+1} &= A_{n+1} B_{n+1} = B_n^2 A_n \end{aligned} \tag{10}$$

where we have introduced $C_n = A_n B_n$ for systematic reasons (cf. the general result of ref. 6 referred to above), although it is not necessary for this example.

For a 2×2 matrix T with $\det(T) = 1$, we have $T^2 = \mathrm{tr}(T)\cdot T - \mathbb{1}$ from the Cayley–Hamilton theorem. As a consequence, the traces

$$x_n = \tfrac{1}{2}\mathrm{tr}(A_n), \qquad y_n = \tfrac{1}{2}\mathrm{tr}(B_n), \qquad z_n = \tfrac{1}{2}\mathrm{tr}(C_n) \tag{11}$$

decouple from the matrix iteration (10), i.e., we can write x_{n+1}, y_{n+1}, and z_{n+1} as functions of x_n, y_n, and z_n only. This decoupling gives rise to a 3D dynamical system

$$F_1: \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} y \\ z \\ 2yz - x \end{pmatrix} \tag{12}$$

The mapping F_1 is the Fibonacci trace map.[2,3] It is a diffeomorphism of $\mathbb{C}^3$ and possesses an integral of motion

$$\hat{I}(x, y, z) = x^2 + y^2 + z^2 - 2xyz - 1 \tag{13}$$

in the sense that $\hat{I}(F_1 \mathbf{x}) = \hat{I}(\mathbf{x})$ for all $\mathbf{x} = (x, y, z) \in \mathbb{C}^3$. This can easily be checked by direct substitution. The (unnecessary) constant term in (13) is kept for historical reasons as well as compatibility with other publications. $\hat{I}$ in (13) is a simple example of a so-called "Fricke character."[4,20,21,11] We will frequently use this term or the term "Fricke invariant" (although it may in fact be a misnomer, as pointed out in ref. 11, because of prior work by Vogt[22]). Motivated by the various physical applications mentioned in the Introduction, we will now investigate the mapping F_1 mainly as a diffeomorphism of $\mathbb{R}^3$, although it certainly has interesting properties as a complex mapping.

Possession of an integral or invariant by a 3D mapping has important dynamical implications because it confines the motion to its family of 2D level sets. Thus one obtains an induced one-parameter family of mappings over the 2D surfaces, the parameter being the value of the invariant. Let us briefly describe the invariant level sets of F_1 given by

$$\mathcal{M}_\mu := \{(x, y, z) \in \mathbb{R}^3 \mid \hat{I}(x, y, z) = \mu\} = \mathcal{M}_\mu^c \cup \mathcal{M}_\mu^{nc} \tag{14}$$

The last part of Eq. (14) indicates the division of $\mathcal{M}_\mu$ into its compact (c) and noncompact (nc) parts, a distinction that is useful for $-1 \leqslant \mu \leqslant 0$. The 2D surfaces $\mathcal{M}_\mu$ have tetrahedral symmetry, i.e., they are invariant under any permutation of the three coordinates and under any pairwise sign change like $(x, y, z) \mapsto (-x, -y, z)$. For $\mu < 0$, $\mathcal{M}_\mu^{nc}$ consists of four disconnected noncompact cones going to infinity. They are confined to four of the eight octants of $\mathbb{R}^3$, and are external to the cube (or box) $\mathcal{B}$ bounded by the planes $\{x = \pm 1\}$, $\{y = \pm 1\}$, and $\{z = \pm 1\}$, i.e.,

$$\mathcal{B} := \{(x, y, z) \mid |x| \leqslant 1, |y| \leqslant 1, |z| \leqslant 1\} \tag{15}$$

That is, for $\mu<0$, we can introduce a notation for the cones and write

$$\begin{aligned}\mathscr{M}_\mu^{nc} &:= \bigcup \mathscr{C}_\mu^{(\varepsilon_x,\varepsilon_y,\varepsilon_z)}\\ \mathscr{C}_\mu^{(\varepsilon_x,\varepsilon_y,\varepsilon_z)} &:= \mathscr{M}_\mu \cap \{xyz \geqslant 1 \mid \operatorname{sgn}(x)=\varepsilon_x,\ \operatorname{sgn}(y)=\varepsilon_y,\ \operatorname{sgn}(z)=\varepsilon_z\}\end{aligned} \tag{16}$$

Here ε_x, ε_y, $\varepsilon_z \in \{-1, 1\}$ with $\varepsilon_x\varepsilon_y\varepsilon_z=1$. We will use the abbreviations $+(-)$ for $+1\ (-1)$ in what follows. Only the cone with positive coordinates, $\mathscr{C}_\mu^{(+,+,+)}$, is an invariant set of F_1 itself, while the other three cones are mapped cyclically between themselves. For $\mu<-1$, $\mathscr{M}_\mu^c$ is empty and the four cones of $\mathscr{M}_\mu^{nc}$ constitute $\mathscr{M}_\mu$; see Fig. 1a. However, for $-1<\mu<0$, we additionally have a nonempty $\mathscr{M}_\mu^c$ which is equal to a compact and invariant set around the origin diffeomorphic to a sphere. It is wholly contained within the cube $\mathscr{B}$; see Fig. 1b. Since we will need this compact object several times, we define, for $\mu \in [-1, 0]$,

$$\mathscr{M}_\mu^c := \mathscr{M}_\mu \cap \mathscr{B} = \{(x, y, z) \in \mathscr{B} \mid \hat{I}(x, y, z) = \mu\} \tag{17}$$

$\mathscr{M}_\mu^c$ degenerates to the origin at $\mu=-1$. At $\mu=0$, it reaches the four cones of $\mathscr{M}_0^{nc}$, for which we also use the notation of Eq. (16), in one point each, namely in four of the vertices of cube $\mathscr{B}$, which we will refer to as the set $\mathscr{P}$ of 'pinches':

$$\begin{aligned}\mathscr{P} &:= \mathscr{M}_0^c \cap \mathscr{M}_0^{nc} = \{(1, 1, 1)\} \cup \hat{\mathscr{P}},\\ \hat{\mathscr{P}} &:= \{(1, -1, -1), (-1, 1, -1), (-1, -1, 1)\}\end{aligned} \tag{18}$$

See Fig. 1c. In the case $\mu=0$, the compact set is only homeomorphic to a sphere. Finally, for $\mu>0$, $\mathscr{M}_\mu^c$ is again empty and $\mathscr{M}_\mu = \mathscr{M}_\mu^{nc}$ is always a connected, noncompact C^∞-manifold; see Fig. 1d for a typical picture.[4]

Let us remark that we have described and depicted the invariant sets $\mathscr{M}_\mu$ for all ranges of μ, in keeping with our interest in this paper of investigating the dynamics on them for all ranges of μ. Different physical applications of trace maps lead one to consider different regimes in μ. For example, for the Fibonacci sequence, the tight-binding model mentioned in the Introduction "chooses" the regime $\mu>0$, whereas the kicked two-level system "chooses" the regime $\mu \leqslant 0$; for details see 8 and references therein.

We will be particularly interested in studying the periodic orbits of period n, or *n-cycles* of F_1 (and later its generalizations). In this respect we note some consequences for the n-cycles of a 3D mapping with an arbitrary invariant $I(x, y, z)$ (we use the term *eigenvalue spectrum* of an n-cycle for the set of eigenvâlues of the linearisation dF^n evaluated at a point of the cycle, and henceforth 'OR' and 'OP' stand for, respectively, orientation-reversing and orientation-preserving):

[4] We are grateful to D. Joseph for providing these pictures.

Proposition 1. Let $F: \mathbb{R}^3 \to \mathbb{R}^3$ be a 3D diffeomorphism with a continuously differentiable invariant $I = I(x, y, z)$ such that $I(F\mathbf{x}) = I(\mathbf{x})$, $\mathbf{x} = (x, y, z)$. Let $\mathbf{p}_0 = (x_0, y_0, z_0)$ be a point of an n-cycle of F with $I(\mathbf{p}_0) = \mu_0$ and $\nabla I(\mathbf{p}_0) \neq \mathbf{0}$. Then:

(i) $\{+1\}$ is an eigenvalue of $dF^n(\mathbf{p}_0)$. If F is also volume-preserving, the eigenvalue spectrum is $\{1, \lambda, \lambda^{-1}\}$ if F is OR with n even or OP, while it is $\{1, \lambda, -\lambda^{-1}\}$ if F is OR with n odd.

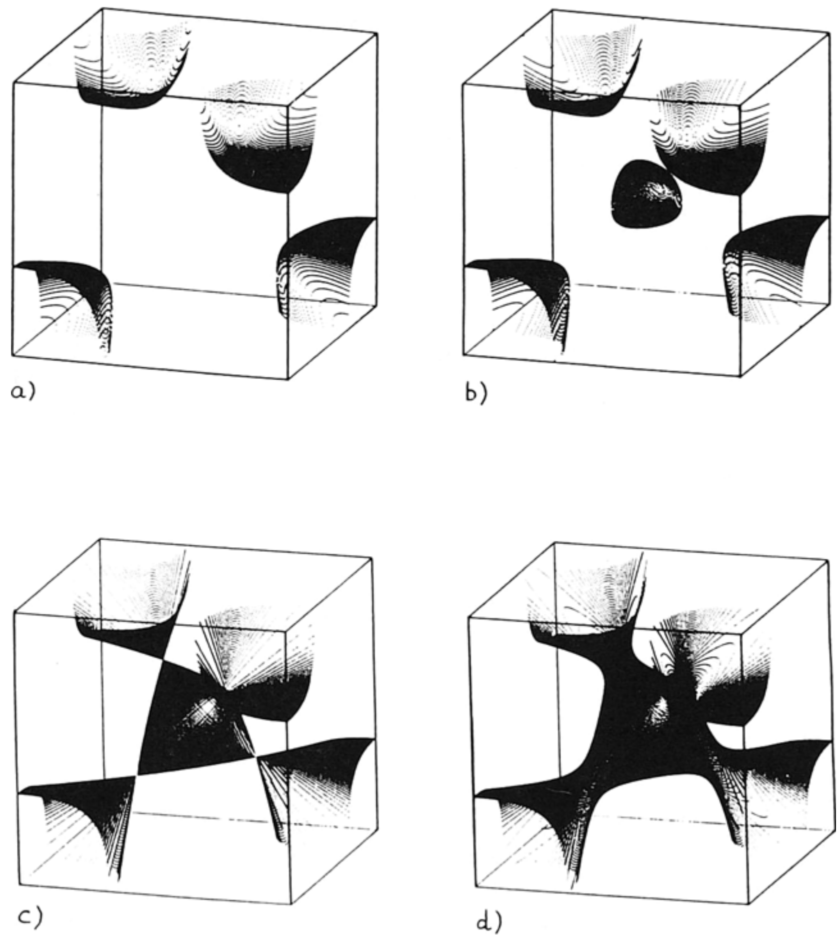

Fig. 1. The structure of the level sets $\mathcal{M}_\mu$ for (a) $\mu = -1.01 < -1$, (b) $-1 < \mu = -0.5 < 0$, (c) $\mu = 0$, and (d) $\mu = 0.3 > 0$.

(ii) If $\{+1\}$ is an eigenvalue of multiplicity 1, the point $\mathbf{p}_0$ is embedded locally into a unique curve $\mathbf{p}(\mu) = (x(\mu), y(\mu), z(\mu))$ that crosses the level set $I(\mathbf{x}) = \mu_0$, and whose points are points of n-cycles of F on adjacent level sets.

(iii) The curve $\mathbf{p}(\mu)$ can intersect a curve of points of a $(j \cdot n)$-cycle at $\mathbf{p}_0$ only if $dF^n(\mathbf{p}_0)$ has an eigenvalue equal to a jth root of unity.

The proof of this proposition follows from the local equivalence between a 3D mapping with an invariant and a one-parameter 2D mapping, which is outlined in the Appendix. The two eigenvalues typically different from $+1$ describe the linearized motion around the n-cycle tangent to the level set. When the mapping is volume preserving, it follows that the periodic orbit on the surface is typically *hyperbolic* with real eigenvalues $\{\lambda, \lambda^{-1}\}$ or $\{\lambda, -\lambda^{-1}\}$ with $\lambda \neq \pm 1$, or *elliptic* with complex eigenvalues $\{e^{i\theta}, e^{-i\theta}\}$ (in the OR case, elliptic cycles are necessarily of even period). Part (iii) of the proposition shows that a curve of points of n-cycles cannot intersect any other curve of the same or different period when the cycles are hyperbolic.

Note that the determinant of the Jacobian matrix of (12) is $\det dF_1 = -1$, so that F_1 is volume-preserving and orientation-reversing. From (12) it is also seen that the origin and $(1, 1, 1)$ are the fixed points of F_1, and that the set $\hat{\mathscr{P}}$ above forms the one 3-cycle of F_1. It is further checked from (13) that these five points are precisely those at which $\nabla \hat{I} = \mathbf{0}$, and so apart from these points, all periodic orbits of F_1 are guaranteed to include $+1$ in their eigenvalue spectrum from Proposition 1. In fact, one derives $\{-1, e^{i\pi/3}, e^{-i\pi/3}\}$ for the eigenvalue spectrum of $(0, 0, 0)$, $\{-1, (3+\sqrt{5})/2, (3-\sqrt{5})/2)\}$ for $(1, 1, 1)$, and $\{-1, 9+4\sqrt{5}, 9-4\sqrt{5}\}$ for the 3-cycle. The eigenvalue spectrum of $(1, 1, 1)$ and the 3-cycle formed from $\hat{\mathscr{P}}$ are nevertheless related because of a symmetry property of the dynamics of F_1 to be discussed next. Before that we note that the typical picture of a mapping described by Proposition 1 is for its 3D phase space to contain "strings" of n-cycles which can only intersect at points which satisfy the linearization condition of (iii). Indeed, when this condition is met for F_1, we will see in the next Section that intersections can, and typically do, occur.

Recalling that the invariant surfaces M_μ have tetrahedral symmetry, it is natural to look at the way the mapping F_1 transforms under the tetrahedral group. Similar to ref. 28, we say that a mapping L *commutes with a (symmetry) group* $\mathscr{H}$ of mappings if for all $h \in \mathscr{H}$ there exists $h' \in \mathscr{H}$ such that

$$L \circ h = h' \circ L \tag{19}$$

The term "to commute with a group" is sometimes, e.g., in bifurcation theory, used in a more restricted sense to mean that L commutes with each element of $\mathscr{H}$ separately; cf. p. 13 of ref. 29). Our definition (19), however, will lead later, in the case of trace maps, to the familiar group-theoretic concept of $\mathscr{H}$ being a normal subgroup of a larger group of mappings containing L and $\mathscr{H}$. In these more algebraic terms, (19) means that L induces a homomorphism or an automorphism $h' = \varphi_L(h)$ on $\mathscr{H}$; compare Chapter 9 of ref. 30 for this and related concepts. Furthermore, if φ_L is an automorphism on a finite group $\mathscr{H}$ of order N, it is useful to attach a permutation $\pi \in S_N$ to L which summarizes the action. If in the case of F_1 we define the subgroup Σ of the tetrahedral group by

$$\Sigma = \{Id, \sigma_1, \sigma_2, \sigma_3\} \simeq \mathbb{Z}_2 \otimes \mathbb{Z}_2 \tag{20}$$

where σ_i are the pairwise sign changes, e.g., $\sigma_1: (x, y, z) \mapsto (x, -y, -z)$, etc., we obtain the following result:

Proposition 2. The Fibonacci trace map F_1 commutes with the group Σ via

$$F_1 \circ \sigma_i = \sigma_{i-1} \circ F_1 \tag{21}$$

with $\sigma_0 := \sigma_3$. Thus, F_1^3 commutes with each element of Σ.

The proof is straightforward by direct verification of (21). Referring to above, the permutation induced by (21) is

$$\pi = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \in S_3$$

(S_3 rather than S_4 suffices since the action on Id is trivial). One consequence is that if $\mathbf{x}$ is a point of a j-cycle, then $\sigma_i \mathbf{x}$ is a point of a j-cycle, a $(3j)$-cycle, or a $(j/3)$-cycle (if $3 \mid j$). Moreover, the eigenvalue spectra of the two cycles containing $\mathbf{x}$ and $\sigma_i \mathbf{x}$ are related correspondingly. With $j = 1$ and $\mathbf{x} = (1, 1, 1)$, we see why the set $\hat{\mathscr{P}}$ is a 3-cycle, and why the eigenvalues of its linearization are cubes of the eigenvalues belonging to the fixed point $\mathbf{x}$.

Among the invariant sets $\mathscr{M}_\mu$, the case of $\mathscr{M}_0^c$ is particularly interesting because the compact part of the invariant surface can be parametrized with two variables ϑ_1 and ϑ_2,[23,24] via

$$S: \begin{pmatrix} \vartheta_1 \\ \vartheta_2 \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \cos 2\pi\vartheta_1 \\ \cos 2\pi\vartheta_2 \\ \cos 2\pi(\vartheta_1 + \vartheta_2) \end{pmatrix} \tag{22}$$

With this parametrization, we find

$$F_1|_{\mathcal{M}_0^c} \circ S = S \circ R_1 \tag{23}$$

where $F_1|_{\mathcal{M}_0^c}$ is the dynamics on $\mathcal{M}_0^c$ and R_1 is the substitution matrix for the Fibonacci chain.[8] [Of course, Eq. (23) can be used to solve the dynamics on the whole of $\mathcal{M}_0$ if we allow complex ϑ_i. But, as will become clear in a moment, we wish to treat the case $\mathcal{M}_0^c$ separately.] Because cos $2\pi\vartheta$ identifies ϑ and $\vartheta + m$, $m \in \mathbb{Z}$, the action of the integer matrix R_1 on the right-hand side of Eq. (23) can be considered to be acting on the torus (i.e., take ϑ mod 1). Consequently, we can replace R_1 in (23) by L_{R_1}, where

$$L_{R_1}: \begin{pmatrix} \vartheta_1 \\ \vartheta_2 \end{pmatrix} \mapsto \begin{pmatrix} \vartheta_2 \\ \vartheta_1 + \vartheta_2 \end{pmatrix} \pmod 1 \tag{24}$$

and $\vartheta_1, \vartheta_2 \in (-1/2, 1/2]$. That is, the dynamics of F_1 on $\mathcal{M}_0^c$ is *semi-conjugate* to that of L_{R_1}, which is an example of a hyperbolic toral automorphism, or *Anosov* system. This makes its dynamics on this surface solvable. The periodic orbits of hyperbolic toral automorphisms are dense in the phase space and the motion is homogeneously chaotic; see, e.g., ref. 25 for more details. However, on $\mathcal{M}_0^c$, it should be stressed that we only have a *semiconjugacy* between F_1 and L_{R_1} and not a conjugacy (or equivalence) as suggested by some previous authors.[26,27] This is because the transformation S is not uniquely invertible, in the sense that we must allow a choice of the two solutions in $(-1/2, 1/2]$ of cos $2\pi\vartheta_1 = x$ when $x \in [-1, 1]$, in order to be able to cover via the parametrization S both of the z values on $\mathcal{M}_0$ for given values of x and y. In other words, the transformation S further identifies points of the torus, namely $(\vartheta_1, \vartheta_2)$ and $(-\vartheta_1, -\vartheta_2)$. Consequently, the action of F_1 on $\mathcal{M}_0^c$ is described by the quotient of a toral automorphism by reflection through the origin $(0, 0)$ and we find the following results:

Proposition 3. On $\mathcal{M}_0^c$, the Fibonacci trace map F_1 acts as a *pseudo-Anosov* mapping. The number of fixed points of F_1^n on $\mathcal{M}_0^c$ is given by $\mathrm{tr}(R_1^n)$, where R_1 is the substitution matrix (8), and there are periodic orbits of every period. The eigenvalue spectrum of an n-cycle on $\mathcal{M}_0^c$, excepting the set $\mathcal{P}$ of Eq. (18), is either $\{1, \tau^n, (1-\tau)^n\}$ or $\{1, -\tau^n, -(1-\tau)^n\}$, where τ is the golden ratio.

