Chapter 12

Z_2-systolic freedom and quantum codes

Michael H. Freedman, David A. Meyer, and Feng Luo

Abstract A closely coupled pair of conjectures/questions—one in differential geometry (by M. Gromov), the other in quantum information theory—are both answered in the negative. The answer derives from a certain metrical flexibility of manifolds and a corresponding improvement to the theoretical efficiency of existing local quantum codes. We exhibit this effect by constructing a family of metrics on $S^2 \times S^1$, and other three and four dimensional manifolds. Quantitatively, the explicit "freedom" exhibited is too weak (a $\log^{1/2}$ factor in the natural scaling) to yield practical codes but we cannot rule out the possibility of other families of geometries with more dramatic freedom.

12.0 Preliminaries and statement of results

We define the *p-systole* of a closed Riemannian manifold M to be:

$$\inf_{\alpha \neq 0} p\text{-area}(\alpha)$$

where α is a smooth oriented p-cycle whose class $[\alpha] \neq 0 \in H_p(M;\mathbb{Z})$. Similarly, we define:

$$\mathbb{Z}_2 - p\text{-systole}(M) = \inf_{\alpha \neq 0} p\text{-area}(\alpha)$$

where α is a smooth unoriented cycle whose class $[\alpha] \neq 0 \in H_p(M;\mathbb{Z}_2)$. In dimension 2, for surfaces different from the 2-sphere, it is known* that

$$(1\text{-systole})^2 \leq \frac{\pi}{2}\,(2\text{-systole}) = \frac{\pi}{2}\text{area}$$

(equality holds for the round projective plane). Moreover, the same inequality holds for $\mathbb{Z}_2$-systoles.

In dimension 3 and greater, oriented and unoriented systoles are quite distinct. The oriented theory has been very well developed (for example see [3, 4] and [14], and "systolic freedom" has been established in many cases.

DEFINITION 12.1 *M^d is (p,q)-free, $p+q=d$, if:*

$$\inf_g \frac{(d\text{-systole})}{(p\text{-systole})(q\text{-systole})} = 0$$

where g varies over Riemannian metrics on M. For example, it has been proven [3]) and ([16] that: (i) every $M^d, d \geq 3$ with $b_1(M) \geq 1$ is $(1, d-1)$-free, and (ii) every simply connected smooth 4-manifold is $(2,2)$-free.

In contrast, it has been conjectured by Gromov (see [16]) that in all dimensions, when $\mathbb{Z}_2$-systoles are considered, rigidity as in Loewner's theorem rather than freedom should prevail. We construct a counterexample:

THEOREM 12.1
There exists a family of metrics on $S^2 \times S^1$ exhibiting $\mathbb{Z}_2$-$(1,2)$-systolic freedom.

*M. Katz points out that this follows by combining Loewner's (unpublished) result for T^2, Berger's for $\mathbb{R}P^2$ [1], and Gromov's for other closed surfaces [14].

THEOREM 12.2
There exists a family of metrics on $S^2 \times S^2$ exhibiting $\mathbb{Z}_2$-$(2,2)$-systolic freedom.

REMARK 12.1 When examined quantitatively, we will see that our family only reaches freedom by a highly iterated log factor when measured in the natural scaling, whereas the standard examples in the $\mathbb{Z}$ case, due to Gromov, are much more robust, exhibiting a definite power or even an exponential [19]. We find a bit more freedom, scaling like $(\log)^{1/2}$ in "weak" families with variable topology (see Section 12.2). We do not know if there are families or even weak families with greater than logarithmic freedom in the $\mathbb{Z}_2$-setting. For application to practical local quantum codes, this is a critical question. ∎

Quantum codes were invented in 1995 [21] as the beginning of a solution to the problem of protecting quantum information from errors induced by unwanted interactions of a quantum computer with its environment. Functionally, they are similar to classical codes which protect bit strings (elements of $\mathbb{Z}_2^k$) by encoding them as longer bit strings $\big((b_1, \ldots, b_n) \in \mathbb{Z}_2^n,\, n > k\big)$ in such a way that noise induced errors which flip some of the bits ($b_i \mapsto b_i + 1$) can be corrected. A *quantum bit* (*qubit*) is an element in two dimensional Hilbert space $\mathbb{C}^2 = \mathbb{C}^{\mathbb{Z}_2}$. An n-qubit code for k-qubits is a 2^k dimensional subspace of $(\mathbb{C}^2)^{\otimes n} = \mathbb{C}^{\mathbb{Z}_2^n}$. For a complete discussion of quantum error correcting codes see Gottesman's thesis [13]; here we will simply posit that a quantum code should allow recovery from a set of errors defined by Hermitian operators of the form $\sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n}$, where $\sigma_{i_j} \in \{\mathrm{id}, \sigma_x, \sigma_y, \sigma_z\}$ and

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\ \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},\ \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

are the *Pauli matrices* acting on $\mathbb{C}^2$ in a fixed *"computational" basis* written $(|0\rangle, |1\rangle)$.

Classical linear codes encode k-bit strings into n-bit strings by a linear map $\mathbb{Z}_2^k \longrightarrow \mathbb{Z}_2^n$. The image of this map is the kernel of a linear map $P : \mathbb{Z}_2^n \to \mathbb{Z}_2^{n-k}$ with maximal rank. P is the *parity check matrix* for the code; each row of P is a *parity check* whose $\mathbb{Z}_2$-inner product with each valid *codeword* vanishes. The parity checks form an abelian group under addition, with rank $n - k$. An *error* is an element $e \in \mathbb{Z}_2^n$ with $e_i = 1$ at the bits which are flipped by the noise, so that a codeword $b \in \mathbb{Z}_2^n$ becomes $b + e$. Since $P(b + e) = Pe$, if Pe is different for each of some set of possible errors, each error can be identified by the results of the parity checks and then corrected—independently of the codeword.

The quantum analogue of a set of parity checks that define a classical code is a set of *stabilizer operators*. These are mutually commuting Hermitian operators, which form an abelian group under multiplication. They have the same form as the error operators; so each has eigenvalues ± 1. Just as classical codewords vanish under the action of parity checks, elements of a quantum code are fixed by each stabilizer.

We consider a special class of stabilizer codes: the CSS (Calderbank–Shor [8] and Steane [23]) codes. The group of stabilizers for a CSS code has a set of generators each of which involves only σ_x or only σ_z operators as its nonidentity tensor factors. As σ_x acts on a qubit to interchange the computational basis vectors, it can be described as a bit flip and there is a naturally corresponding classical parity check. Thus to a set of s_x bit flip (σ_x) stabilizers there corresponds a parity check matrix $P_x : \mathbb{Z}_2^n \longrightarrow \mathbb{Z}_2^{s_x}$ with entries set to 1 or 0 corresponding to the presence or absence of a σ_x factor. We can also construct a parity check matrix $P_z : \widehat{\mathbb{Z}}_2^n \to \mathbb{Z}_2^{s_z}$ corresponding to the s_z "phase flip" stabilizers, now with elements set to 1 corresponding to the presence of σ_z factors. Here $\widehat{\ }$ denotes the Fourier dual $H : \mathbb{Z}_2 \to \widehat{\mathbb{Z}}_2$, where

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

is the Hadamard, or 2-dimensional discrete Fourier, transform.

Each of the bit flip stabilizers commutes with the others, and similarly for the phase flip stabilizers. Since $\sigma_x \sigma_z = -\sigma_z \sigma_x$, a bit flip stabilizer commutes with a phase flip stabilizer when the set of qubits on which both act nontrivially is even. Thus the condition that all the stabilizers of a CSS code commute (and therefore have a nonempty fixed set) can be written $P_z P_x^{\mathrm{T}} = 0$. That is, the sequence of maps

$$\mathbb{Z}_2^{s_x} \xrightarrow{P_x^{\mathrm{T}}} \widehat{\mathbb{Z}}_2^n \xrightarrow{P_z} \mathbb{Z}_2^{s_z}$$

satisfies: $\mathrm{im} P_x^{\mathrm{T}} \subset \mathrm{ker} P_z$. The elements of $\mathrm{ker} P_z$ label a basis for the subspace of $(\mathbb{C}^2)^{\otimes n}$ fixed by the σ_z stabilizers, while $\mathbb{Z}_2$-addition by the elements of $\mathrm{im} P_x^{\mathrm{T}}$ defines the orbit of the action of the σ_x stabilizers. Averaging over the orbits gives the subspace simultaneously fixed by *all* the stabilizers. Thus the code subspace is identified with the $\mathbb{Z}_2$-homology $\ker P_z^{\mathrm{T}}/\mathrm{im} P_x^{\mathrm{T}}$ and is $2^{n-s_x-s_z}$ dimensional.*

It is physically natural to realize the topology of CSS codes in a geometrical setting in which there is a *locality* condition requiring each stabilizer to involve only a bounded number of σ_x or σ_z. We will derive such local codes from the Riemannian geometry of a closed smooth manifold M^d. The construction is a straightforward generalization of Kitaev's "toric codes" [12].