Proof. The surface $\mathcal{M}_0^c$ can be mapped onto the sphere $\mathcal{S}$ of radius $\sqrt{3}$ around the origin (on which the set $\mathcal{P}$ lies) by radial projection $\Pi: \mathcal{M}_0^c \mapsto \mathcal{S}$. Then $F_1|_{\mathcal{M}_0^c}$ induces a homeomorphism $f_0: \mathcal{S} \mapsto \mathcal{S}$, where

$$f_0 = \Pi \circ F_1|_{\mathcal{M}_0^c} \circ \Pi^{-1}$$

Equation (23) shows that $f_0 \circ \Pi S = \Pi S \circ R_1$, where $\Pi S := \Pi \circ S$ inherits from S the identification of $(\vartheta_1, \vartheta_2)$ and $(-\vartheta_1, -\vartheta_2)$. It follows that f_0 is a homeomorphism of the two-sphere $\mathscr{S}$ equal to the quotient of the hyperbolic toral automorphism or Anosov map, L_{R_1}, by reflection through the origin $(\vartheta_1, \vartheta_2)$. As such, it is therefore a pseudo-Anosov map; see ref. 28, and refs. 50 and 14 for the general definition and properties of such mappings. The fixed points, or singularities, of the quotient by the reflection through the origin are given by

$$(\vartheta_1, \vartheta_2) = \{(0,0), (\tfrac{1}{2}, \tfrac{1}{2}), (0, \tfrac{1}{2}), (\tfrac{1}{2}, 0)\}$$

and these map under $\mathscr{S}$ and ΠS to the set $\mathscr{P}$ of pinches, and are the so-called "1-prongs" of the pseudo-Anosov map. The fixed points of F_1^n on $\mathscr{M}_0^c$, i.e., $\{\mathbf{x} \in \mathscr{M}_0^c \mid F_1^n \mathbf{x} = \mathbf{x}\}$, follow from

$$F_1^n|_{\mathscr{M}_0^c} \circ S = S \circ R_1^n \tag{25}$$

which is a simple consequence of the semiconjugacy (23). From (25), they are given by $\mathbf{x} = S\boldsymbol{\vartheta}$, where

$$R_1^n \boldsymbol{\vartheta} = \pm \boldsymbol{\vartheta} \pmod 1 \tag{26}$$

The number l_n of solutions to (26), after identification of solutions $\boldsymbol{\vartheta}$ and $-\boldsymbol{\vartheta}$, is $l_n = \mathrm{tr}(R_1^n)$. Now a fixed point of F_1^n need not be a point of an n-cycle; it may belong to any k-cycle with $k \mid n$. However, a further consequence of Eq. (26), together with ref. 28, is that $F_1|_{\mathscr{M}_0^c}$ has periodic orbits of all periods, i.e., possesses a genuine n-cycle for every $n \in \mathbb{N}$. The eigenvalue spectrum of cycles of F_1 on $\mathscr{M}_0^c$ follows from differentiating both sides of (25) and evaluating at $\boldsymbol{\vartheta}$ satifying (26). No conclusions can be drawn for $dF_1^n|_{\mathscr{M}_0^c}(S\boldsymbol{\vartheta})$ if $dS(\boldsymbol{\vartheta})$ vanishes, which happens precisely at those points that are the singularities of the quotient given above, and that correspond to the four pinches. Otherwise, we find if λ is an eigenvalue of R_1^n, then $\lambda(-\lambda)$ is an eigenvalue of $dF_1^n|_{\mathscr{M}_0^c}(S\boldsymbol{\vartheta})$ according as $\boldsymbol{\vartheta}$ satisfies (26) with the $+(-)$ sign. The additional eigenvalue $+1$ follows from Proposition 1. ■

Note that if f_0 in the above proof were really Anosov rather than pseudo-Anosov, then because Anosov systems are structurally stable, dynamics on an open set of nearby level sets $\mathscr{M}_\mu^c$ would also be Anosov. But pseudo-Anosov systems are not structurally stable, so we cannot draw this conclusion.

Using Eq. (26) and $\mathbf{x} = S\boldsymbol{\vartheta}$, the points of periodic orbits on $\mathscr{M}_0^c$ can be calculated explicitly if desired. If we just want to count the number p_n of n-cycles on $F_1|_{\mathscr{M}_0^c}$, we can use Proposition 3. We know that

$l_n = \mathrm{tr}(R_1^n) = \tau^n + (1-\tau)^n$ gives the total number of fixed points of $F_1^n|_{\mathcal{M}_0^c}$. Here l_n are nothing but the Lucas numbers,[31] defined through $l_0 = 2$, $l_1 = 1$, and $l_{n+1} = l_n + l_{n-1}$. To get p_n we must subtract the number of periodic points belonging to k-cycles with $k \mid n$ but $k < n$, and then divide by n. In other words,

$$l_n = \sum_{k \mid n} k p_k \tag{27}$$

From the Möbius inversion formula[32,31] we then get the formula

$$p_n = \frac{1}{n} \sum_{k \mid n} \mu\left(\frac{n}{k}\right) l_k \tag{28}$$

where $\mu(d)$ is the Möbius function (not to be mixed up with the μ for the level sets used elsewhere). The first numbers (starting with p_1) for the Fibonacci case explicitly read 1, 1, 1, 1, 2, 2, 4, 5, 8, 11, 18, 25,.... Certainly there are periodic orbits of F_1 that do not lie on $\mathcal{M}_0^c$, though they might belong to one-parameter families of periodic orbits originating on $\mathcal{M}_0^c$; compare Proposition 1. This way, the numbers in (28) provide useful lower bounds on the number of families of n-cycles of F_1.

Let us close this Section with several results on existence versus non-existence of bounded orbits of the Fibonacci trace map on the surfaces $\mathcal{M}_\mu$ for $\mu \leqslant 0$. Since the compact surface $\mathcal{M}_\mu^c$, $-1 \leqslant \mu \leqslant 0$, is an invariant set and since it is bounded by the box $\mathcal{B}$, we immediately have the following result.

Proposition 4. For $-1 \leqslant \mu \leqslant 0$, all orbits that start on the compact surface $\mathcal{M}_\mu^c$ stay there and are bounded, both under forward and backward iteration.

The distinction between periodic and quasiperiodic orbits on $\mathcal{M}_\mu^c$ is an interesting question to which we will return later in the context of bifurcations with μ as the bifurcation parameter. Less obvious than Proposition 4 is the following observation:

Proposition 5. All orbits on the surfaces $\mathcal{C}_\mu^{(\varepsilon_x, \varepsilon_y, \varepsilon_z)}$, $\mu < 0$, $\varepsilon_x \varepsilon_y \varepsilon_z = 1$, are unbounded. Therefore, no periodic orbits exist on the disconnected cones.

Proof. Observe from (11) and (12) with $x_{n+1} = y_n$ and $x_{n+2} = z_n$ that the dynamics of F_1 is equivalent to the difference equation $x_{n+1} + x_{n-2} = 2x_n x_{n-1}$. We have $|x_n| > 1$ for all n on the cones because they do not touch the box $\mathcal{B}$ for $\mu < 0$. If, for any $n \in \mathbb{N}$, we have

$|x_n x_{n-1}| \geqslant |x_{n-2}|$, the orbit diverges. To see this, we generalize an argument given in ref. 33. Assume the inequality holds, whence

$$\begin{aligned}|x_{n+1}| &= |2x_n x_{n-1} - x_{n-2}| \\ &\geqslant 2\,|x_n x_{n-1}| - |x_{n-2}| \geqslant |x_n x_{n-1}| > \max\{|x_n|, |x_{n-1}|\} \geqslant |x_n| > 1\end{aligned}$$

But this also implies $|x_{n+1}| > |x_{n-1}|$ and thus the relation $|x_{n+1} x_n| > |x_{n-1}|$, which is a stronger version of our original assumption involving the next triple of successive iterates. Repeating the argument, we get the inequality $|x_{m+1} x_m| > |x_{m-1}| > 1$ for all $m \geqslant n$. From the inequality used above, we then obtain $|x_{m+1}| > |x_m x_{m-1}|$ for $m > n$. Taking logarithms, i.a., $a_m = \ln(|x_m|) > 0$, we find $a_{m+1} > a_m + a_{m-1}$ and hence a sequence that diverges faster than a Fibonacci sequence (positive initial conditions!). Now, assuming that there *be* a bounded orbit, we had to conclude that $|x_n x_{n-1}| < |x_{n-2}|$ for all $n > 2$. But then we would get a monotonically decreasing sequence that is bounded from below by 1. But such a sequence must converge—the limit either corresponding to a fixed point of F (on $\mathscr{C}_\mu^{(+++)}$) or a 3-cycle (distributed over the other cones)! Since neither of them exists on the cones for $\mu < 0$, we have a contradiction: no bounded orbit (and thus no periodic orbit) can exist on the disconnected cones. ■

Slightly different in nature is the structure of the orbits on the cones $\mathscr{C}_0^{(\varepsilon_x, \varepsilon_y, \varepsilon_z)}$, i.e., for $\mu = 0$. As already mentioned, the transformation S of (22) is also a parametrization of part of $\mathscr{M}_0$ if we take ϑ_i imaginary, equivalently if we make the replacement $\cos 2\pi\vartheta \to \cosh \vartheta$. We will call this transformed version of S the *hyperbolic parametrization* S'. It parametrizes the cone $\mathscr{C}_0^{(+++)}$, which, as already mentioned, is an invariant set [wherefore the orbits which start from it always keep signature $(+\,+\,+)$]. Recalling the tetrahedral symmetry of $\mathscr{M}_\mu$, in particular $\mathscr{M}_0^{nc}$, the parametrizations for the three remaining cones are given, respectively, by the composition of σ_1, σ_2, and σ_3 of (20) with S'. For the dynamics on these cones we find the following result.

Proposition 6. The only forward (backward) bounded orbits on the noncompact cone $\mathscr{C}_0^{(+++)}$ are obtained from the images under the hyperbolic parametrization of the sets of points $\{(\vartheta_{n-1}, \vartheta_n)\}$ with $\vartheta_n = (1-\tau)^n\, \vartheta_0$ $(\vartheta_n = \tau^n \vartheta_0)$. These points also generate the only forward (backward) bounded motion among the three remaining cones of $\mathscr{M}_0^{nc}$, with the modification that the points hop cyclically between the three cones as a consequence of the symmetry transformations under the group Σ.

Proof. Equation (23) also holds with S replaced by S', the hyperbolic parametrization, and with $F_1|_{\mathcal{M}_0^c}$ replaced by $F_1|_{\mathscr{C}_0^{(+++)}}$. Now R_1 must be considered as a linear mapping of the entire plane, i.e., we *cannot* take the ϑ_i mod 1. The only forward (backward) bounded orbits of the latter linear hyperbolic mapping are those on the stable (unstable) manifold of the origin (0, 0), which is the only fixed point of R_1. These are given, respectively, by the sets of points listed in the Proposition. For the dynamics on the three remaining cones, parametrized by $\sigma_i \circ S'$, $\{i = 1, 2, 3\}$, using the above and Proposition 2 gives $F_1 \circ (\sigma_i \circ S') = \sigma_{i-1} \circ (F_1 \circ S') = (\sigma_{i-1} \circ S') \circ R_1$. Again the dynamics is generated by the hyperbolic linear map corresponding to R_1, except that the image point under this map is sent onto a different cone to the initial point. Every third iterate returns to the same cone because F_1^3 commutes with σ_i. ■

Note that the bounded orbits of the above proposition converge under forward (backward) iteration to precisely the four pinches $\mathscr{P}$ of (18), which are those four vertices of the cube $\mathscr{B}$ where the cones $\mathscr{C}_0^{(\varepsilon_x, \varepsilon_y, \varepsilon_z)}$ touch the compact surface $\mathcal{M}_0^c$. One physical consequence of Proposition 6 is the exact localization of the critical point of the Ising quantum chain on the Fibonacci quasilattice.[34,35]

4. REVERSIBILITY AND EXISTENCE OF AN INVARIANT: DYNAMICAL IMPLICATIONS

Having shown the existence of the invariant (13) or the Fibonacci trace map (12) and some of its dynamical consequences, we now discuss another important dynamical property of the trace map, namely its time-reversal symmetry. Although this symmetry has been noticed before,[36,23,24] its consequences have escaped a systematic study so far. The symmetry can be seen, for instance, from writing out the ubiquitous 6-cycle of (12), which plays an important role in determining scaling in the electronic band structure.[26] This 6-cycle is

$$\begin{aligned} P_1(0, a, 0) &\to P_2(a, 0, 0) \to P_3(0, 0, -a) \\ &\to P_4(0, -a, 0) \to P_5(-a, 0, 0) \to P_6(0, 0, a) \end{aligned} \tag{29}$$

where a is an arbitrary number. Notice that the first point P_1 and half-way point P_4 of the 6-cycle lie on the plane $\{x = z\}$. Moreover, the forward-going points from P_1, namely P_2 and P_3, are reflections in this plane of the backward-going points from P_1, namely P_6 and P_5. These observations can be explained by noting that F_1 in (12) can be written as the composition $F_1 = H_1 \circ G_1$, where

$$H_1 \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} y \\ x \\ 2xy - z \end{pmatrix}, \qquad G_1: \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} z \\ y \\ x \end{pmatrix} \tag{30}$$

are involutions, i.e., $H_1 \circ H_1 = G_1 \circ G_1 = Id$. Mappings like the Fibonacci trace map which can be written as the product of involutions are called *reversible* mappings. Periodic orbits like (29) that are left invariant under G_1 and H_1 are called *symmetric*.

More generally, a mapping $L: \mathbb{R}^n \to \mathbb{R}^n$ is called *reversible*[37,38,12] if there exists a mapping G satisfying

$$L \circ G \circ L = G, \qquad G \circ G = Id \tag{31}$$

This is equivalent to the definition given above of L being able to be written as the product of two involutions

$$L = H \circ G, \qquad H \circ H = G \circ G = Id \tag{32}$$

with $H := L \circ G$. An involution G that satisfies (31) for a given mapping L is called a *reversing symmetry* of L. This is because G creates a conjugacy between L and its inverse via

$$L^{-1} = G \circ L \circ G^{-1} \tag{33}$$

noting $G^{-1} = G$. Equation (33) is a third equivalent definition of reversibility and indicates why reversibility is a generalization of time-reversal symmetry. The application of G to a trajectory of L is also a trajectory of L when followed in the opposite sense.

Reversible mappings have been studied since the pioneering work of Birkhoff,[39] and reversibility has strong dynamical consequences. We now recall some of the properties of reversible mappings which we will use below; for details and derivations we refer to a recent review.[12]

R1 If L is reversible with reversing symmetry G, then $L^i \circ G$, $i \in \mathbb{Z}$, is also a reversing symmetry of L. We call $\{L^i \circ G\}$ the family of reversing symmetries generated by G.

R2 Any power of a reversible mapping L is reversible with at least the same reversing symmetries as the mapping, i.e., $L^i \circ G \circ L^i = G$, $i \in \mathbb{Z}$.

R3 Reversibility is preserved under coordinate transformations, or conjugacies: the transformed mapping $P \circ L \circ P^{-1}$ has reversing symmetry $P \circ G \circ P^{-1}$.

R4 A reversible mapping L may possess more than one family of reversing symmetries, in which case it is called multiply reversible. For instance, if L commutes with a mapping T, then $G' = G \circ T$ or $G' = T \circ G$ are reversing symmetries if they are involutions, and are distinct from the family $\{L^i \circ G\}$ provided $T \neq L^i$, $i \in \mathbb{Z}$.

R5 If Γ is an invariant set of a reversible mapping L (i.e., $L\Gamma = \Gamma$), then $G\Gamma$ is also invariant, and $G\Gamma$ equals the image of Γ under all the symmetries of the family $\{L^i \circ G\}$. If $G\Gamma = \Gamma$ then Γ is called *symmetric* (more specifically, G-symmetric). The eigenvalue spectrum of a symmetric n-cycle comprises reciprocal pairs, i.e., if λ is an eigenvalue, so is λ^{-1}. Periodic orbits that are not symmetric are called *asymmetric*, and come in pairs related by G. If $\mathbf{p}_0$ is a periodic point belonging to one of the asymmetric cycles, then $G\mathbf{p}_0$ belongs to the other cycle of the pair, and the eigenvalues of dL^n at $\mathbf{p}_0$ are the reciprocals of those at $G\mathbf{p}_0$.

Reversibility facilitates the location of symmetric periodic orbits, which results in significant computational advantages.[12] We observed above that the 6-cycle (29) of the Fibonacci mapping is invariant under the reversing symmetry G_1 in (30), and so is symmetric. It contains two points P_1 and P_4 which are fixed points of G_1. More generally, with $\mathrm{Fix}(G) := \{\mathbf{x} \mid G\mathbf{x} = \mathbf{x}\}$, etc., we can say the following:

R6 An orbit of L is a symmetric *even* $2j$-cycle if and only if it contains precisely two points $\mathbf{x}$ and $L^j\mathbf{x}$ in $Fix(G)$, or precisely two such points in $Fix(L \circ G)$. An orbit of L is a symmetric *odd*$(2j+1)$-cycle if and only if it contains precisely one point $\mathbf{x}$ in $Fix(G)$ and precisely one point $L^{j+1}\mathbf{x}$ in $Fix(L \circ G)$.

An equivalent characterization of symmetric cycles is in terms of intersections of the fixed-point sets of the full family of reversing symmetries, defined by $\mathcal{T}_i := \{\mathbf{x} \mid L^i \circ G\mathbf{x} = \mathbf{x}\}$. The $\mathcal{T}_i$ can be generated from $\mathcal{T}_0 := Fix(G)$ and $\mathcal{T}_1 := Fix(L \circ G)$ via $\mathcal{T}_{2j} = L^j \mathcal{T}_0$ and $\mathcal{T}_{2j+1} = L^j \mathcal{T}_1$.