Let $\mathcal{C}$ be a piecewise smooth cellulation of M and let $\mathcal{C}^*$ be its dual cellulation. Let C_p denote the $\mathbb{Z}_2$-chain group of p-cells of $\mathcal{C}$. Then the Hilbert space of qubits associated to the p-cells is isomorphic to $(\mathbb{C}^2)^{\otimes n}$ where n is the number of p-cells, which is more functorially written as the functions on the chain group: $\mathbb{C}^{C_p}$. For each $(p-1)$-cell a and $(p+1)$-cell b we define stabilizers A_a and B_b acting on $\mathbb{C}^{C_p}$ with nonidentity tensor factors acting on the qubits associated to the p-cells, π, whose boundary contains a and which lie in the boundary of b, respectively:

$$A_a = \bigotimes_{a\in\partial\pi} \sigma_z^\pi \otimes \bigotimes_{a\notin\partial\pi} \mathrm{id}^\pi \quad \text{and} \quad B_b = \bigotimes_{\pi\in\partial b} \sigma_x^\pi \otimes \bigotimes_{\pi\notin\partial b} \mathrm{id}^\pi.$$

These stabilizers define a CSS code which is the $\mathbb{Z}_2$-homology of the manifold, $H_p = H_p(M;\mathbb{Z}_2)$. The common fixed set of the A operators, $\mathrm{fix}A := \bigcap_{(p-1)-\text{cells } a} \mathrm{fix}\, A_a$, is naturally identified with $\mathbb{C}^{Z_p}$ where Z_p denotes the $\mathbb{Z}_2$-cycles in C_p. The operators B_b act on $\mathbb{Z}_p$ by addition of ∂b and therefore on $\mathrm{fix}A$ by the induced action on functions. The common fixed space $\mathrm{fix}A \cap \mathrm{fix}B$ (where $\mathrm{fix}B := \bigcap_{(p+1)-\text{cells } b} \mathrm{fix}B_b$) under this identification is the space of boundary invariant functions from Z_p to $\mathbb{C}$, i.e., $\mathbb{C}^{H_p}$.

$\mathbb{C}^{H_p}$ is the protected space of a quantum code that protects $\mathrm{rank}(H_p)$ qubits inside n qubits against $\lfloor \frac{t}{2} \rfloor$ errors where t is the minimum of (1) the fewest number of p-cells in an essential $\mathbb{Z}_2$-p-cycle of $\mathcal{C}$, and (2) the fewest number of dual $(d-p)$-cells in an essential $\mathbb{Z}_2$-$(d-p)$-cycle of $\mathcal{C}^*$.

To understand this estimate, a straightforward generalization of the estimate for toric codes [12], we must discuss the error-recovery procedure. We have noted that $\mathrm{fix}A$ is the space of superpositions of cycles, $\mathbb{C}^{Z_p} \subset \mathbb{C}^{C_p} \equiv \mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2$, one copy labeled by each p-cell. The basis $(|0\rangle, |1\rangle)$ for $\mathbb{C}^2$ is used to fix the equivalence, $\equiv$. Since $H\sigma_z H^{-1} = \sigma_x$, conjugating by the Hadamard transformation H on each factor transforms each B_b into $A_{\widehat{b}} = \bigotimes_{\widehat{b}\in\partial\theta} \sigma_x^\theta \otimes \bigotimes_{\widehat{b}\notin\partial\theta} \mathrm{id}^\theta$ where $\widehat{b}$ is the dual q-cell,

*This perspective on CSS codes is implicit in [12] and has been advocated by Greg Kupperberg.