R7 If $\mathbf{y} \in \mathcal{T}_p \cap \mathcal{T}_q$ and $(p-q)$ is odd, then $\mathbf{y}$ belongs to a symmetric odd-period cycle. If $(p-q)$ is even, then $\mathbf{y}$ belongs to a symmetric even-period cycle or else $L^{(p-q)/2}\mathbf{y} = \mathbf{y}$. The dimension of the sets $\mathcal{T}_p$ and $\mathcal{T}_q$ and their intersection set gives an indication of the dimension of the set of symmetric periodic orbits of a particular period.

Here we apply and demonstrate these reversibility properties on 3D reversible mappings, in particular the Fibonacci mapping. A decomposition of the Fibonacci mapping (12) into involutions is provided by (30). With reference to R4, we can ask if F_1 is multiply reversible because it has nontrivial commutators. It is easy to verify, for instance, that (12) commutes with no nontrivial linear maps. However, we have seen via Proposition 2 that F_1^3 commutes with the linear maps σ_i belonging to the group Σ of (20). Also it is reversible with reversing symmetry G_1 from R2. One readily calculates that $G_1 \circ \sigma_2 = \sigma_2 \circ G_1 : (x, y, z) \mapsto (-z, y, -x)$ is an involution, whereas $G_1 \circ \sigma_1 = \sigma_3 \circ G_1 : (x, y, z) \mapsto (-z, -y, x)$ and $G_1 \circ \sigma_3 = \sigma_1 \circ G_1$: $(x, y, z) \mapsto (z, -y, -x)$ are diffeomorphisms of order 4, the squares of which equal σ_2. From R4 above, F_1^3 is multiply reversible with the family of reversing symmetries $\{F_1^{3i} \circ G'\}$ with $G' = G_1 \circ \sigma_2$.

While $G_1 \circ \sigma_1$ and $G_1 \circ \sigma_3$ are not reversing symmetries of F_1^3 by the above definition, they motivate considering a generalization of reversibility. A mapping is called *weakly reversible*[38] if it satisfies $L \circ G \circ L = G$, equivalently (33), where G is an arbitrary homeomorphism, i.e., not necessarily an involution. We call such a G a *weak reversing symmetry*. All reversible mappings are weakly reversible but not vice versa. Nevertheless, the properties R1–R7 meaningfully extend to weakly reversible mappings.[40] Significantly in R4, without the restriction that G be an involution, the composition of a commuting mapping T and a weak reversing symmetry of a weakly reversible mapping is always a weak reversing symmetry. This is why F_1^3 is weakly reversible with respect to $G_1 \circ \sigma_1$ and $G_1 \circ \sigma_3$. Also in general, the composition of two weak reversing symmetries is a commutator for L. One can then study the group built up from commuting maps and weakly reversing symmetries.[40],[5] In particular, even powers G^{2j} of a weak reversing symmetry G typically generate nontrivial commutators for L, whereas odd powers G^{2j+1} are weak reversing symmetries belonging to different families. If the weak reversing symmetry is of finite order and the mapping L is not an involution, then G is necessarily of even order. The fact that this order need not be 2 as in reversibility means that, with reference to property R5 for weakly reversible mappings, there may be

[5] We use the adjective "weak" for a noninvolutory reversing symmetry (unlike ref. 40) only to mirror Sevryuk's previous terminology of "weakly reversible."[38] On a purely algebraic level, it is most natural to consider the conjugacy (33) between L and L^{-1} without restricting G to be an involution. In the present context, it will turn out that if a trace map has a genuine weak reversing symmetry, then it typically also has a true reversing symmetry, i.e., an involutory one.

more than 2 asymmetric cycles with reciprocal eigenvalues—every odd power of G can map a cycle to a different partner, for instance.

In the present case, at the level of F_1 instead of F_1^3, it is interesting to note that F_1 is not (weakly) reversible with respect to any of $G_1 \circ \sigma_i$. We have an example of a more general reversibility situation again. F_1 is an example of a mapping L which commutes with a group $\mathscr{H}$ via (19) and also satisfies $L \circ G = G \circ L^{-1}$. It follows that for any element of the form $k = h \circ G$ there exists an element $k' = h' \circ G$ such that

$$L \circ k = k' \circ L^{-1} \tag{34}$$

Equation (34) represents the reversibility analog of (19), see also (41). The algebraic structure of (34) is not our purpose here, but will be given elsewhere.[42] One of our interests later in the paper will be to investigate, among a large class of trace maps, how commonly a commuting group and a (weak) reversing symmetry occur, these being the ingredients of this generalized structure. We also note that the generalizations (19) and (34) of symmetry and reversing symmetry have also been noted and investigated recently in some other dynamical systems.[41]

We concentrate now on the symmetric cycles of F_1 of (12) with respect to G_1 and its family. From (30), $Fix(G_1)$ is a two-dimensional set, the plane $\{(x, y, x) \mid x, y \in \mathbb{R}\}$, whereas $Fix(H_1)$ is a one-dimensional set, the curve $\{(y, y, y^2) \mid y \in \mathbb{R}\}$. The sets intersect in the two points $(0, 0, 0)$ and $(1, 1, 1)$, which are consequently (symmetric) fixed points of F_1. In property R7, the dimensions of the fixed sets $\mathscr{T}_{2j}$ and $\mathscr{T}_{2j+1}$, respectively, are 2 and 1 for all j. We expect isolated symmetric odd-cycles containing the point of intersection of a 2D surface $\mathscr{T}_{2j}$ and a 1D curve $\mathscr{T}_{2k+1}$, and curves of nonisolated even-cycles generated by the intersections of two of the 2D surfaces $\mathscr{T}_{2j}$. We do not expect any even-cycles from intersections of the 1D curves of $\mathscr{T}_{2j+1}$. We observed from Proposition 1 that in 3D mappings with an invariant, curves of periodic orbits would also be expected. However, in that case, one would not typically expect some odd cycles to be isolated. This feature stems from the reversibility property. By possessing curves of symmetric even-cycles but isolated symmetric odd-cycles, the Fibonacci mapping is typical of 3D orientation-reversing reversible mappings, most of which of course do not possess an integral of motion like F_1. At the level of asymmetric cycles in such reversible mappings, we do not expect a distinction between even and odd cycles.

We illustrate property R6 explicitly in Table I, where we show the symmetric cycles of (12) up to period 7. We find the even $2n$-cycles by trialling a point of the form (a, b, a) and demanding that the (n)th iterate also have $x = z$. We find one-parameter curves of even-cycles as predicted.

Table I. Symmetric *n*-Cycles of the Fibonacci Trace Map F_1[a]

n	Coordinates	Eigenvalue spectrum
1	$(0, 0, 0)$	$\{-1, e^{i\pi/3}, e^{-i\pi/3}\}$
1	$(1, 1, 1)$	$\left\{-1, \frac{3+\sqrt{5}}{2}, \frac{3-\sqrt{5}}{2}\right\}$
2	$\left(a, \frac{a}{2a-1}, a\right), \left(\frac{a}{2a-1}, a, \frac{a}{2a-1}\right)$	$\{1, \lambda, \lambda^{-1}\}$: $\lambda+\lambda^{-1}=\frac{8a^2-2a+1}{2a-1}$
3	$(-1, 1, -1), (-1, -1, 1)$	$\{-1, 9+4\sqrt{5}, 9-4\sqrt{5}\}$
4	$(-\frac{1}{2}, a, -\frac{1}{2}), (-\frac{1}{2}, \frac{1}{2}-a, -\frac{1}{2})$	$\{1, \lambda, \lambda^{-1}\}$: $\lambda+\lambda^{-1}=8a(1-2a)+1$
5	$\left(b^2, \frac{b+1}{2b}, b^2\right), (b, b, b^2)$	$\{-1, 1, 1\}$
6	$(0, a, 0), (0, -a, 0)$	$\{1, \lambda, \lambda^{-1}\}$: $\lambda+\lambda^{-1}=2(8a^4+1)$
6	$\left(a, \frac{a}{2a+1}, a\right), \left(\frac{-a}{2a+1}, -a, \frac{-a}{2a+1}\right)$	$\{1, \lambda, \lambda^{-1}\}$: $\lambda+\lambda^{-1}=\frac{2(8a^2+2a+1)(-32a^4-16a^3-4a^2+4a+1)}{(2a+1)^3}$
7	$\left(2c^3-c, \frac{2c^2+c-1}{4c^2-2}, 2c^3-c\right), (c, c, c^2)$	$\{-1, 1, 1\}$

[a] For even cycles, the coordinates listed are the two points in $\mathrm{Fix}(G_1)$. For odd cycles, the coordinates listed are the point in $\mathrm{Fix}(G_1)$ followed by the point in $\mathrm{Fix}(H_1)$. The variable a is arbitrary, while b for $n=5$ is any root of $4b^3+4b^2+2b+1=0$ and c for $n=7$ is any root of $16c^6+16c^5-2c-1=0$.

Symmetric odd $(2n+1)$-cycles, on the other hand, are found by starting with a point (b, b, b^2) and demanding that the (n)th iterate satisfy $x=z$. Symmetric odd-cycles are isolated, again as expected from the dimension of $Fix(H_1)$. Beyond the 7-cycle, analytical calculation of symmetric cycles becomes increasingly cumbersome. At the level of the 8-cycles, irrational expressions start to appear. There are two families of 8-cycles generated by the curves $(a, d_+(a), a)$ and $(a, d_-(a), a)$, where

$$d_\pm(a) = \frac{a(2a-1)(4a+1) \pm [a(2a-1)(2a^2+3a+2)]^{1/2}}{4a(4a^2-1)} \tag{35}$$

Table I also shows the eigenvalue spectrum for each symmetric n-cycle. The structure of the spectrum for each cycle is a consequence of the combination of F_1 in (12) being both reversible and possessing an invariant [note that G_1 and H_1 also preserve $\hat{I}$ of (13)]. The consequences of the latter property were given in Proposition 1(i) of the previous Section, whereas reversibility gives the following result.

Proposition 7. In a 3D (weakly) reversible mapping, the eigenvalues of symmetric even-cycles are $\{1, \lambda, \lambda^{-1}\}$; for symmetric odd-cycles, the eigenvalues are $\{1, \lambda, \lambda^{-1}\}$ if the mapping is OP and $\{-1, \lambda, \lambda^{-1}\}$ if the mapping is OR. If the mapping is (weakly) reversible and also has an invariant I, the eigenvalues are not further restricted, except for symmetric odd-cycles of OR mappings for which ∇I is nonvanishing, which must have spectrum $\{-1, 1, 1\}$.

Proof. This result is a straightforward consequence of the reciprocal pairing of eigenvalues of symmetric cycles given in R5, which extends to weakly reversible mappings, together with the fact that in odd dimensions this forces one eigenvalue to be self-reciprocal, i.e., ± 1. When the (weakly) reversible mapping also has an invariant, odd-cycles are forced to have an eigenvalue $\{+1\}$ from Proposition 1, which leads to their degenerate spectrum. ■

Note that we could also conclude the spectrum $\{-1, 1, 1\}$ for odd-cycles of OR (weakly) reversible 3D mappings with an invariant by using the fact that they are isolated. From Proposition 1(ii), a cycle can only be isolated by having more than one eigenvalue equal to $\{+1\}$.

From Table I, we can further observe the effects of the symmetry property of F_1 given by Proposition 2 and the ensuing discussion. Apart from linking the fixed point $(1, 1, 1)$ and the 3-cycle, as described previously, we find that Σ links the 2-cycle and the second of the 6-cycles. For example, the image of $(a, a/(2a-1), a)$ by σ_2 becomes $(a, a/(2a+1), a)$

when $a \to -a$. Furthermore, the eigenvalues of the 6-cycle are cubes of those of the 2-cycle. This linking of cycles continues throughout the periodic orbits of F_1—the image of points of the 4-cycle under σ_i give points of a 12-cycle. Some cycles like the origin and the first-listed 6-cycle are self-invariant under the elements of Σ. Some of these observations of cycles linked by the symmetry property of F_1 were made previously in ref. 7.

It is interesting to use Table I to look at the intersections of the curves of symmetric even-cycles. For example, the curve of symmetric 2-cycles intersects the curves of symmetric 4-cycles in the points $(-1/2, 1/4, -1/2)$ and $(1/4, -1/2, 1/4)$. From (13) these points of intersection lie on the level set (14) with $\mu_1 = -9/16 = -0.5625$. Similarly, the curves of symmetric 4-cycles intersect the family of curves of symmetric 8-cycles generated from the family $(a, d_+(a), a)$ in four points at which $\mu_2 = -5/16 = -0.3125$. Since each G_1-symmetric even-cycle has two points on the symmetry plane $\{x = z\}$, we can plot the intersecting curves of 2-, 4-, and 8-cycles on the projection of this plane onto the plane $\{z = 0\}$. Likewise we can plot the intersection of the compact invariant surfaces $\mathcal{M}_\mu^c$ for $\mu \in [-1, 0]$ with the symmetry plane. This produces the contours

$$\{(x, y) \in \mathbb{R}^2 \mid 2x^2(1-y) + y^2 - 1 = \mu\} \tag{36}$$

Figure 2 shows these pictures.

An alternative way to calculate the point of intersection of the 2-cycle and the 4-cycle in Table I is to calculate dF_1^2 evaluated at a point of the 2-cycle curve and to determine, as a function of position on the curve, when the eigenvalue spectrum $\{1, \lambda, \lambda^{-1}\}$ equals $\{1, -1, -1\}$, i.e., satisfies $\lambda + \lambda^{-1} = -2$. This is equivalent to $\mathrm{tr}(dF_1^2) = -1$. Since position on the curve of 2-cycles is a function of μ, this process determines a value of μ and hence a surface $\mathcal{M}_\mu$ on which the stability of the 2-cycle is transitional between being elliptic and hyperbolic. The correspondence between this and the intersection of the 2-cycle and 4-cycle follows from Proposition 1(iii), which shows that the only possibility for a point of intersection with a curve of cycles of double the period is if the linearization contains $\{-1\}$.

We have exploited the reversibility of the Fibonacci trace map to continue to calculate numerically the period-doubling intersections of 2^n-cycles and 2^{n+1}-cycles. Starting with the 8-cycle, we track numerically the 2^n-cycle along its curve on the symmetry plane $\{x = z\}$ until $\mathrm{tr}(dF_1^{2^n})$ evaluated on the curve equals -1, and note the value $\mu = \mu_n$ at which this occurs. Just beyond this μ value we find two points of a 2^{2+1}-cycle emanating on the symmetry plane from one of the two points of the 2^n-cycle on the plane. Iterating the process, we find

$$\{\mu_3 = -0.2818352255..., \mu_4 = -0.2782223174..., \mu_5 = -0.2778064018..., \\ \mu_6 = -0.2777586826, \mu_7 = -0.2777532105...\}$$

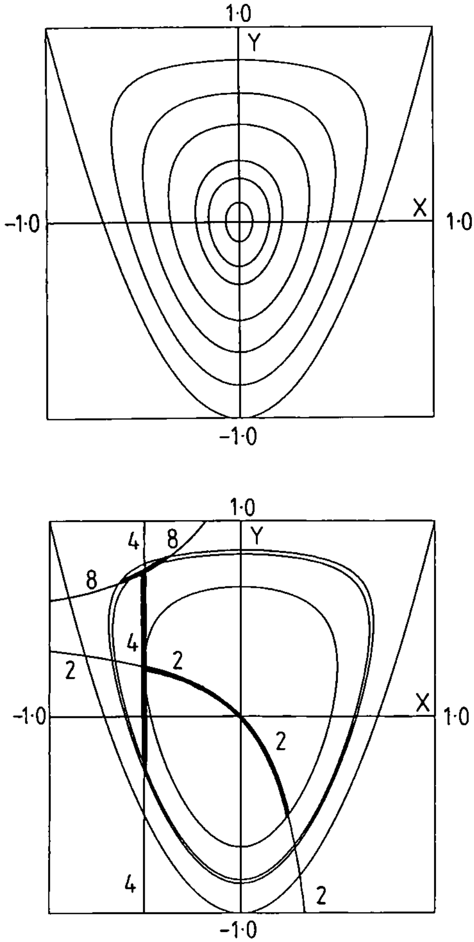

Fig. 2. Top: The projection, onto the plane $\{z=0\}$, of $\mathcal{M}_\mu^c$ intersected with $\{x=z\}=:\mathrm{Fix}(G_1)$, corresponding to μ values $\{-0.99, -0.95, -0.90, -0.75, -0.5625, -0.3125\}$ (from the origin outward). The projection of $\mathcal{M}_0^c \cap \{x=z\}$ comprises the outer parabolic curve and the line $\{y=1\}$. Bottom: Part of the top figure, showing the projections of the curves of points of symmetric 2-, 4-, and 8-cycles that lie on the plane $\{x=z\}$. The thick part of the curves indicates where the corresponding cycles are elliptic. Additionally, the projection of $\mathcal{M}_\mu^c \cap \{x=z\}$ with $\mu = -0.281835...$ is shown where a branch of the symmetric 8-cycle curve intersects symmetric 16-cycles (not shown).

Regarding the value μ_n of the invariant set (14) as a parameter, we find the scaling of this period-doubling sequence is given by

$$\delta_n = \frac{\mu_n - \mu_{n-1}}{\mu_{n+1} - \mu_n} \to 8.721... \tag{37}$$

e.g., $\delta_6 = 8.722....$ Furthermore, an estimate of the accumulation value of $\{\mu_n\}$ is $\mu_\infty = -0.27775....$

The number on the right-hand side of (37) is that associated with period-doubling sequences in 2D area-preserving or 2D reversible mappings.[12] This is not so surprising because effectively the presence of the invariant turns this into a 2D problem. The 3D mapping F_1 is related to a 2D mapping in the following way. The equation $\hat{I}(x, y, z) = \mu$ with $\hat{I}$ given by (13) defines two z values for given x, y, and μ satisfying $\mu > -1$ and $x^2 + y^2 < 1 + \mu$. They are given by

$$z_\pm = J_\pm(x, y, \mu) = xy \pm [(x^2 - 1)(y^2 - 1) + \mu]^{1/2} \tag{38}$$

The plane $\{z = 0\}$ divides these two values and for $\mu \in (-1, 0]$ divides the "balls" $\mathcal{M}_\mu^c$ into an *upper* branch on which $z > 0$ and a *lower* branch on which $z < 0$. Note that from Table I and Fig. 2 the part of the curve of symmetric 2-cycles that intersects the symmetric 4-cycles is that part that cuts across the lower branches of $\mathcal{M}_\mu^c$. If we consider F_1^2, this portion of the curve becomes a curve of symmetric fixed points extending downward from the origin across these lower branches. It can be checked that, in numerically following the period-doubling of F_1 as described above, the cycles in the cascade considered as cycles of F_1^2 continue to have all their points on the lower branches of $\mathcal{M}_\mu^c$. If we define the transformations $P: (x, y, z) \mapsto (x, y, \mu = \hat{I}(x, y, z))$ and $P^{-1}: (x, y, \mu) \mapsto (x, y, z = J_-(x, y, \mu))$, then $M = P \circ F_1^2 \circ P^{-1}$ has the form

$$M: \begin{pmatrix} x \\ y \\ \mu \end{pmatrix} \mapsto \begin{pmatrix} f(x, y, J_-(x, y, \mu)) \\ g(x, y, J_-(x, y, \mu)) \\ \mu \end{pmatrix} =: \begin{pmatrix} F(x, y, \mu) \\ G(x, y, \mu) \\ \mu \end{pmatrix} \tag{39}$$

with $f(x, y, z) = z$ and $g(x, y, z) = 2yz - x$. The mapping M is a well-defined transformation of F_1^2 in a region in which the dynamics of F_1^2 remains wholly on the lower branches of $\mathcal{M}_\mu^c$. Such a region, albeit small, can be found in a neighborhood of the hierarchy of curves of 2^n-cycles of F_1^2. In this region, M is conjugate to the reversible volume-preserving mapping F_1^2, which means that its nontrivial part, a 2D mapping, is reversible and measure-preserving (conjugate to area-preserving[12]). This is an extension of the local conjugacy argument used in the Appendix. The intersections of the curves of cycles of F_1 is a manifestation of the bifurcation behavior of area-preserving reversible 2D mappings,[12] and the latter theory implies in the present context that generically, for example, a curve of 2^n-cycles *will* intersect a curve of 2^{n+1}-cycles when the eigenvalues corresponding to motion on the level set are both equal to $\{-1\}$. Although we have not studied it numerically, we would expect via a similar argument to get cascades of *period-tupling* intersections between the curves of even symmetric cycles, with the tupling exponents of 2D area-preserving reversible mappings. Note that the mapping M above is similar to what would be obtained by trivially lifting a one-parameter 2D mapping to 3D by calling the parameter a new coordinate. As we have seen, locally and to some extent globally, such a mapping explains much of the dynamics of F_1.

We now make some remarks about the significance of $\mathcal{M}_0^c$ in the reversible dynamics of F_1.

The period-doubling process just described occurs and accumulates in the region of $\mathbb{R}^3$ foliated by the surfaces $\mathcal{M}_\mu^c$ of (14) with $-1<\mu<0$. The surface $\mathcal{M}_0^c$ provides a barrier to the process because from Proposition 3 all cycles on this surface "off" the pinches are hyperbolic. As such, they must remain hyperbolic in some range of μ either side of $\mu=0$, whereas a 2^n-cycle is elliptic on the level surfaces prior to the one where it intersects a 2^{n+1}-cycle. The hyperbolicity of the cycles of F_1 on $\mathcal{M}_0^c$ also means, via Proposition 1(ii), that all cycles lie on curves that extend to adjacent level sets, and that a given curve does not intersect any other for some range over these level sets. Studies of the energy spectrum of the tight-binding problem with a Fibonacci quasiperiodic potential suggest that on the surfaces $\mathcal{M}_\mu$ with $\mu>0$, all periodic orbits of F_1 are hyperbolic. (This suggestion follows from the fact that the spectrum has been shown to be a Cantor set.[43] On the other hand, an elliptic n-cycle in the regime $\mu>0$ would be surrounded by a finite-size ball of bounded motion via KAM theory, and would thus contribute a finite-width band to the spectrum.) The implication that all cycles for $\mu>0$ are hyperbolic again implies from Proposition 1(ii) that there are no intersections of periodic orbit curves in this range of μ. Results in ref. 36 also suggest regions in $\mathbb{R}^3$ where the curves can and cannot be in the regime $\mu>0$.

The additional fact that on $\mathcal{M}_0^c$ we have the dynamics of F_1 semiconjugate to a chaotic system, a toral automorphism, is reminiscent of an analogous scenario in 1D. In the logistic map $x \to \mu x(1-x)$, there is a period-doubling cascade from the nonzero fixed point beginning at $\mu=3$ and accumulating at $\mu_\infty=3.5699...$. For $\mu=4$, the logistic map is solvable and semiconjugate to the chaotic map of the circle $\theta \to 2\theta$ by the transformation $s\colon \theta \to \cos\theta$; cf. ref. 25, §1.8. Of course, the scaling δ characterizing the period doubling in the logistic map is the *dissipative* value 4.669....

In the 2D case we have investigated above, the period doubling δ is the *conservative* value 8.721.... It would be interesting to see if this conservative δ could be verified experimentally. There are several important physical problems which choose the regime $-1<\mu<0$ in which the period doubling occurs, the most prominent being the quasiperiodically kicked two-level system.[8]

The fact that F_1 has the invariant of motion (13) and is reversible with a reversing symmetry G_1 that preserves this invariant creates a one-parameter family of reversible dynamical systems on the 2D level sets $\mathcal{M}_\mu$ of $\hat{I}$, the parameter being μ. It is interesting to see how the reversibility works on the particular surface $\mathcal{M}_0^c$ where we have the advantage of the parametrization S in (22). Because G_1, H_1, and σ_i of the group Σ preserve this surface, we can write

$$G_1|_{\mathcal{M}_0^c} \circ S = S \circ R_{G_1}, \quad H_1|_{\mathcal{M}_0^c} \circ S = S \circ R_{H_1}, \quad \sigma_i|_{\mathcal{M}_0^c} \circ S = S \circ f_{\sigma_i} \tag{40}$$

where

$$R_{G_1}\colon \ (\vartheta_1, \vartheta_2) \mapsto (\vartheta_1 + \vartheta_2, -\vartheta_2), \qquad R_{H_1}\colon \ (\vartheta_1, \vartheta_2) \mapsto (-\vartheta_2, \vartheta_1) \tag{41}$$

and

$$\begin{aligned} f_{\sigma_1}\colon &\ (\vartheta_1, \vartheta_2) \mapsto (\vartheta_1, \vartheta_2 - \tfrac{1}{2}) \\ f_{\sigma_2}\colon &\ (\vartheta_1, \vartheta_2) \mapsto (\tfrac{1}{2} - \vartheta_1, -\vartheta_2) \\ f_{\sigma_3}\colon &\ (\vartheta_1, \vartheta_2) \mapsto (\tfrac{1}{2} - \vartheta_1, \tfrac{1}{2} - \vartheta_2) \end{aligned} \tag{42}$$

All these transformations are considered on the torus (i.e., taken mod 1) and one can alternatively take the negative of any of these transformations because of the identification of $(\vartheta_1, \vartheta_2)$ and $(-\vartheta_1, -\vartheta_2)$ by S. Noting also the representation of $F_1|_{\mathcal{M}_0^c}$ from (23) and (24), we find that the reversibility relation $F_1 \circ G_1 \circ F_1 = G_1$ and the semiconjugacies involving S induce the relation $S \circ (R_1 \circ R_{G_1} \circ R_1) = S \circ R_{G_1}$ for the 2D mappings on $\mathcal{M}_0^c$. Indeed one can choose to satisfy this by taking $R_1 \circ R_{G_1} \circ R_1 = -R_{G_1}$, which indicates the generalizations in the notion of reversibility that arise when a reversible system is mapped under a semiconjugacy (cf. also ref. 42). Note also that the representations of G_1 and H_1 on $\mathcal{M}_0^c$ are linear, but this is not true of σ_i. This is because G_1 and H_1 are themselves trace maps for appropriately chosen substitution rules, and this explains the suggestive "R" notation used for their representations, whereas σ_i can never be trace maps. We will return to this point in Section 6.

We conclude this Section with some remarks about asymmetric cycles of the Fibonacci trace map. Without the benefits of symmetry, finding asymmetric cycles in a 3D reversible mapping in general necessitates a 3D search. Because of the invariant, Proposition 1 still holds and we expect asymmetric cycles to lie on curves, with their intersections governed by (iii). We do not expect a difference between odd- and even-cycles, as occurred in the symmetric case, because of the different dimensions of $Fix(G_1)$ and $Fix(H_1)$. Indeed, considering G_1-asymmetric cycles of F_1, it is easy to check that all the fixed points through 4-cycles of F_1 are given precisely by the G_1-symmetric cycles of Table I. The first asymmetric cycles of F_1 occur at the level of 5-cycles. One way to see this is to observe that Table I produces only one (symmetric) 5-cycle which is clearly not on $\mathcal{M}_0$, whereas from counting the 5-cycles on $\mathcal{M}_0^c$ using (28) we have $p_5 = 2$. The latter 5-cycles must be an asymmetric pair. In general, the asymmetric 5-cycle pair is generated by the two curves $(a, b_+(a), c_+(a))$ and $(a, b_-(a), c_-(a))$, where $c_\pm(a)$ are solutions of the quadratic equation

$$8ac^2 + (8a^2 - 2)\,c - 2a - 1 = 0 \tag{43}$$

and

$$b_\pm = -\tfrac{1}{2}[1 + 2(a + c_\pm)]^{-1} \tag{44}$$

Note that (43) and (44) are symmetric in a and c, reflecting the fact that the image under G_1 of a solution is also a solution.

Even though reversibility cannot be used to find asymmetric cycles, one way numerically to build up curves of asymmetric odd-cycles of F_1 is to start by first finding them on $\mathcal{M}_0^c$. This is because of the following result.

Proposition 8. All odd n-cycles of F_1 on $\mathcal{M}_0^c$ for $n \geqslant 5$ are asymmetric with respect to any (weak) reversing symmetry of F_1. The number p_n of such n-cycles is even.

Proof. Because F_1 is OR, the eigenvalue spectrum of an odd-cycle which is symmetric with respect to any (weak) reversing symmetry of F_1 is $\{-1, 1, 1\}$ from Proposition 7. However, from Proposition 3, the eigenvalue spectrum of all cycles on $\mathcal{M}_0^c$ off the set $\mathcal{P}$ of pinches, in particular odd n-cycles with $n \geqslant 5$, is known to be different from this. Hence any such odd n-cycle is asymmetric with respect to a (weak) reversing symmetry. Now F_1 has a reversing involution preserving $\mathcal{M}_0^c$, e.g. G_1, and an involution acting on the set of n-cycles divides them into G_1-invariant pairs. Any cycle not in a pair would be necessarily invariant under G_1, contradicting the fact that it is asymmetric. ■

The great advantage of $\mathcal{M}_0^c$ is, of course, that on this surface the dynamics of F_1 is solvable and this can be exploited to find the odd-cycles. The curves of odd-cycles in which they are embedded can then be numerically followed off this surface. For example, at the level of 5-cycles on $\mathcal{M}_0^c$, we can solve Eq. (26) for $n=5$ to find that $\vartheta=(1/11, 4/11)$ is a solution in the $(+)$ case, and that $\vartheta=(4/11, 10/11)$ is a solution in the $(-)$ case. These points generate, respectively, a 5-cycle and a 10-cycle of L_{R_1} of (24), which are checked to be linked to each other by the representations R_{G_1} and R_{H_1} of (41). Taking the images of these two cycles under S of (22) produces the two asymmetric 5-cycles on $\mathcal{M}_0^c$.

With respect to the second statement of the above Proposition, one can verify for F_1 that p_n is even when n is odd using the formula (28). From a number-theoretic point of view, Proposition 8 says that when n is odd, the right-hand side of (28) is divisible by two. We will see this more generally in Section 6. If we consider asymmetry with respect to G_1, we can say even more about asymmetric odd n-cycles:

Proposition 9. All odd n-cycles of F_1 on $\mathcal{M}_\mu$ for $\mu>0$ are G_1-asymmetric.

Proof. From R5–R6, an odd-cycle of F_1 is either G_1-symmetric or G_1-asymmetric. In the former case, the cycle must contain a point $\{y, y, y^2\}$ in $Fix(H_1)$. From (13), $I(y, y, y^2)=-(y^2-1)^2\leqslant 0$. The only G_1-symmetric odd-cycles of F_1 with $\mu=0$ are the fixed point and 3-cycle of the pinches (18). Otherwise, in view of Proposition 5, these cycles lie on $\mathcal{M}_\mu^c$, $-1<\mu<0$, and an odd-cycle of F_1 on $\mathcal{M}_\mu$ for $\mu>0$ is necessarily asymmetric with respect to G_1. ■

It is worth noting that, although asymmetric odd cycles typically lie on curves, most curves belonging to different asymmetric odd cycles cannot intersect. This is because F_1 is OR with an invariant and the spectrum $\{1, \lambda, -\lambda^{-1}\}$ of an odd n-cycle precludes it having an eigenvalue equal to a jth root of unity when $j>2$; cf. Proposition 1(iii). This would be true of odd n-cycles in any 3D OR volume-preserving mapping with an invariant, irrespective of it being reversible. In any case, one could study numerically period-doubling intersections of curves of asymmetric odd cycles or asymmetric even cycles (or period-tupling of asymmetric even cycles). One would typically expect such intersections where Proposition 1(iii) permits them, again irrespective of reversibility, because the intersection theory is described locally by the conjugacy to a 2D conservative mapping whose bifurcation theory, irrespective of reversibility or symmetric or asymmetric, is given by ref. 45. Also one would expect the μ values at which intersections occur to scale with the 2D conservative values. It is also worth being mindful of the fact that ref. 44 has pointed out some anomalous behavior that can occur in period-doubling and period-tupling of 2D conservative mappings that commute with a group as in Eq. (19), which could also appear in mappings like F_1.

In the next Section we give some further examples of 3D trace maps that are both reversible and have an invariant.

5. SOME GENERALIZATIONS OF THE FIBONACCI TRACE MAP

Following the above discussion of the Fibonacci trace map, we now list three classes of trace maps of increasing complexity and then show that they retain many of the structural properties found in F_1. To calculate each of the trace maps, we use the prescription at the end of Section 2. That is, we use the substitution rule ρ acting on $Sl(2, \mathbb{C})$ matrices A and B to induce the associated trace map F_ρ via

$$F_\rho: \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\mathrm{tr}(A) \\ \frac{1}{2}\mathrm{tr}(B) \\ \frac{1}{2}\mathrm{tr}(AB) \end{pmatrix} \mapsto \begin{pmatrix} \frac{1}{2}\mathrm{tr}(\rho(A)) \\ \frac{1}{2}\mathrm{tr}(\rho(B)) \\ \frac{1}{2}\mathrm{tr}(\rho(AB)) \end{pmatrix} = \begin{pmatrix} f_\rho(x, y, z) \\ g_\rho(x, y, z) \\ h_\rho(x, y, z) \end{pmatrix} \tag{45}$$

Example 1. Our first example is a class of volume-preserving, orientation-reversing trace maps derived from the substitution rule ($l \in \mathbb{N}$)

$$\rho_2: \begin{array}{l} a \to b \\ b \to b^l a \end{array} \tag{46}$$

which is invertible via $\rho_2^{-1}: a \to a^{-l}b, \quad b \to a$. The corresponding substitution matrix reads

$$R_2 = \begin{pmatrix} 0 & 1 \\ 1 & l \end{pmatrix} \tag{47}$$

This matrix has irrational eigenvalues $\lambda_\pm^{(l)} = [l \pm (l^2+4)^{1/2}]/2$ which are the so-called *metallic* means. The infinite words made from the rule (46) are again aperiodic.

The resulting trace map reads

$$F_2: \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} y \\ zU_{l-1}(y) - xU_{l-2}(y) \\ zU_l(y) - xU_{l-1}(y) \end{pmatrix} \tag{48}$$

an expression given before in ref. 27. Here, $U_l(y)$ are Chebyshev's polynomials of the second kind. They can be defined by the recurrence relation

$U_{l+1}(y) = 2yU_l(y) - U_{l-1}(y)$ together with the initial conditions $U_{-1} \equiv 0$ and $U_0 \equiv 1$. Note that they are thus well defined for all integer l, although in the context of orthogonal polynomials one only needs $l \geqslant 0$.[46] Indeed, Eq. (48) is valid as the trace map of the substitution (46) for all $l \in \mathbb{Z}$.

Example 2. Consider the substitution rule

$$\rho_3: \begin{array}{l} a \to (ba)^{m-1} b \\ b \to (ba)^m b \end{array} \tag{49}$$

It is invertible via $\rho_3^{-1}: a \to (a^{-1}b)^m a^{-1}$, $b \to (ab^{-1})^{m-1} a$ and belongs to the substitution matrix R_3,

$$R_3 = \begin{pmatrix} m-1 & m \\ m & m+1 \end{pmatrix} \tag{50}$$

with eigenvalues $\lambda_{\pm}^{(m)} = m \pm (m^2+1)^{1/2}$ and det $(R_3) = -1$. This gives rise to the following 3D volume-preserving, orientation-reversing trace map:

$$F_3: \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} yU_{m-1}(z) - xU_{m-2}(z) \\ yU_m(z) - xU_{m-1}(z) \\ 2\{yU_{m-1}(z) - xU_{m-2}(z)\}\{yU_m(z) - xU_{m-1}(z)\} \end{pmatrix} \tag{51}$$

which is of slightly more complex form than (48). Note that there is an overlap: $m = 1$ in (51) coincides with $l = 2$ in (48) after a simple change of coordinates. One therefore knows F_3, for $m = 1$, to have the same dynamical properties as F_2, for $l = 2$.

Example 3. Our third example is obtained from the substitution matrix

$$R_4 = \begin{pmatrix} 1 & 1 \\ l & l+1 \end{pmatrix} \tag{52}$$

which, unlike R_2 and R_3, is not symmetric for $l \neq 1$. The substitution underlying R_4 can be taken to be

$$\rho_4: \begin{array}{l} a \to ba \\ b \to (ba)^l b \end{array} \tag{53}$$

a rule which is also invertible. The inverse is given by $\rho_4^{-1}: a \to b^{-1}a^{l+1}$, $b \to a^{-l}b$. By the same methods as in the previous examples, we obtain the trace map

$$F_4: \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} z \\ yU_l(z) - xU_{l-1}(z) \\ yU_{l+1}(z) - xU_l(z) \end{pmatrix} \tag{54}$$

which is volume- *and* orientation-preserving.

There are many similarities of these examples with the Fibonacci trace map F_1. Let us first state the following result:

Proposition 10. The trace maps F_2 of (48), F_3 of (51), and F_4 of (54) all leave $\hat{I}(x, y, z)$ of (13) invariant. Moreover, they are all reversible with, for example, reversing symmetries H_1 of (30) for F_2 and F_3, and G_1 of (30) for F_4.