$q := d - p$, to b in the dual cellulation $\mathcal{C}^*$ and θ runs over $(q+1)$-dual cells. Thus fix$B$ can be interpreted in a new way: using the Fourier basis $\{\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\}$ to define equivalence, $\equiv$, above, a state is in fix$B$ iff it is a superposition of dual cycles. All errors may be projected into σ_x and σ_z errors operating on various qubits [13]. The location of such errors may be deduced from the commuting system of measurements $\{A_a, B_b\}$. If the number of σ_x (σ_z) errors $e_x(e_z)$ satisfies: $e_x < \lfloor \frac{t}{2} \rfloor$ ($e_x < \lfloor \frac{t}{2} \rfloor$) then it will be possible to reconstruct all cycles (dual cycles) of the superposition viewed in the $\{|0\rangle, |1\rangle\}$-basis $\big(\{\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\}$-basis$\big)$ by applying σ_z (σ_z) at no more than $\lfloor \frac{t}{2} \rfloor$ p- (q-) cells (dual cells). Doing this cannot change the class (dual class) in H_p^e ($H_q^{e^*}$) since any cycle (dual cycle) will be changed in fewer than $\lfloor \frac{t}{2} - \frac{1}{2} \rfloor + \lfloor \frac{t}{2} \rfloor < t$ cells (dual cells). So after repair, the state is changed by two operators, a null cycle C of σ_xs and by a null dual cycle C^* of σ_zs. Note that nullity implies an even number of intersections $C \cap C^*$; so the two operators commute and their order of application is irrelevant. Nullity further implies that the first operator is a composition of several B_b while the second operator is a composition of several A_a, indexed by the respective coboundary chains. Thus the two operators have no effect on the code space fix$A \cap$ fix$B$ and the repair has been successful.

If a closed Riemannian surface M is given a fine triangulation of bounded geometry—(bounded edge lengths and angles of the triangles)—a condition that ensures locality of the corresponding code based on 1-cell labeled qubits, then up to multiplicative constants, $t \approx 1-$systole(M) and $n \approx$ area(M). Loewner's theorem and its generalizations tell us that no bounded geometry surface code with n qubits can do better than to protect against $\lfloor \frac{t}{2} \rfloor$ worst case errors, where $t \leq C\ n^{1/2}$ and $C = C(V)$ depends only on the valence V of the triangulation. It has been asked whether this square root relation between t and n is intrinsic—coming from the dichotomy of "bit" versus "phase" error—or is merely a feature of "surface codes." We will see that in the manifold context the p- and q-volumes of dual $\mathbb{Z}_2$-cycles can escape (narrowly) from systolic inequalities, and this "freedom" is then mapped back to the world of local quantum codes. We show that higher dimensional codes offer—through the phenomenon of systolic freedom—some slight but larger than constant improvement:

$$t \geq \text{constant} \cdot n^{1/2} \log^{1/2} n \qquad (12.1)$$

in the distance of the code. While an improvement so slight is of only conceptual interest, it is quite open whether better families of metrics can be found that would raise the exponent of n in formula (12.1).

THEOREM 12.3

There is a family of local stabilizer quantum codes of one qubit into n which protect against $\lfloor \frac{t}{2} \rfloor$ worst case errors for $t \geq \text{constant}\cdot n^{1/2} \log^{1/2} n$.

Before leaving surface codes, we note that in the theory of Fuchsian groups the trace of a group element α, $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in SL(2, \mathbb{Z})$, satisfies

$$\mathrm{tr}(\alpha) = 2\cosh\left(\frac{L(\alpha)}{2}\right)$$

where $L(\alpha)$ is the translation length of α. This formula allows the 1-systoles (and the number of curves representing the 1-systole) of many families of Fuchsian quotients to be computed. For example, in [20] the $N \to \infty$ asymptotics of the 1-systole for the quotients of $\Gamma_{-1,p}(N) \subset SL(\mathbb{Z}, R)$ for prime $p \equiv 3(\text{mod } 4)$ are computed for

$$\Gamma_{-1,p}(N) = \left\{ \begin{pmatrix} 1 + N(a + b\sqrt{p}) & N(-C + d\sqrt{p}) \\ N(c + d\sqrt{p}) & 1 + N(a - b\sqrt{p}) \end{pmatrix} : \begin{array}{c} \det = 1 \text{ and} \\ a, b, c, d \in \mathbb{Z} \end{array} \right\}.$$