Proof. It is known that F_2 has $\hat{I}(x, y, z)$ of Eq. (13) as an invariant,[27] but it is also true of F_3 and F_4. This fact can most easily be seen from Nielsen's theorem[47,13]: Since all substitutions in question are automorphisms of the free group $\mathscr{F}_2$ of two generators, the group commutator $K(A, B) = ABA^{-1}B^{-1}$ is mapped, under ρ, to a conjugate of itself or of its inverse, i.e., $K(\rho(A), \rho(B)) = W[K(A, B)]^{\pm 1} W^{-1}$ with suitable W; see refs. 13, 10, and 49 for details. The exponent on the right-hand side coincides with the determinant of the substitution matrix.[8] The result then follows from the observation that, with $x = \frac{1}{2}\mathrm{tr}(A)$, etc., as above, $\mathrm{tr}(K(A, B)) = 4\hat{I}(x, y, z) + 2$ for matrices in $Sl(2, \mathbb{C})$, which can be shown by the Cayley–Hamilton theorem and has been well known since the last century.[4,5,10]

The relations $F_2 \circ H_1 \circ F_2 = H_1$ and $F_3 \circ H_1 \circ F_3 = H_1$ can be verified by direct computation using the recurrence relation of the Chebyshev polynomials and the identities $(U_{l-1}U_{l-2} - U_lU_{l-3})(y) = 2y$ and $U_l^2 - U_{l-1}U_{l+1} \equiv 1$, respectively. These identities can be checked inductively by means of the recurrence relations of the Chebyshev polynomials. The latter identity is also needed to verify $F_4 \circ G_1 \circ F_4 = G_1$. ∎

Since F_2, F_3, and F_4 are reversible with reversing symmetries as indicated above, they automatically have a whole family of reversing symmetries; compare property R1. In particular, they can be written as a product of involutions, e.g., $F_2 = H_1 \circ (H_1 \circ F_2)$. Note that for $l = 1$, one has $H_1 \circ F_2 = G_1$ of (30).

At this point we can expect many similarities of the dynamics of (48) with that of the Fibonacci trace map (12), the latter being contained for $l=1$. This has partly been studied by ref. 48 in a slightly different coordinate system which has, however, the disadvantage of losing the volume preservation and introducing unnecessary singularities. In view of the results of the Appendix and the comments about the Fibonacci trace map, we expect period-doubling and -tupling cascades in (48) to be typical, culminating in a pseudo-Anosov system on the compact part $\mathcal{M}_0^c$ of the invariant surface for $\mu=0$. There, the dynamics is again semiconjugate to a hyperbolic toral automorphism given by the matrix R_2 of Eq. (47) and has the Lyapunov exponents $\pm\log(\lambda_+^{(l)})$.

The trace map F_3 has fixed points $(0, 0, 0)$, $(1, 1, 1)$, and $(-1, -1, 1)$. For m even, also $(1, -1, -1)$ and $(-1, 1, -1)$ are fixed points, while the latter form a symmetric 2-cycle for m odd. Furthermore, $(0, 0, z)$ gives a one-parameter *family* of symmetric 2-cycles which intersects with the fixed point $(0, 0, 0)$. Here again, one can expect a period doubling cascade, and also the invariant set $\mathcal{M}_0^c$ plays a similar role as above: the motion on it is chaotic and leads again to a pseudo-Anosov system, by the semiconjugacy to a hyperbolic toral automorphism described by the substitution matrix R_3 of (50). The Lyapunov exponents are again given by $\pm\log(\lambda_+^{(m)})$. Completely analogous statements can be formulated for F_4, but we will not go into further details here.

It is obvious that a more systematic structure is behind the common features of the above examples and we will now turn to its description.

6. GENERATORS FOR VOLUME-PRESERVING TRACE MAPS AND REVERSIBILITY

In this Section we explore the prevalence of reversibility and other dynamical features in trace maps, as suggested by the above examples. We first do this in a constructive way by using the fact that trace maps belonging to a certain class of substitution rule, namely invertible substitution rules, can be built up from simple generator trace maps which are easily studied. Then, using the full power of the algebraic structure of such trace maps, we show how reversibility can be deduced directly from the associated substitution matrix.

The importance of the theory of free groups to the understanding of 1D nonperiodic chains obtained by substitution rules is well known.[10,17,51,54] More specifically, as mentioned in Section 2, the set of all finite words $w(a, b, a^{-1}, b^{-1})$ constitutes the free group $\mathcal{F}_2$ generated by the two-letter alphabet $\{a, b\}$. Substitution rules are homomorphisms from $\mathcal{F}_2$ to itself. Invertible homomorphisms are *automorphisms* of $\mathcal{F}_2$, and they

Table II. Generators U, P, S for the Group Φ_2 of Automorphisms ρ of the Free Group $\mathscr{F}_2$, with Their Corresponding Substitution Matrices and Trace Maps

ρ	R_ρ	F_ρ
U: $a \to ab$, $b \to b$	$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \to \begin{pmatrix} z \\ y \\ 2yz - x \end{pmatrix}$
P: $a \to b$, $b \to a$	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \to \begin{pmatrix} y \\ x \\ z \end{pmatrix}$
S: $a \to a^{-1}$, $b \to b$	$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \to \begin{pmatrix} x \\ y \\ 2xy - z \end{pmatrix}$

form a group Φ_2 under the product (4). Invertible substitution rules like (7), (46), (49), and (53) above are elements of Φ_2. Nielsen showed that the group Φ_2 is *finitely* generated by three automorphisms labeled U, P, and S.[(13),6] In Table II we list these generating automorphisms together with their corresponding substitution matrices, and trace maps written with respect to the standard basis $(x, y, z) = (\frac{1}{2}\text{tr}(A), \frac{1}{2}\text{tr}(B), \frac{1}{2}\text{tr}(AB))$, where A and B are 2×2 unimodular matrices identified with the letters a and b.

The existence of the three generators means that arbitrary invertible substitution rules can be written as words $W(U, P, S)$ in U, P, and S. For example, in Table III, we do this for the rules (7), (46), (49), and (53). Via (9), trace maps corresponding to invertible substitution rules, which we call *Nielsen trace maps*, can be written as words in F_U, F_P, and F_S, and their substitution matrices can be written as a product of a string of the matrices R_U, R_P, and R_S. In fact, from (5) with (9), it follows that the inverse of a Nielsen trace map corresponding to an invertible substitution rule exists and is a Nielsen trace map, and that the corresponding (integer) substitution matrix has an inverse which is also an (integer) substitution matrix. Consequently, the mapping from the invertible substitution rules to their trace maps is a homomorphism ψ (from Φ_2 to the group of all trace maps of the Nielsen class, which we denote by $\mathscr{G}$), as is the mapping φ from the substitution rules to their substitution matrices [from Φ_2 to the group $Gl(2, \mathbb{Z})$]. Because the latter mapping is onto, it follows that $Gl(2, \mathbb{Z})$

[6] Note that the generator σ of ref. 13 is S in our notation.

Table III. Some Invertible Substitution Rules and Their Representations As Words $W(U, P, S)$ in the Nielsen Generators U, P, S of Table II[a]

ρ	$W(U, P, S)$
$\rho_1: a \to b, b \to ba$	PUT
$\rho_2: a \to b, b \to b^l a$	$P(UT)^l$
$\rho_3: a \to (ba)^{m-1} b, b \to (ba)^m b$	$USP(UT)^m PUT$
$\rho_4: a \to ba, b \to (ba)^l b$	$P(UT)^l PUT$

[a] For brevity, we use $T = (SU^{-1})^2$: $a \to bab^{-1}$, $b \to b$.

can be generated from R_U, R_P, and R_S of Table II. The situation is summarized by the following diagram:

$$\begin{tikzcd} & \rho \in \Phi_2 \arrow[dl, "\varphi"'] \arrow[dr, "\psi"] & \\ R_\rho \in Gl(2, \mathbb{Z}) \arrow[rr, "\chi"] & & F_\rho \in \mathcal{G} \simeq PGl(2, \mathbb{Z}) \end{tikzcd} \tag{55}$$

Note that one can also find a homomorphism χ from $Gl(2, \mathbb{Z})$ to the group of Nielsen trace maps $\mathcal{G} = \langle\langle F_U, F_P, F_S \rangle\rangle$ mapping R_U, R_P, R_S onto F_U, F_P, F_S, respectively. Indeed this homomorphism χ has kernel $\{\pm \mathbb{1}\}$, as can easily be checked. Consequently, the group $\mathcal{G}$ of Nielsen trace maps is isomorphic with $PGl(2, \mathbb{Z})$ as also indicated in the diagram (55).[52,51]

Some general properties of the group $\mathcal{G}$ of Nielsen trace maps can now be deduced by looking at properties of the component generator trace maps (cf. also ref. 54 for a similar approach, where slightly different generators are used). The following properties for F_U, F_P, and F_S of Table II are easily verified:

Proposition 11. The generator trace maps F_U, F_P, and F_S of Table II satisfy:

1. All are volume-preserving: $\det dF_U = 1$, $\det dF_P = -1$, and $\det dF_S = -1$.
2. All preserve the Fricke character $\hat{I}(x, y, z)$ of Eq. (13).
3. All fix the points $(0, 0, 0)$ and $(1, 1, 1)$.

4. All commute with the symmetry group Σ of Eq. (20), specifically

$$\begin{aligned} F_U\circ\sigma_1&=\sigma_3\circ F_U, \quad F_U\circ\sigma_2=\sigma_2\circ F_U, \quad F_U\circ\sigma_3=\sigma_1\circ F_U \\ F_P\circ\sigma_1&=\sigma_2\circ F_P, \quad F_P\circ\sigma_2=\sigma_1\circ F_P, \quad F_P\circ\sigma_3=\sigma_3\circ F_P \end{aligned} \tag{56}$$

   and F_S commutes with each σ_i.

5. The generators F_P and F_S are involutions and commute. Their product is

$$F_P\circ F_S=F_S\circ F_P=H_1 \tag{57}$$

   where H_1 is the involution in Eq. (30). The generators F_U and F_S satisfy

$$F_S\circ F_U=G_1 \tag{58}$$

   where G_1 is the involution in Eq. (30).

Because every Nielsen trace map is expressible as a trace map word in F_U, F_P, and F_S (and, of course, their inverses) via the homomorphism (9), properties common to the generators that are retained under functional composition will hold for all trace maps in the Nielsen class. We find the following result:

Proposition 12. Consider any trace map F that is a composition of F_U, F_P, and F_S.

1. It is a polynomial diffeomorphism of $\mathbb{R}^3$ to itself with integer coefficients.
2. It is volume-preserving and its linearization has determinant ± 1 which coincides with the determinant of the corresponding substitution matrix $R\in Gl(2,\mathbb{Z})$.
3. It preserves the Fricke character $\hat{I}(x, y, z)$ of Eq. (13).
4. It has fixed points $(0, 0, 0)$ and $(1, 1, 1)$.
5. It keeps invariant the set $\hat{\mathscr{P}}$ of Eq. (18).
6. It commutes with the symmetry group Σ of Eq. (20).

It should be stressed that statements 1–5 of Proposition 12 are certainly known already in the literature.[5,11,6,17,54,51,10,8] We list them here as a summary of properties of Nielsen trace maps, and to indicate the power of the generator approach. On the other hand, we are not aware of previous consideration of the symmetries and reversing symmetries of Nielsen trace maps, as expressed by Proposition 11 (4, 5) and Proposition 12 (6), which will be a key point in this Section.

Prior to this, however, let us point out that the preservation of the Fricke character $\hat{I}$ by all Nielsen trace maps has some important consequences. Proposition 12 (3) means that the sets $\mathcal{M}_\mu$ defined by Eq. (14), and depicted in Fig. 1 are invariant sets for *any* Nielsen trace map F. Different trace maps in this class will induce different dynamics on the family of 2D invariant surfaces, but certain dynamical features will be common to all. For instance, by elementary topological arguments, the "spherical" part of the surface in Fig. 1b will be separately invariant under any F, and the four cones will be mapped between themselves. Other dynamical features noted in the Fibonacci trace map that extend to the entire Nielsen class are described in the next two Propositions.

Proposition 13. Apart from the fixed points $(0, 0, 0)$ and $(1, 1, 1)$ and the elements of the invariant set $\hat{\mathcal{P}}$ given by (18), which are all points at which $\nabla \hat{I}$ vanishes, all n-cycles of a Nielsen trace map F have the properties (i)–(iii) of Proposition 1. The bifurcation conditions for curves of one period to intersect curves of another period are described by the bifurcation theory of OP or OR area-preserving mappings. In a cascade of such bifurcations, the sequence of μ-values of the level sets $\mathcal{M}_\mu$ where intersections occur scales according to the universality class of area-preserving mappings.

This Proposition follows directly from the results of Proposition 1 and the Appendix. These show a local conjugacy between F and a one-parameter 2D measure-preserving mapping, so that intersections of curves of periodic orbits are expected precisely where Proposition 1 allows them to happen.[45] As we did with the Fibonacci map [cf. (39)], the local conjugacy can be extended to a small neighborhood of a cascade of intersections lying wholly on upper or lower branches of the invariant. This may require taking a power of F and/or restricting to high-period cycles. As period-doubling and n-tupling become spatially-localized processes if this is done, one always expects to be able to confine to one set of branches. Note that this conjugacy is possible because of the invariance of $\hat{I}$ and is independent of reversibility. Also, as mentioned at the end of Section 4, the presence of a commuting group for a Nielsen trace map may give rise to some anomalous bifurcation behavior.[44]

As with the Fibonacci map and the other examples above, we can also characterize the motion on $\mathcal{M}_0$.

Proposition 14. Let F_ρ be a Nielsen trace map corresponding to the invertible substitution rule ρ with associated substitution matrix $R_\rho \in Gl(2, \mathbb{Z})$. Then on $\mathcal{M}_0$ we have

$$F_\rho|_{\mathcal{M}_0} \circ S = S \circ R_\rho \tag{59}$$

where the parametrization S is given by (22) on the compact part $\mathcal{M}_0^c$, and on $\mathcal{M}_0^{nc}$ by S with $\cos(2\pi\vartheta)$ replaced by $\cosh(\vartheta)$.

Assume R_ρ is hyperbolic. Then the motion on $\mathcal{M}_0^c$ is that of a pseudo-Anosov mapping, the number of fixed points of F_ρ^n on $\mathcal{M}_0^c$ is $|\mathrm{tr}(R_\rho^n)|$, there are periodic orbits of every period, and the eigenvalue spectrum of an n-cycle off the pinches is $\{1, \pm\lambda_\rho^n, \pm\lambda_\rho^{-n}\}(\{1, \pm\lambda_\rho^n, \mp\lambda_\rho^{-n}\})$, where λ_ρ is an eigenvalue of R_ρ, and F_ρ is OP or OR with n even (OR with n odd).

The motion on $\mathcal{M}_0^{nc}$, the noncompact cones $\mathcal{C}_0^{(\varepsilon_x, \varepsilon_y, \varepsilon_z)}$, $\varepsilon_x\varepsilon_y\varepsilon_z = 1$, is unbounded in forward and backward time, except for the images under the parametrization on the cones of the stable and unstable manifolds of the linear hyperbolic mapping generated by R_ρ.

If $R_\rho \in Gl(2, \mathbb{Z})$, $R_\rho \neq \pm\mathbb{1}$, is not hyperbolic and does not have $\mathrm{tr}(R_\rho) = \pm 2$ with $\det(R_\rho) = 1$, then F_ρ is of finite order, i.e., $F_\rho^k = Id$ with $k = 1, 2$, or 3.

Proof. Proposition 14 rests upon showing the semiconjugacy (59). This result follows immediately from the fact that, without loss of generality, on $\mathcal{M}_0$ we can consider the substitution rule ρ to act on diagonal 2×2 unimodular matrices A and B (cf. the Appendix in ref. 8 and ref. 56). Given (59), it follows when R_ρ is hyperbolic that the dynamics of F_ρ on $\mathcal{M}_0^c$ is equivalent to the quotient of the hyperbolic toral automorphism L_{R_ρ} induced by R_ρ, by the identification of $(\vartheta_1, \vartheta_2)$ and $(-\vartheta_1, -\vartheta_2)$. Hence F acts as a pseudo-Anosov mapping and the results of ref. 28 can be applied, exactly analogous to the proof of Proposition 3. The dynamics of F_ρ on $\mathcal{M}_0^{nc}$ is deduced in a similar way to the proof of Proposition 6, i.e., determining it on $\mathcal{C}_0^{(+++)}$ and then using the knowledge of the way F_ρ commutes with Σ. Finally, R_ρ nonhyperbolic and not of the type $\mathrm{tr}(R_\rho) = \pm 2$ with $\det(R_\rho) = 1$ means R_ρ is traceless with $\det(R_\rho) = \mp 1$ (whence of order 2 or 4) or $\mathrm{tr}(R_\rho) = \pm 1$ with $\det(R_\rho) = 1$ (whence of order 3 or 6). The only other elements of finite order are $\pm\mathbb{1}$. Via the homomorphism χ of (55) with kernel $\{\pm\mathbb{1}\}$, we find the corresponding Nielsen trace maps are of finite order $\{1, 2, 3\}$, and no other finite orders are possible. ∎

We turn now to the action of a Nielsen trace map on the symmetry group Σ of (20). In addition to Proposition 12 (6), we see even from the previous Proposition that it is desirable to determine this action explicitly. As discussed above and below Proposition 2, the action of F_ρ on Σ will induce a permutation $\pi \in S_3$ on the elements $\{\sigma_1, \sigma_2, \sigma_3\}$. In turn, this will have the dynamical effect of linking periodic orbits of F_ρ and their corresponding eigenvalue spectra. Which permutation is induced can in fact be found from the corresponding substitution matrix (cf. also ref. 42):

Proposition 15. A Nielsen trace map F_ρ with substitution matrix $R_\rho = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ commutes with σ_1 of (20) if and only if b even and d odd, with σ_2 if and only if a odd and c even, and with σ_3 if and only if both $(a+b)$ and $(c+d)$ odd.

Proof. Nielsen trace maps and all transformations in Σ preserve the level sets of $\hat{I}(x, y, z)$, in particular, $\mathcal{M}_0^c$. On this surface, we have the parametrization S of (22). Hence we get the semiconjugacies (59) for F_ρ and (40), (42) for the σ's which provide simple 2D representations of these mappings. Now, $F_\rho \circ \sigma_1 = \sigma_1 \circ F_\rho$ on $\mathcal{M}_0^c$ if and only if $S \circ (R_\rho \circ f_{\sigma_1}) = S \circ (f_{\sigma_1} \circ R_\rho)$, which gives b even and d odd, etc. Finally, as we know from the generators for the Nielsen trace maps, the commutation property extends from $\mathcal{M}_0^c$ to the entire space. ■

Of course, the way that F_ρ permutes the elements of $\hat{\mathcal{P}}$ of Eq. (18), and the cones of $\mathcal{M}_0^{nc}$ attached to them, is a reflection of the way it permutes the elements of the group Σ. The action of F on $\hat{\mathcal{P}}$ is also studied in ref. 56 with an implicitly equivalent result to Proposition 15, but the conclusions for the existence of global symmetries are not drawn.

We now consider the reversibility of arbitrary Nielsen trace maps. Following the study of other properties above, we start by considering the reversibility of the generator trace maps F_U, F_P and F_S. This also provides a partial characterization of trace map words. For brevity from this point onward it is convenient to introduce an easier notation for the generator trace maps via

$$F_U \to u, \qquad F_P \to p, \qquad F_S \to s \tag{60}$$

We then consider trace map words $w(u, p, s)$ which are strings in the letters u, p, s and their (integral) powers, where the letters are identified with the generator trace maps of Table II via (60). Juxtaposition of letters in a word corresponds to functional composition of the trace maps, so that we drop the "$\circ$" symbol.

Proposition . Dividing trace map words $w(u, p, s)$ into one-, two-, and three-letter words, then:

1. All one-letter words are reversible with reversing symmetry s. In particular, $w(p) = p$ or Id, $w(s) = s$ or Id, and the word $w(u) = u^n$, $n \in \mathbb{Z}$, satisfies

$$u^n s = s u^{-n} \tag{61}$$

2. A two-letter word $w(p, s)$ is equal to one of the involutions p, s, or ps and so is reversible. A two-letter word $w(u, s)$ is equal to an involution of the form su^n, $n \in \mathbb{Z}$, or to some power of u, and so is reversible.

Proof. For the one-letter words, s is an involution and any involution is trivially reversible with respect to itself. Then use the fact that $p = (ps)s$ and $u = s(su)$ and that ps and su are also involutions from (57) and (58), respectively. Because u is reversible with reversing symmetry s, so is u^n, which gives Eq. (61). For the two-letter words $w(p, s)$ use the fact that p and s commute and are involutions to separate the two types of letters within the word and reduce their number. Similarly, for $w(u, s)$, use (61) to separate the two types of letters, and then reduce the string of s's to s or e, the identity letter. ■

The trace map words not considered in Proposition 16 are the two-letter words $w(u, p)$ and the general three-letter words $w(u, p, s)$. These remaining possibilities are closely related because $w(u, p, s)$ can always be reduced to a two-letter word $w(u, p)$ or to its composition with s by means of (61) and the commutativity of p and s. Even without the use of reversibility as in the above proposition, s could always be eliminated from a trace map word $w(u, p, s)$ because it is expressible in terms of the generators u and p via

$$s = upu^{-1}pup \tag{62}$$

That is, the group $\mathscr{G}$ of Nielsen trace maps is generated by u and p alone, as is $PGl(2, \mathbb{Z})$ by the equivalence classes of the two matrices R_U and R_P of Table II; compare ref. 52. Nevertheless, using reversibility relations between the generators to pull the s's together in the word and annihilate them in pairs leads to simpler words than using (62) to substitute for every occurrence of s. We think it is worthwhile for historical and practical reasons to continue considering the set of three generators of $\mathscr{G}$.

Proposition 16 then highlights the need to consider the reversibility of the two-letter words $w(u, p)$ and the three-letter words formed by composing them with an s. This is nontrivial even given the reversibility of the component generator trace maps p and u. Most of Proposition 12 followed by just noting which properties of the generators were retained under composition. For reversibility, however, the situation is more complicated, as the following additional property of reversible mappings indicates:

R8 The composition of two reversible mappings L and M which share a reversing symmetry G is a reversible mapping, but typically the composition $L \circ M$ or $M \circ L$ does not have G as a reversing symmetry. The latter is true if and only if L and M commute.

Because p and u do not commute, a word $w(u, p)$ is not reversible with respect to s even, though p and u are. Despite this, a word $w(u, p)$ may still of course be reversible with a reversing symmetry different from s. For example, although it is somewhat *ad hoc*, we can systematically consider the reversibility of successively more complicated short words $w(u, p)$ and $w(u, p)\,s$. For instance, the following words $w(u, p)$ that start with p are the products of two bracketted involutions as indicated:

$$\begin{aligned} pu^j &= (ps)(su^j) \\ pu^jp &= (pu^jps)(s) \\ pu^jpu^k &= (pu^jps)(su^k) \\ pu^jpu^kp &= (pu^jpu^jps)(psu^{k-j}p) \\ pu^jpu^kpu^lp &= (pu^jpu^kpu^jps)(pu^{j-l}sp) \end{aligned} \tag{63}$$

The decomposition into involutions of these words is aided by the fact that if L has reversing symmetry G, then $L^i \circ G = G \circ L^{-i}$ is also a reversing symmetry (cf. R1, Section 4). This allows longer and longer involution words to be constructed, e.g., the reversibility of u with respect to s shows that su^j is an involution [cf. (61)], and the reversibility of pu^j with respect to ps shows that pu^jps, pu^jpu^jps, and $pspu^jpu^j$ are also involutions.

The reversibility of the first few words $w(u, p)$ starting with u^j follows by "sandwiching" each of the above words beteen p and $p^{-1} = p$ and using property R3 of Section 4. Similarly, some short reversible trace map words of the form $w(u, p)\,s$ include

$$\begin{aligned} pu^js &= (p)(u^js) \\ pu^jps &= (pu^jps) \\ pu^jpu^kps &= (pu^jpu^{-j}p)(pu^{j+k}ps) \end{aligned} \tag{64}$$

Already missing from the natural progression of the words in the two short lists above are words $w(u, p)$ like

$$w_1(u, p) = pu^jpu^kpu^l \tag{65}$$

and words $w(u, p)\,s$ like

$$w_1(u, p, s) = pu^jpu^ks \tag{66}$$

whose reversibility appears harder to decide for completely arbitrary powers of u. Some special cases can be seen to be reversible, e.g. for the first word,

$$\begin{aligned} pu^j pu^j pu^l &= (pu^j pu^j ps)(su^l) \\ pu^j pu^k pu^k &= (ps)(su^j pu^k pu^k) \\ pu^{-1} pu^j pu &= u^{-1}(psu^{j-1}p)\, u \end{aligned} \tag{67}$$

and for the second word,

$$pu^j pu^j s = (pu^j pu^j ps)(p) \tag{68}$$

Note that in the last case of (67), we use $pu^{-1}p = u^{-1}psu^{-1}$, which follows from Eq. (62).

Nevertheless, the above constructive approach recognized many reversible trace maps and involutory trace maps which act as reversing symmetries for them. In particular, it allows us to see by inspection why the substitution rules (7), (46), (49), and (53) considered above led to reversible trace maps with the reversing symmetries previously noted. In Table IV, we give the trace map words $w(u, p, s)$ corresponding to these rules, which follow from their words $W(U, P, S)$ in Table III, the homomorphism (9), and the simplifications that can be made to condense the trace map word using Proposition 16. Noting (57) and (58), we understand the prevalence of reversibility in these examples with respect to G_1 or H_1. Also listed in Table IV is the action of each example on the symmetry group Σ.

Table IV. Trace Map Words $w(u, s)$ for the Trace Maps Corresponding to the Substitution Rules of Table III, with Their Transformation under the Group Σ of Proposition 2

F_p	$w(u, p, s)$	$F_p \circ \Sigma$
F_1	pu	$F_1 \circ \sigma_i = \sigma_{i-1} \circ F_1$
F_2	pu^l	l odd: same as F_1 l even: same as p
F_3	$su^{-1}pu^m pu$	m odd: same as p m even: same as s
F_4	$pu^l pu$	l odd: same as F_1^{-1} l even: same as u

A more systematic approach can be taken to asking what trace maps in the Nielsen class are reversible with respect to reversing symmetries which are themselves Nielsen trace maps. Because of the homomorphism χ in (55), we can link a trace map F directly with a substitution matrix, now denoted R_F.

Proposition 17. Let F be a Nielsen trace map and let $R_F = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in Gl(2, \mathbb{Z})$ be the corresponding substitution matrix. Then F is reversible with a reversing symmetry that is also a Nielsen trace map if and only if

(1) $a + d = 0$ (in which case $F^2 = Id$) or

(2) there is a solution of $\alpha(a-d) + \beta c + \gamma b = 0$, $\alpha, \beta, \gamma \in \mathbb{Z}$, with $\alpha^2 + \beta\gamma = \pm 1$.

Proof. The proof is straightforward if one observes that $\{\pm\mathbb{1}\}$ is the kernel of the homomorphism χ from the matrix group $Gl(2, \mathbb{Z})$ to the group $\mathcal{G}$ of trace maps of the Nielsen class; cf. (55). Then the condition $F \circ G \circ F = G$ can be calculated with 2×2 matrices modulo the kernel, i.e., we have to solve $R_F \cdot R_G \cdot R_F = \pm R_G$ with R_F, $R_G \in Gl(2, \mathbb{Z})$ and $R_G^2 = \pm\mathbb{1}$ [equivalently, solve reversibility of matrices in $PGl(2, \mathbb{Z})$]. To do so, we observe that—in $Gl(2, \mathbb{Z})$—nontrivial involutions (i.e., $R_G^2 = \mathbb{1}$, but $R_G \neq \pm\mathbb{1}$) are only possible for $\Delta = \det(R_G) = -1$ and are of the form $\left(\begin{smallmatrix} \alpha & \beta \\ \gamma & -\alpha \end{smallmatrix}\right)$ with $\alpha^2 + \beta\gamma = 1$, while anti-involutions (i.e., $R_G^2 = -\mathbb{1}$) have the same matrix form but require $\Delta = 1$, i.e., $\alpha^2 + \beta\gamma = -1$. A direct substitution into $R_F \cdot R_G \cdot R_F = \pm R_G$ results in the statement of the proposition. Finally, the substitution matrices R_G involved have to be put into a (uniquely defined!) trace map to obtain the necessary involution for reversibility. ∎

Note that the "only if" part of the above Proposition can also be deduced by considering the dynamics induced on $\mathcal{M}_0^c$. Because Nielsen trace maps preserve this surface, a necessary condition for reversibility within the Nielsen class is reversibility on this particular surface. Consideration of the latter is made simpler using the representation (59) on $\mathcal{M}_0^c$. The reversibility relation $F \circ G \circ F = G$ for two trace maps F and G then induces the relation $S \circ (R_F \cdot R_G \cdot R_F) = S \circ R_G$, i.e., $R_F \cdot R_G \cdot R_F = \pm R_G$. We showed how this worked for the Fibonacci mapping F_1 in Section 3; cf. (40) and the surrounding discussion. From Proposition 17 we observe that this mapping is reversible with a trace map with substitution matrix $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. From (41), we see that this matrix corresponds to the representation of H_1 on $\mathcal{M}_0^c$.

Before we come back to the general case, let us state two consequences of Proposition 17.

Proposition 18. Let F be a trace map with substitution matrix $F=\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in Gl(2, \mathbb{Z})$. Then, any of the following conditions is sufficient for reversibility:

(1) $a=d$ or $a=-d$.

(2) $b=c$ or $b=-c$.

(3) b or c divides $(d-a)$.

Proof. A direct substitution shows that $R_G=\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ gives the case $a=d$. Now, $a=-d$ is case (1) of Proposition 17, while $R_G=\begin{pmatrix} 0 & \pm 1 \\ 1 & 0 \end{pmatrix}$ yields condition (2). Finally, if $b \mid (d-a)$, one can choose $\alpha=1, \beta=0$, and $\gamma=(d-a)/b$ in Proposition 17, while $c \mid (d-a)$ requires $\alpha=1$, $\beta=(d-a)/c$, and $\gamma=0$. ■

Proposition 19. Any Nielsen trace map F of finite order is reversible. Also, F is reversible if the eigenvalues of the corresponding substitution matrix R_F are rational.

Proof. Finite order in $PGl(2, \mathbb{Z})$ means order 1, 2, or 3; compare ref. 52 and Proposition 14. The statement is trivial for orders 1 and 2. Order 3, in turn, requires $R \in Gl(2, \mathbb{Z})$ with $\mathrm{tr}(R)=\pm 1$ and $\det(R)=1$. W.l.o.g., we can choose $\mathrm{tr}(R)=1$, i.e., $R=\begin{pmatrix} a & b \\ c & 1-a \end{pmatrix}$, where $bc=a(1-a)-1$. Next, we observe that the greatest common divisor (gcd) of $2a-1$ and a^2-a+1 is either 1 or 3, whence either $\gcd(2a-1, b)=1$ or $\gcd(2a-1, c)=1$. In either case, it is clear how to write down the general solution of the linear Diophantine equation in Proposition 17 (2). If one then inserts this into the remaining nonlinear Diophantine equation, one reduces the problem to the question of whether 1 is representable by an integer binary quadratic form of discriminant -3, the answer to which is always yes because the corresponding class number is one.[53]

Rational eigenvalues for R_F means $\mathrm{tr}(R_F)=0$ with $\det(R_F)=-1$ (in which case F is an involution) or $\mathrm{tr}(R_F)=\pm 2$ with $\det(R_F)=1$. The conjugacy classes [in $Gl(2, \mathbb{Z})$] of the latter case[42] are faithfully represented by the matrices $\begin{pmatrix} \pm 1 & m \\ 0 & \pm 1 \end{pmatrix}$ with $m \in \mathbb{N}_0$. The corresponding trace maps are obviously reversible by Proposition 18 (1), and reversibility is preserved under conjugacy. ■

Though the criteria of Proposition 18 seem a bit restricted, it will nevertheless turn out that—up to conjugacy within the set of Nielsen trace maps—they completely characterize reversibility; see refs. 42 and 55 for details. These simple criteria allow us to see immediately the reversibility of all our examples in Section 5, and the corresponding reversing symmetries are obtained from R_G via the generators of Table II. Furthermore, the appearance of integral quadratic forms in the proof of Proposition 19 is not at all an accident. The general case (2) of Proposition 17 can be reformulated as a representation problem of ± 1 by (generally indefinite) quadratic forms. This results in a complete decidability of the reversible cases, but a detailed description is beyond the scope of the present article and is given in ref. 42.

One example that escapes both conditions of Proposition 17 is $\begin{pmatrix} 1 & 5 \\ 3 & 14 \end{pmatrix}$. Notice that this belongs to the set of substitution matrices

$$R = \begin{pmatrix} 1 & -k \\ j & -(jk+1) \end{pmatrix}$$

corresponding to the trace map words $w(u, p, s)$ of (66), whose reversibility could not be decided in general via the constructive approach. Consequently, the trace map $pu^3pu^{-5}s = pu^3psu^5$ is not reversible in the class of Nielsen trace maps. This trace map is similar to the Fibonacci map in that it commutes with no elements of Σ, but its cube commutes with all elements—consequently, it keeps $(1, 1, 1)$ fixed and the set $\hat{\mathscr{P}}$ of (18) is a 3-cycle.

We have mentioned weak reversibility in Section 4 as a generalization of reversibility. Within the class of Nielsen trace maps, we can state the following result:

Proposition 20. Let F, G be Nielsen trace maps with $F \circ G \circ F = G$. Then $G^2 = Id$ or $F^2 = Id$, and in both cases F is reversible. In other words, weak reversibility automatically implies reversibility within the class of Nielsen trace maps.

Proof. This is again a statement about $PGl(2, \mathbb{Z})$ matrices. Starting from integer 2×2 matrices with $R_F \cdot R_G = \pm R_G \cdot R_F^{-1}$ as above, one finds either the cases of Proposition 17 or the condition $R_F^2 = \pm \mathbb{1}$. In the former case, one has $R_G^2 = \pm \mathbb{1}$ and thus $G^2 = Id$, while in the latter case, F itself is an involution and hence trivially reversible. ∎

So far, Propositions 17 and 20 restrict the search for reversing symmetries to involutions that are Nielsen trace maps themselves, i.e., are elements of $\mathscr{G}$. As such, these reversing symmetries will have all the properties detailed in Proposition 12, e.g., they preserve the Fricke invariant, $(0, 0, 0)$, $(1, 1, 1)$, and the set $\hat{\mathscr{P}}$ of (18) are fixed, and they commute with the group Σ. Nevertheless, obviously other involutions that are not themselves trace maps could act as reversing symmetries in this case. That is to say, an even larger class of mappings than those identified in Proposition 17 could be reversible.

We will now enlarge the search for reversing symmetries to the group

$\mathcal{A}$ generated by all elements of Σ and $\mathcal{G}$. This will turn out to be a very pertinent set to look at, and it has the following structure:

Proposition 21. The group Σ is a normal subgroup of $\mathcal{A}$ and we have $\mathcal{A} = \Sigma \oplus_s \mathcal{G}$, i.e., $\mathcal{A}$ is the semidirect product of its subgroups Σ and $\mathcal{G}$.

Proof. Σ normal in $\mathcal{A}$ means $a\Sigma a^{-1} = \Sigma$ for all $a \in \mathcal{A}$, which is an immediate consequence of Proposition 12 (6). On the other hand, $\sigma_i \in \Sigma$ cannot be a trace map, because it does not fix the point (1, 1, 1); cf. Proposition 12 (4). Thus $\Sigma \cap \mathcal{G} = \{Id\}$, whereby the statement follows. ∎

Let us remark that the result that Σ is a normal subgroup of $\mathcal{A}$ is just a more algebraic expression of the fact that every Nielsen trace map commutes with Σ. The action of a trace map F on Σ is given by $F \circ \sigma_i \circ F^{-1} = \sigma_{\pi_F(i)}$, where π_F is the permutation deduced using Proposition 15. This way, the structure of the semidirect product is completely known.

The set $\mathcal{A}$ turns out to be identical with the set of all polynomial mappings of $\mathbb{R}^3$ or $\mathbb{C}^3$ (with real or complex coefficients) that leave $\hat{I}(x, y, z)$ of (13) invariant.[51] Significantly, the fact that $\mathcal{A}$ is a group means that possession of this property by a polynomial mapping is sufficient to guarantee automatically that the mapping is invertible and that its inverse also preserves $\hat{I}$. The subgroup $\mathcal{G} \simeq PGl(2, \mathbb{Z})$ can be specified by fixing (1, 1, 1), i.e., $\mathcal{G} = \{a \in \mathcal{A} \mid a(1, 1, 1) = (1, 1, 1)\}$, while Σ together with the elements $(x, y, z) \mapsto (y, x, z)$ and $(x, y, z) \mapsto (z, x, y)$ from $\mathcal{G}$ generate the full tetrahedral group, T_d, with 24 elements. They are the only linear (in fact, the only affine) transformations that leave $\hat{I}$ invariant. They constitute what one would call the ordinary symmetry group of the level sets.

Considering reversibility of Nielsen trace maps with respect to reversing symmetries in $\mathcal{A}$, we note that elements $\sigma \circ g$, where $\sigma \in \Sigma$ and $g \in \mathcal{G}$, need not be involutions, because σ and g might not commute. So *a priori* we should consider weak reversibility with respect to elements of $\mathcal{A}$, but then we find:

Proposition 22. $F \in \mathcal{G}$, $F^2 \neq Id$, is weakly reversible with some element $G \in \mathcal{A}$, where $G = \sigma \circ g$ with $\sigma \in \Sigma$ and $g \in \mathcal{G}$, if and only if F is reversible with $g \in \mathcal{G}$ and commutes with σ.