Expressing the result in terms of the genus $g = g(N)$ of the quotient $\mathbb{H}^2/\Gamma_{-1,p}(N)$ one obtains:

$$\text{1-systole} > \text{constant} \cdot \frac{4}{3} \log g.$$

Taking fine triangulations with vertex valence bounded by 7, one may translate this geometric result into codes:

THEOREM 12.4

There is a constant $C_0 > 0$ so that for any positive integer n, there exists a 7-local CSS code that can protect $2g$ qubits $(\mathbb{C}^2)^{\otimes 2g}$ by imbedding them into $\lfloor C_0 ng \rfloor$ qubits $(\mathbb{C}^2)^{\otimes \lfloor C_0 ng \rfloor}$ so as to protect against $\log g\ n^{1/2}$ worst case errors.

The "$\log g$" reflects the "economy of scale" common in coding theory: "encryption is cheaper by the dozen."

12.1 Mapping torus constructions

To understand the difficulty of producing systolically free families of metrics with $\mathbb{Z}_2$-coefficients (or any torsion coefficients), consider the following low dimensional picture which readily generalizes to all dimensions. Let $A = D^2 \times [0,\epsilon]/(\rho,\theta,0) = (\rho,\theta+\pi,\epsilon)$ be the unit 2-disk cross a short interval—a penny—with opposite faces identified by a π-twist. Any factor disk $D^2 \times t$, $t \in [0,\epsilon]$ is least area among essential relative 2-cycles with integer coefficients. The proof can be based on integral geometry or the divergence theorem applied to $\frac{\partial}{\partial t}$, often called a *calibrating field*. However, if we consider $\mathbb{Z}_2$-coefficients, much smaller essential 2-cycles with the topology of a punctured Klein bottle K^- are present, as shown in Figure 12.1.

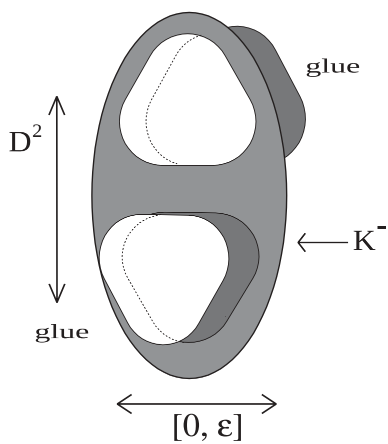

FIGURE 12.1
The topology of a punctured Klein bottle K^-.

Consider the problem of computing a useful lower bound on the (closed) 3-systole of a metric on $M \times S^1$, where M is a closed 3-manifold. There is a resource for this, the “co-area” inequality. In most examples, such as Gromov’s original family of $\mathbb{Z}$-systolically free metrics on $S^3 \times S^1$ [14], the arithmetic of the co-area argument *fails* to force systolically free scaling of the $\mathbb{Z}_2$-systoles. We begin by presenting an extended example of a $\mathbb{Z}$-systolically free family of metrics on $S^3 \times S^1$ for which the co-area argument *fails* to prove $\mathbb{Z}_2$-systolic freedom and a higher dimensional tubing construction analogous to the tube drawn in Figure 12.1 shows that this family of metrics is, in fact, $\mathbb{Z}_2$-systolically rigid. The purpose of the example is to introduce the co-area argument first in families of manifolds $\{S^3_R\}$ and to see how the isoperimetric exponent of the S^3_R "fibers" enters the arithmetic. This will motivate our switch to an example based on arithmetic surfaces whose underlying hyperbolic geometry possesses a linear isoperimetric inequality. This, together with the logarithmic growth (with genus) of the systoles of arithmetic surfaces (see Section 12.0), puts us over the boundary from rigidity to systolic freedom.

Example 12.1

Set $Q_R = S^3_R \times [0, R^{1/3}]/\text{twist}$, where twist means an $R^{2/3}$ isometric rotation along the Hopf fibers of the 3-sphere of radius R, S^3_R. Again, applying the divergence theorem to the calibrating field $\frac{\partial}{\partial t}$ shows that the factor S^3_R realizes the $\mathbb{Z}-3$-systole. The $\mathbb{Z}-1$-systole is approximately realized both by the generator of $H_1(S^3 \times S^1; \mathbb{Z})$ and by its $R^{1/3}$ power, each of which has length scaling like $R^{2/3}$. Thus, the $\mathbb{Z}$-systoles scale like:

$$\frac{\mathbb{Z}-4\text{-systole} \sim R^{10/3}}{\mathbb{Z}-3\text{-systole} \sim R^3 \quad \mathbb{Z}-1\text{-systole} \sim R^{2/3}}.$$

Since $3 + \frac{2}{3} > 3\frac{1}{3}$, this family is $\mathbb{Z}$-systolically free.