Proof. The "if" part of the statement follows from the weak reversibility property that the composition of a weak reversing symmetry and a commuting map is again a weak reversing symmetry. Consider the "only if" part. Let $F \circ G = G \circ F^{-1}$ and $G = \sigma \circ g$. Then,

$$F \circ \sigma \circ g = (F \circ \sigma \circ F^{-1}) \circ (F \circ g) = \sigma \circ (g \circ F^{-1})$$

But from $\mathscr{A} = \Sigma \otimes_s \mathscr{G}$ we know that $\sigma \circ g = \sigma' \circ g'$ implies $\sigma = \sigma'$ and $g = g'$ (factorization is unique). Since $F \circ \sigma \circ F^{-1} \in \Sigma$, we have to conclude $F \circ g = g \circ F^{-1}$ and $F \circ \sigma = \sigma \circ F$. The first condition gives weak reversibility—and thus reversibility, see Proposition 20—and the second one shows that σ is a symmetry of F. ∎

This Proposition relates to our earlier observation in Section 4 that F_1^3 possesses the weak reversing symmetries $G_1 \circ \sigma_1$ and $G_1 \circ \sigma_3$, and the reversing symmetry $G_1 \circ \sigma_2$, all of which belong to $\mathscr{A}$, but that these are not (weak) reversing symmetries for F_1. We can see now that the latter would only be possible if F_1 itself commuted with σ_1, σ_2, or σ_3. Proposition 22 then makes it easy to create Nielsen trace maps F that are not (weakly) reversible with respect to $G \in \mathscr{A}$, but their second or third power is. Provided F is reversible with respect to g in the Nielsen class via Proposition 17 and does not commute with every element of Σ, then take $G = g \circ \sigma_i$, where F does not commute with σ_i.

One obvious question is how general the above results are w.r.t. failure of reversibility. From Proposition 19 we know that a nonreversible Nielsen trace map F must have a hyperbolic substitution matrix R_F. Let us now briefly discuss some properties that a more general (weak) reversing symmetry (outside $\mathscr{A}$) must possess. The most general class of transformations that are meaningful in this context are *homeomorphisms*. Hence, let us consider $G \circ F \circ G^{-1} = F^{-1}$ with homeomorphism G and hyperbolic trace map F.

From R5 of Section 4, if Γ is an invariant set of F, then so is $G\Gamma$. In particular, $\mathscr{M}_0^c$ is an invariant set which contains a dense orbit (due to the pseudo-Anosov structure). Consequently, $G\mathscr{M}_0^c$ must have the same property and must be of the same topological type. Furthermore, $(0, 0, 0)$ is an isolated fixed point of F, which is then also true of $G(0, 0, 0)$, and a similar situation is met by the pinches, $\mathscr{P}$. A couple of these considerations together with the connectivity properties of the level sets and the potential consequences of transversal deformations indicate that the following situation is, at least, typical (a proof or decision would require a more complete knowledge of the closure of possible invariant sets):

- **H1** G preserves the foliation of $\mathbb{R}^3$ by the surfaces $\{\mathscr{M}_\mu\}$. In particular, it maps the surface $\mathscr{M}_\mu$, $\mu > 0$, to $\mathscr{M}_{\mu'}$, for some $\mu' > 0$. Similarly, G can at most permute the compact sets $\mathscr{M}_\mu^c$, $-1 < \mu < 0$, and the noncompact cones belonging to $\mathscr{M}_\mu^{nc}$, $\mu < 0$.
- **H2** G leaves $\mathscr{M}_0$ invariant, and in particular $\mathscr{M}_0^c$.
- **H3** The point $(0, 0, 0)$ is fixed by G, and so is necessarily a symmetric fixed point of F. The set of four "pinches" on $\mathscr{M}_0^c$, given by $\mathscr{P}$ of (18), is also fixed by G.

Let us therefore continue our discussion restricting to homeomorphisms G that obey the three properties above. But then we find:

Proposition 23. Let F be a Nielsen trace map with (weak) reversing symmetry G. Let G be of finite order, map $\mathbb{R}^3$ into itself, and have properties H1–H3. Then $G\mathcal{M}_\mu = \mathcal{M}_\mu$ and G preserves the Fricke character $\hat{I}$ of (13). If, in addition, G is a polynomial mapping, then $G \in \mathscr{A} = \Sigma \otimes_s \mathscr{G}$.

Proof. The action of G in permuting the level sets $\mathcal{M}_\mu$ induces a 1D real mapping $f := \mu \to \mu' = f(\mu)$ via $G\mathcal{M}_\mu = \mathcal{M}_{\mu'}$. Because G is a homeomorphism, f is a homeomorphism and the invariance of $\mathcal{M}_0$ and $\mathcal{M}_{-1}$ under G further imply that $f(0) = 0$ and $f(-1) = -1$. It follows that $f: \mathbb{R} \to \mathbb{R}$ is an increasing function. Because *a priori* G may permute the cones and balls of $\mathcal{M}_\mu$, $\mu < 0$, in different ways, it is conceivable that in this range we may need to consider different 1D homeomorphisms. In any case, $G^k = Id$ in $\mathbb{R}^3$ implies $f^k = Id$ in $\mathbb{R}$, and the only possible increasing 1D homeomorphism satisfying the latter is $f \equiv Id$. Hence $G\mathcal{M}_\mu = \mathcal{M}_\mu$, $\forall \mu$, which implies $\hat{I}(G\mathbf{x}) = \hat{I}(\mathbf{x})$, $\forall \mathbf{x} \in \mathbb{R}^3$. The second statement of the proposition follows immediately because, as mentioned above, ref. 51 has shown that the set of real polynomial mappings that leave $\hat{I}$ invariant is precisely $\mathscr{A}$. ■

The previous proposition highlights the importance of the set $\mathscr{A}$ if we confine our search for reversing symmetries to polynomial mappings of finite order. We can bring together some of the above results in the following chain of implications:

$$
\begin{aligned}
&F\text{ (weakly) reversible with polynomial mapping } G\\
&\quad \text{of finite order and properties H1–H3}\\
&\quad \Rightarrow G \in \mathscr{A} \qquad \text{(Proposition 23)}\\
&\quad \Rightarrow F \text{ reversible in } \mathscr{G} \qquad \text{(Proposition 22)}\\
&\quad \Rightarrow R_F \text{ satisfies Proposition 17}
\end{aligned}
\tag{69}
$$

We see then that a Nielsen trace map whose related substitution matrix does not satisfy one of the two conditions of Proposition 17 is not (weakly) reversible with any polynomial (weak) reversing symmetry of finite order which also respects H1–H3. From above, such an example is the trace map $pu^3pu^{-5}s = pu^3psu^5$ with substitution matrix $\left(\begin{smallmatrix} 1 & 5 \\ 3 & 14 \end{smallmatrix}\right)$.

We are led to ask more generally how to decide if a given Nielsen trace map is reversible with an arbitrary nonpolynomial (weak) reversing symmetry.

The above discussion suggests a couple of practical ways to test

Nielsen trace maps for irreversibility. The first is to check the eigenvalue spectra of the linearized trace map at (0, 0, 0), (1, 1, 1), and the cycles that make up the set $\mathscr{P}$, recalling that the spectra of symmetric cycles must include $\{\pm 1\}$. The spectrum can be calculated from combining the linearizations of the generator trace maps, using the form of the trace map word. Second, the number of odd-period cycles on OR Nielsen trace maps on $\mathscr{M}_0^c$ can be explicitly found using the traces of powers of the substitution matrix via Proposition 14. We can combine this knowledge with the generalization of Proposition 8 of Section 4:

Proposition 24. If a Nielsen trace map F is OR with associated hyperbolic substitution matrix R_F and is also (weakly) reversible with (weak) reversing symmetry G which is a diffeomorphism and respects properties H1–H3, then all odd n-cycles on $\mathscr{M}_0^c$ off the "pinches" $\mathscr{P}$ are asymmetric. In the event of reversibility, where G is an involution, the number p_n of such n-cycles is even if $n \geqslant 5$.

Proof. If F is OR, we know from Proposition 14 that each odd-period orbit on $\mathscr{M}_0^c$ which does not contain one of the four "pinches" has an eigenvalue spectrum different from $\{1, 1, -1\}$, which is necessary for a symmetric odd-period orbit. Hence such an odd n-cycle is asymmetric and its image under G also lies on $\mathscr{M}_0^c$ off the "pinches" via properties H1–H3. If G is an involution, then each odd n-cycle and its image are a G-invariant pair. If $n \geqslant 5$, there are no odd n-cycles within the set of pinches, and hence the total number of odd n-cycles on $\mathscr{M}_0^c$ is the sum of such pairs. ■

Note that there is an explicit expression for p_n above, analogous to that for the Fibonacci trace map in (28), namely

$$p_n = \frac{1}{n} \sum_{k \mid m} \mu\left(\frac{n}{k}\right) |\operatorname{tr}(R_F^k)| \tag{70}$$

The above Proposition is then a statement of the divisibility of this expression by 2 when $n \geqslant 5$ in the case of an OR reversible trace map. It turns out that, as a test for reversibility in this case, this divisibility property is vacuous because it always holds.[57] Similarly, using the above implications on the linearizations of (0, 0, 0) and the pinches to find possible irreversible candidates among trace maps also appears to be a "nontest."

The significance of the first part of Proposition 23 is that it shows, for reversibility and weak reversibility with G of finite order and with properties H1–H3, that one has a 2D (weakly) reversible mapping induced on each level set of $\hat{I}$ because G preserves each such set. Consequently, one can test for irreversibility of the trace map on any of these surfaces. Testing 2D mappings for (weak) reversibility and identifying nonreversible mappings has been discussed in refs. 12 and 58. The problem is particularly subtle if the mapping is OP and area-preserving. However, if the mapping is OR, the problem is much easier because possible symmetric periodic orbits are atypical, and then possible asymmetric periodic orbits would need to have reciprocal eigenvalues. We see from above that there are many OR Nielsen trace maps. It seems that, within this class, examples can be found that are not reversible with respect to any (not necessarily polynomial) G.[42] Numerically, it is useful to exploit again the solvability of the dynamics on $\mathcal{M}_0^c$ and to "track" periodic orbits off this surface.

7. GENERALIZATIONS IN HIGHER DIMENSIONS

Having presented in some detail various 3D trace maps that are both reversible and possess an invariant, we will now consider some higher-dimensional generalizations. We stress that these generalizations are chosen from a dynamical point of view, incorporating the two features of an invariant and of reversibility which we have concentrated on above. Our generalizations are not necessarily related to the recent pursuit of higher-dimensional trace maps; see ref. 59 and references within. Also, trace maps of n-letter substitution rules for $n > 2$ do not seem to have invariants analogous to the Fricke invariant $\hat{I}(x, y, z)$.[60]

Considering the mapping

$$F\colon \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} f(x, y) - z \\ x \\ y \end{pmatrix} \tag{71}$$

one finds that F is reversible with involution G, $G(x, y, z) = (z, y, x)$, if and only if $f(x, y)$ is symmetric, i.e., if and only if

$$f(y, x) = f(x, y) \tag{72}$$

This is a rather large class of reversible mappings and contains the Fibonacci trace map of (12) via $f(x, y) = 2xy$. (We have interchanged the roles of x and z hare, which obviously does not matter.) There we also had an invariant, namely $\hat{I}(x, y, z) = x^2 + y^2 + z^2 - 2xyz - 1$. But this is not the only such example in the class defined by Eq. (71), because also

$$f(x, y) = a + b \cdot (x + y) + c \cdot xy \tag{73}$$

leads to a reversible mapping (with the same G *nota bene*), and one can directly verify the invariance of the following expression[7]:

$$I(x, y, z) = x^2 + y^2 + z^2 - a \cdot (x + y + z) - b \cdot (xy + yz + xz) - c \cdot xyz \tag{74}$$

Note that we have dropped now the constant contribution which was kept in $\hat{I}$ above for the sake of compatibility with other articles.

A closer look shows that one can easily rewrite the expressions (73) and (74) with elementary symmetric polynomials. For m variables, they are defined as

$$\sigma_l^{(m)}(x_1,..., x_m) = \frac{1}{l!\,(m-l)!} \sum_{\pi \in S_m} \left(\prod_{k=1}^{l} x_{\pi(k)} \right) \tag{75}$$

where S_m denotes the symmetric or permutation group. The polynomial $\sigma_l^{(m)}$ is homogeneous of degree l. Explicitly, we have $\sigma_0^{(m)} \equiv 1$, $\sigma_1^{(m)} = x_1 + \cdots + x_m$, $\sigma_2^{(m)} = x_1 x_2 + x_1 x_3 + \cdots + x_{m-1} x_m$, etc., up to $\sigma_m^{(m)} = x_1 \cdot ... \cdot x_m$; for details we refer to ref. 61.

Let us now consider the following generalization of Eq. (71) defined by a mapping F: $\mathbb{R}^n \to \mathbb{R}^n$, $n \geqslant 2$, via

$$F: \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \mapsto \begin{pmatrix} f(x_1,..., x_{n-1}) - x_n \\ x_1 \\ \vdots \\ x_{n-1} \end{pmatrix} \tag{76}$$

where f: $\mathbb{R}^{n-1} \to \mathbb{R}$ is a scalar function of $n-1$ variables. Note that F is volume-preserving, and orientation-preserving (-reversing) if n is even (odd). That mappings with invariants exist in this class can directly be seen from F with $f \equiv 0$, which leaves $I(x_1,..., x_n) = x_1^2 + \cdots + x_n^2 = |\mathbf{x}|^2$ invariant. If we extend the involution considered for (71) to

$$G: \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \mapsto \begin{pmatrix} x_n \\ x_{n-1} \\ \vdots \\ x_1 \end{pmatrix} \tag{77}$$

we can show the following:

[7] This observation is due to G. R. W. Quispel.

Proposition 25. The mapping F of (76) is reversible with involution G of (77) if and only if $f(x_1, x_2,..., x_{n-1}) = f(x_{n-1}, x_{n-2},..., x_1)$. The subclass of such reversible mappings with

$$f(x_1,..., x_{n-1}) = \sigma_l^{(n-1)}(x_1,..., x_{n-1}) \tag{78}$$

and $0 \leqslant l \leqslant n-1$, also has an invariant given by

$$I(x_1,..., x_n) = |\mathbf{x}|^2 - \sigma_{l+1}^{(n)}(x_1,..., x_n) \tag{79}$$

Proof. One shows the reversibility by direct computation of $F \circ G \circ F = G$. For a proof of the invariance of (79) by the subclass defined through (78), we need the following identity $(0 \leqslant l \leqslant n-1)$:

$$\sigma_{l+1}^{(n)}(y_1,..., y_n) = y_1 \cdot \sigma_l^{(n-1)}(y_2,..., y_n) + \sigma_{l+1}^{(n-1)}(y_2,..., y_n) \tag{80}$$

which follows immediately from the definition of the elementary symmetric polynomials. We then obtain, with $|\mathbf{x}|^2 = x_1^2 + \cdots + x_n^2$,

$$\begin{aligned}
&I(F(x_1,..., x_n)) \\
&\quad = \{\sigma_l^{(n-1)}(x_1,..., x_{n-1})\}^2 - 2x_n \cdot \sigma_l^{(n-1)}(x_1,..., x_{n-1}) + |\mathbf{x}|^2 \\
&\qquad - \sigma_{l+1}^{(n)}(\sigma_l^{(n-1)}(x_1,..., x_{n-1}) - x_n, x_1, x_2,..., x_{n-1}) \\
&\quad = \{\sigma_l^{(n-1)}(x_1,..., x_{n-1})\}^2 - 2x_n \cdot \sigma_l^{(n-1)}(x_1,..., x_{n-1}) + |\mathbf{x}|^2 \\
&\qquad - \{\sigma_l^{(n-1)}(x_1,..., x_{n-1}) - x_n\} \cdot \sigma_l^{(n-1)}(x_1,..., x_{n-1}) \\
&\qquad - \sigma_{l+1}^{(n-1)}(x_1,..., x_{n-1}) \\
&\quad = |\mathbf{x}|^2 - x_n \cdot \sigma_l^{(n-1)}(x_1,..., x_{n-1}) - \sigma_{l+1}^{(n-1)}(x_1,..., x_{n-1}) \\
&\quad = |\mathbf{x}|^2 - \sigma_{l+1}^{(n)}(x_1,..., x_n)
\end{aligned}$$

This proves the invariance of I in (79) under F. ■

Now, looking back to Eq. (74), one could suspect that a linear combination of *different* elementary symmetric polynomials is also possible, and if we define

$$f(x_1,..., x_{n-1}) = \sum_{l=0}^{n-1} \alpha_l \cdot \sigma_l^{(n-1)}(x_1,..., x_{n-1}) \tag{81}$$

where the α_l are arbitrary constants, the corresponding F is reversible according to Proposition 25 and still possesses an invariant, namely

$$I(x_1,..., x_n) = |\mathbf{x}|^2 - \sum_{l=0}^{n-1} \alpha_l \cdot \sigma_{l+1}^{(n)}(x_1,..., x_n) \tag{82}$$

The proof can again be given through a direct calculation as above, and we omit it here. We summarize our findings in the following result.

Proposition 26. The mapping $F: \mathbb{R}^n \to \mathbb{R}^n$, $n \geqslant 2$, defined by Eq. (76) with f chosen according to Eq. (81), is reversible with G of (77) and has an invariant given by Eq. (82).

For any scalar function of $n-1$ variables, we can—via symmetrization w.r.t. the permutation

$$\pi = \begin{pmatrix} 1 & 2 & \cdots & n-1 \\ n-1 & n-2 & \cdots & 1 \end{pmatrix}$$

—construct a reversible mapping according to Proposition 25, but we will in general not find an invariant. Of course, the class of reversible mappings which also have an invariant may be larger than that indicated by Proposition 26.

It is worth noting that many of the properties we identified for 3D mappings with an invariant extend trivially to n dimensions—in particular, the result of the Appendix, which gives a local transformation to an effective $(n-1)$-dimensional map depending on a parameter which is the value of the invariant. This leads to straightforward generalizations of Proposition 1. Furthermore, the reversibility properties R1–R8 are obviously independent of dimension. On the other hand, the dynamical structure for the mappings of Proposition 26 with $n \geqslant 4$ seems to be rather "poor," at least w.r.t. periodic orbits and bifurcation with the value of the invariant as a parameter. The Fibonacci trace map (where $n=3$) is an interesting exception, but in general one would expect more complicated invariants, e.g., rational functions, for reversible systems with more internal structure. Nevertheless, a closer investigation is in progress and we postpone further details.

8. CONCLUDING REMARKS

In this paper, we have studied trace maps derived from invertible two-letter substitution rules with focus on dynamical properties; compare also ref. 9. The key example was the well-known Fibonacci mapping. Many of its properties extend to the more general class of Nielsen trace maps which are 3D volume-preserving dynamical systems with one constant of motion. In suitable coordinates, this constant of motion is the Fricke character $\hat{I}$ of (13) for the entire class. We have given some dynamical consequences of the invariance of $\hat{I}$ due to (local) conjugacies to 2D conservative mappings: the existence of curves of periodic orbits and the expectation of 2D area-preserving period-doubling exponents. Furthermore, we have analyzed one level set of $\hat{I}$ on which the motion is explicitly solvable and pseudo-Anosov. It turns out that trace maps provide simple but interesting examples of such systems. It might be worth mentioning that this establishes a link between the theory of free groups and their automorphisms and the classification of automorphisms of orientable, compact surfaces—both being greatly influenced by Nielsen.

All Nielsen trace maps have the symmetry property of commuting with the group Σ of Proposition 2, which links periodic orbits and their linearizations. Furthermore, a large subclass of them is reversible, which facilitates the location of (symmetric) periodic orbits and has many dynamical consequences.

In Section 6, we asked how general is the presence of reversibility (and its generalizations) in Nielsen trace maps, and showed that reversibility with respect to polynomial reversing symmetries is answered by reversibility in the class of matrices $PGl(2, \mathbb{Z})$. The latter problem is fully solved in ref. 42, but the sufficient conditions given in Section 6 already indicate why reversibility is so prevalent in many trace map examples. We believe that typically a reversing symmetry of a reversible Nielsen trace map preserves the Fricke invariant itself, so that Nielsen trace maps provide many examples of a one-parameter family of reversible dynamical systems on 2D manifolds where the dynamics on one manifold is solvable.

In Section 6, we were able to generalize many of the dynamical properties of the Fibonacci trace map to all Nielsen trace maps. A unifying feature of these results is the importance of the substitution matrix R_F in the dynamics of a Nielsen trace map F. This $Gl(2, \mathbb{Z})$ matrix controls the way F commutes with Σ (Proposition 15), it decides whether F is reversible w.r.t. a large class of reversing symmetries [Propositions 17 and 18 and Eq. (69)], and it controls the dynamics on $\mathcal{M}_0^c$ (Proposition 14). In particular, for hyperbolic R_F, it gives the pseudo-Anosov system on $\mathcal{M}_0^c$.

Finally, in Section 7, we have given some ways to construct nD mappings which are reversible and have an invariant of motion. For $n > 3$, they can be seen as one possible generalization of the trace map dynamical systems.

Let us now close with remarks on possible further developments, some of which we are currently investigating.

(1) As we have remarked at various stages, many aspects of the dynamics of Nielsen trace maps in the real regime carry over immediately to the complex regime. Thus, they provide an interesting example of a one-parameter family of mappings on 2D complex manifolds.

(2) There are of course non-Nielsen trace maps which do not

preserve the Fricke character $\hat{I}$ but do preserve the surface $\mathcal{M}_0$, where again solvability and pseudo-Anosov structure come into play. Some examples also have another constant of motion different from the Fricke invariant,[8] but are dissipative 3D mappings. In such examples, period doubling would be expected with the 2D dissipative scalings.

(3) Finally, as mentioned at the very beginning of this article, the trace maps are just a reduced dynamical system derived from a higher-dimensional matrix dynamical system. Studies of the latter should reveal a more complete picture of the dynamical features and their links to the underlying physical problem; compare ref. 10 and references therein.

APPENDIX. LOCAL EQUIVALENCE BETWEEN 3D MAPPINGS WITH AN INVARIANT AND ONE-PARAMETER 2D MAPPINGS

In this Appendix, we adapt and highlight a result from the theory of diffeomorphisms on manifolds to the dynamical situation suggested by the trace maps. Specifically, we show that there is a local transformation that takes a 3D mapping that possesses an invariant I to a one-parameter 2D mapping, the parameter being the value of the invariant on the nearby level sets. This conjugacy helps to explain many of the properties of such 3D mappings.

Proposition A.1. Let $L: \mathbb{R}^3 \to \mathbb{R}^3$ be a diffeomorphism given by

$$L: \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} f(x, y, z) \\ g(x, y, z) \\ h(x, y, z) \end{pmatrix} \tag{A.1}$$

with a continuously differentiable invariant $I = I(x, y, z)$ such that $I(L\mathbf{x}) = I(\mathbf{x})$, $\mathbf{x} = (x, y, z)$. Assume: (i) there exists an open set $V \subset \mathbb{R}^3$ such that at least one component of ∇I is nonzero in V; and (ii) there is an open set $W \subset V$ and a $k_0 \geqslant 1$ such that $L^k W \subset V$ for $0 < k \leqslant k_0$. Then, locally in W, L is equivalent to a one-parameter 2D mapping, the parameter being the value μ of I on the level sets $\mathcal{M}_\mu := \{\mathbf{x} \in \mathbb{R}^3 \mid I(\mathbf{x}) = \mu\}$ that foliate W.

Proof. From Assumption (i), at least one of $\{\partial I/\partial x, \partial I/\partial y, \partial I/\partial z\}$ is nonzero in V, which implies that the equation $I(x, y, z) = \mu$ locally has a unique inverse when solved, correspondingly, for one of $\{x, y, z\}$. Assumption (ii) then implies that for each point in W, the motion of L is restricted to V for some finite number of iterations (at least). For illustrative purposes, suppose that $\partial I/\partial z \neq 0$ in V, whence $I(x, y, z) = \mu$ can be solved locally in a unique way for z with $(x, y, z) \in V$ via $z = J(x, y, \mu)$. For $\mathbf{x} \in W$

we can then define the transformed local mapping M by the following diagram:

$$\begin{array}{ccc} (x, y, z) & \xrightarrow{L} & (x', y', z') \\ \Big\downarrow P & & \Big\downarrow P \\ (x, y, \mu) & \xrightarrow[M]{} & (x', y', \mu') \end{array} \tag{A.2}$$

where primes denote the image of a point and

$$P: \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \\ \mu = I(x, y, z) \end{pmatrix}, P^{-1}: \begin{pmatrix} x \\ y \\ \mu \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \\ z = J(x, y, \mu) \end{pmatrix} \tag{A.3}$$

Consequently $M = P \circ L \circ P^{-1}$ has the form

$$M: \quad \begin{pmatrix} x \\ y \\ \mu \end{pmatrix} \mapsto \begin{pmatrix} f(x, y, J(x, y, \mu)) \\ g(x, y, J(x, y, \mu)) \\ \mu \end{pmatrix} =: \begin{pmatrix} F(x, y, \mu) \\ G(x, y, \mu) \\ \mu \end{pmatrix} \tag{A.4}$$

The domain of M from the above conjugacy diagram (A.2) is $P(W) \subset P(V)$, which defines a domain interval for μ. The nontrivial part of M is the one-parameter 2D mapping

$$T: \quad \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} F(x, y, \mu) \\ G(x, y, \mu) \end{pmatrix} \quad \blacksquare \tag{A.5}$$

We concentrate on using Proposition A.1 in the vicinity of a point $\mathbf{p}_0 = (x_0, y_0, z_0)$ belonging to a j-cycle of a trace map F with invariant I and $I(\mathbf{p}_0) = \mu_0$. We take $L = F^j$, so that $\mathbf{p}_0$ is a fixed point of L. If $\nabla I(\mathbf{p}_0) \neq 0$, then because ∇I is continuous, Assumption (i) is satisfied and there exists an open set $V \ni \mathbf{p}_0$. Also, by continuity of L at $\mathbf{p}_0$, there always exists, for any finite k_0, a $W \ni \mathbf{p}_0$ with $W \subset V$ that satisfies Assumption (ii). Application of Proposition A1 in this case leads to Proposition 1 of Section 3. The results of the latter Proposition are a consequence of the conjugacy of L to a two-dimensional map like T of (A.5), and the use of the (local) theory of 2D mappings (see, for example, Chapter 2 of ref. 12 and references therein) to get implications on L. Because of the conjugacy, fixed points and periodic orbits of T, and hence M, are fixed points and periodic orbits of L and vice versa, and the eigenvalue spectra of the corresponding cycles of M and L are the same. In particular, the trivial third dimension of M gives $\mu' = \mu$ and contributes an eigenvalue $\{+1\}$ to its linearization. Otherwise, the remaining nontrivial part of M is 2D, like

T above. The continued existence of a fixed point $(x_0(\mu_0), y_0(\mu_0))$ of a one-parameter 2D mapping like T in a range of parameter μ around μ_0 and its isolation from other fixed points (k-cycles) are guaranteed by the implicit function theorem if $dT(x_0(\mu_0), y_0(\mu_0))$ does not have an eigenvalue equal to $+1$ (a kth root of unity). Via the conjugacy (A.2), this translates to the existence of curves of fixed points of L (or j-periodic points of F) and the nonintersection of these curves with those of other periods. Note that for 2D mappings like T, possession of an eigenvalue of $dT(x_0(\mu_0), y_0(\mu_0))$ which is an nth root of unity is equivalent to

$$\operatorname{tr} dT^n(x_0(\mu_0), y_0(\mu_0)) = 1 + \det dT^n(x_0(\mu_0), y_0(\mu_0))$$

Again from the conjugacy, this translates to

$$\operatorname{tr} dL^n(x_0(\mu_0), y_0(\mu_0), z_0(\mu_0)) = 2 + det\, dL^n(x_0(\mu_0), y_0(\mu_0), z_0(\mu_0))$$

ACKNOWLEDGMENTS

The authors are grateful to P. Kramer for pointing out the importance of Nielsen's work to the description of two-letter substitution rules and quasicrystals. It is a pleasure to thank T, Bountis, G. Cairns, H. W. Capel, U. Grimm, D. Joseph, J. S. W. Lamb, G. R. W. Quispel, M. Schlottmann, Z.-Y. Wen, and F. Wijnands for interesting discussions and valuable comments, R. S. MacKay for helpful information about pseudo-Anosov systems and a copy of ref. 28,. and G. Cairns for bringing ref. 55 to our attention. We also thank Z.-Y. Wen for pointing out ref. 56 after he read the preprint version of this article (Melbourne preprint 9-92). This work was supported by Deutsche Forschungsgemeinschaft. J.A.G.R. gratefully acknowledges financial support from the Australian Research Council and the Stichting voor Fundamenteel Onderzoek der Materie (FOM), which is supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO). He thanks P. Kramer and his group for their hospitality in Tübingen, where part of this work was done.

NOTE ADDED IN PROOF

In Props. 23 and 24 of Section 6, we consider reversing symmetries G of trace maps F that satisfy properties [H1] to [H3]. The essential part of these properties is that G preserves the foliation of $\mathbb{R}^3$ by the surfaces $\{\mathcal{M}_\mu\}$, because the rest of [H1] to [H3] effectively follows from this. We might ask what would happen to the dynamics of F if G did *not* preserve this foliation. In this case, $\{\mathcal{M}_\mu\}$ and $\{G\mathcal{M}_\mu\}$ would provide two different foliations of some region of $\mathbb{R}^3$ by 2D surfaces invariant under F. The intersection of these two invariant foliations would typically give a foliation of the region by 1D F-invariant curves. As a consequence, the dynamics of the reversible trace map F would be further restricted on a given level set of the Fricke character to lie on these dense families of curves. This restriction would appear to be atypical behaviour in general, as it implies the existence of a second constant of the motion apart from the Fricke character, which lends support to our assumed properties [H1] to [H3].

REFERENCES

1. P. J. Steinhardt and S. Ostlund, eds., *The Physics of Quasicrystals* (World Scientific, Singapore, 1987).
2. M. Kohmoto, L. P. Kadanoff, and C. Tang, *Phys. Rev. Lett.* **50**:1870 (1983) [reprinted in ref. 1].
3. S. Ostlund, R. Pandit, D. Rand, H.-J. Schellnhuber, and E. D. Siggia, *Phys. Rev. Lett.* **50**:1873 (1983) [reprinted in ref. 1].
4. R. Fricke and F. Klein, *Vorlesungen über automorphe Funktionen*, Vol. 1 (Teubner, Leipzig, 1897).
5. R. D. Horowitz, *Trans. Am. Math. Soc.* **208**:41 (1975).
6. J.-P. Allouche and J. Peyrière, *C. R. Acad. Sci. Paris* **302**(II):1135 (1986).
7. B. Sutherland, *Phys. Rev. Lett.* **57**:770 (1986).
8. M. Baake, U. Grimm, and D. Joseph, *Int. J. Mod. Phys. B* **7**:1527 (1993).
9. J. A. G. Roberts and M. Baake, The dynamics of trace maps, in *Hamiltonian Mechanics: Integrability and Chaotic Behaviour*, J. Seimenis, ed. (Plenum Press, New York, in press).
10. P. Kramer, *J. Phys. A* **26**:213, L245 (1993).
11. W. Magnus, *Math. Z.* **170**:91 (1980).
12. J. A. G. Roberts and G. R. W. Quispel, *Phys. Rep.* **216**:63 (1992).
13. W. Magnus, A. Karrass, and D. Solitar, *Combinatorial Group Theory*, 2nd ed. (Dover, New York, 1976).
14. A. J. Casson and S. A. Bleiler, *Automorphisms of Surfaces after Nielsen and Thurston* (Cambridge University Press, Cambridge, 1988).
15. J.-M. Luck, C. Godrèche, A. Janner, and T. Janssen, *J. Phys. A* **26**:1951 (1993).
16. M. Lothaire, *Combinatorics on Word* (Addison-Wesley, Reading, Massachusetts, 1983).
17. J. Peyrière, *J. Stat. Phys.* **62**:411 (1991).
18. F. Wijnands, *J. Phys. A* **22**:3267 (1989).
19. M. Kolar and M. K. Ali, *Phys. Rev. A* **42**:7112 (1990).
20. R. D. Horowitz, *Commun. Pure Appl. Math.* **25**:635 (1972).
21. A. Whittemore, *Proc. Am. Math. Soc.* **40**:383 (1973).
22. M. H. Vogt, *Ann. Sci. Ecole Norm. Sup. (3)* **6** (Suppl. 3) (1889).
23. L. P. Kadanoff, Applications of scaling ideas to dynamics, in *Regular and Chaotic Motions in Dynamical Systems*, G. Velo and A. S. Wightman, eds. (Plenum Press, New York, 1985).
24. L. P. Kadanoff, Analysis of cycles for a volume preserving map, preprint, University of Chicago (1983) [cited in ref. 36].
25. R. L. Devaney, *An Introduction to Chaotic Dynamical Systems*, 2nd ed. (Addison-Wesley, Redwood City, California, 1989).
26. M. Kohmoto and Y. Oono, *Phys. Lett. A* **102**:145 (1984).
27. P. A. Kalugin, A. Yu. Kitaev, and L. S. Levitov, *Sov. Phys. JETP* **64**:410 (1986).
28. J. Llibre and R. S. MacKay, *Math. Proc. Camb. Phil. Soc.* **112**:539 (1992).
29. M. Golubitsky, I. Stewart, and D. C. Schaeffer, *Singularities and Groups in Bifurcation Theory*, Vol. II (Springer, New York, 1988).

30. J. J. Rotman, *An Introduction to the Theory of Groups*, 3rd. ed. (Allyn and Bacon, Boston, 1984).
31. M. R. Schroeder, *Number Theory in Science and Communication*, 2nd ed. (Springer, Berlin, 1990).
32. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 4th ed. (Clarendon, Oxford, 1960).
33. J. Kollar and A. Sütő, *Phys. Lett. A* **117**:203 (1986).
34. V. G. Benza, *Europhys. Lett.* **8**:321 (1989).
35. U. Grimm and M. Baake, Non-periodic Ising quantum chains and conformal invariance, to appear in *J. Stat. Phys.* (March 1994).
36. M. Casdagli, *Commun. Math. Phys.* **107**:295 (1986).
37. R. L. Devaney, *Trans. Am. Math. Soc.* **218**:89 (1976).
38. M. B. Sevryuk, *Reversible Systems* (Springer, Berlin, 1986).
39. G. D. Birkhoff, *Collected Mathematical Papers*, Vols. 1 and 2, American Mathematical Society, Providence, Rhode Island, 1950).
40. J. S. W. Lamb, *J. Phys. A* **25**:925 (1992).
41. J. S. W. Lamb and G. R. W. Quispel, Reversing k-symmetries in dynamical systems, Amsterdam preprint ITFA 93-16.
42. M. Baake and J. A. G. Roberts, Symmetries and reversing symmetries of trace maps, in *Proceedings 3rd International Wigner Symposium* (Oxford, 1993), L. L. Boyle and A. I. Solomon, eds., to appear; M. Baake and J. A. G. Roberts, in preparation.
43. J. Bellissard, B. Iochum, E. Scoppola, and D. Testard, *Commun. Math. Phys.* **125**:527 (1989); A. Sütő, *J. Stat. Phys.* **56**:525 (1989).
44. R. S. MacKay, *Phys. Lett. A* **106**:99 (1984).
45. K. R. Meyer, *Trans. Am. Math. Soc.* **149**:95 (1970).
46. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).
47. J. Nielsen, *Math. Ann.* **78**:385 (1918).
48. M. Holzer, *Phys. Rev. B* **38**:1709 (1988).
49. M. Baake, D. Joseph, and P. Kramer, *Phys. Lett. A* **168**:199 (1992); D. Joseph, M. Baake, and P. Kramer, *J. Non-Cryst. Solids* **153&154**:394 (1992).
50. W. P. Thurston, *Bull. AMS* **19**:417 (1988).
51. J. Peyrière, Wen Zhi-Ying, and Wen Zhi-Xiong, *L'Enseignement Math.* **39**:153 (1993); Algebraic properties of trace mappings associated with substitutive sequences, *Mod. Math.* (China), to appear.
52. H. S. M. Coxeter and W. O. J. Moser, *Generators and Relations for Discrete Groups*, 2nd ed. (Springer, Berlin, 1965).
53. H. Davenport, *The Higher Arithmetic* (Dover, New York, 1983).
54. K. Iguchi, *Phys. Rev. B* **43**:5915, 5919 (1991).
55. E. Ghys and V. Sergiescu, *Topology* **19**:179 (1980).
56. J. Peyrière, Wen Zhi-Xiong, and Wen Zhi-Yiung, in *Nonlinear Problems in Engineering and Science*, Shutie Xiao and Xian-Cheng Hu, eds. (Science Press, Beijing 1992).
57. P. Pleasants, private communication (1992).
58. J. A. G. Roberts and H. W. Capel, *Phys. Lett. A* **162**:243 (1992); and in preparation.
59. Y. Avishai and D. Berend, *J. Phys. A* **26**:2437 (1993); Y. Avishai, D. Berend, and D. Glaubman, Minimum-dimension trace maps for substitution sequences, preprint, Beer-Sheva (1993).
60. D. Berend, private communication (1993).
61. S. Lang, *Algebra*, 2nd ed. (Addison-Wesley, Menlo Park, California, 1984).
62. M. Kohmoto, *Int. J. Mod. Phys. B* **1**:31 (1987).