Because the dimension of the ambient manifold is no more than 7 and we are considering a codimension 1 $\mathbb{Z}_2$-class, we may apply a theorem of Federer [12] to represent the $\mathbb{Z}_2-3$-systole by an embedded (but not necessarily orientable) 3-manifold $X_R \subset Q_R$. Clearly, the $\mathbb{Z}-3$-systole, which infimizes over a smaller set, dominates the $\mathbb{Z}_2-3$-systole; so we have:

$$\text{vol}(S^3_R) \geq \text{vol}(X_R)$$

or if $x = \text{ scaling }\big(\text{vol }(X_R)\big) = \limsup \log_R \text{vol}(X_R)$ is the growth exponent for the $\mathbb{Z}_2-3$-systole, we have $3 \geq x$.

The co-area inequality [6] states

$$\int_{t=0}^{R^{1/3}} \text{area}(X_R \cap S^3_R \times t)\text{d}t \leq \text{vol}(X_R),$$

so

$$A_R = \min \text{ area}(X_R \cap S^3_R \times t) \leq \frac{\text{vol}(X_R)}{R^{1/3}} \qquad t \in [0, R^{1/3}].$$

To get an inequality on growth rates, set $w = \limsup \log_R A_R$, so that $w \leq x - \frac{1}{3}$.

If the scaling of the $\mathbb{Z}-3$-systole is by $x < 3$, we can attempt to reach a contradiction by cutting and gluing X_R to an orientable representative Z for the generator of $H_3(Q_R;\mathbb{Z})$. This uses a topological lemma:

LEMMA 12.1

If $Z \subset S^3 \times S^1$ is an embedded 3-manifold representing the nonzero element of $H_3(S^3 \times S^1; \mathbb{Z}_2)$ and $Z \cap (S^3 \times t) = \emptyset$ for some t, then Z is orientable and represents a nontrivial element, in fact a generator if Z is connected, of $H_3(S^3 \times S^1; \mathbb{Z})$.

PROOF of Lemma 12.1 $Z \subset S^3 \times (0,1) \subset S^3 \times S^1$ and must carry the $\mathbb{Z}_2-3$-cycle in $S^3 \times (0,1)$. It follows that Z separates the two ends of $S^3 \times (0,1)$, is orientable (use a normal vector field to Z to construct the orientation) and therefore is essential in $H_3\big(S^3 \times (0,1); \mathbb{Z}\big)$. The lemma follows. ▯

Let W_R be a t-cross section of X_R with scaling w and let $Y_R \subset S^3_R \times t$ bound W_R. Using the isoperimetric inequality in S^3_R, which asymptotically is Euclidean 3-space, we see that:

$$y = \text{scaling}\left(\text{vol}(Y_R)\right) \leq \frac{3}{2} w.$$

Now $Z_R = \left(X_R\text{-neighborhood } (W_R)\right) \cup 2\text{-copies } (Y_R)$ is $\mathbb{Z}_2$-homologous to X_R and disjoint from $S^3_R \times t$; so by Lemma 12.1, Z_R is orientable and must have volume at least as great as the $\mathbb{Z} - 3$-systole. Thus, if $x < 3$ then most of the volume of Z_R must come from the patch Y_R. The scaling of the patch volume is:

$$\frac{3}{2}\left(x - \frac{1}{3}\right) \geq \frac{3}{2} w \geq 3$$

so $x \geq \frac{7}{3}$.

The scaling as estimated (and, in fact, true) for the $\mathbb{Z}_2 -$ 4-; 3- and 1-systoles of Q_R as powers of R are:

$$\frac{R^{10/3}}{R^{7/3} \cdot R^{2/3}}.$$

We have only proved $\frac{7}{3}$ is a lower bound, but a construction similar to the "penny" example realizes that exponent. Since $\frac{7}{3} + \frac{2}{3} \leq \frac{10}{3}$, the family Q_R is not $\mathbb{Z}_2$-systolically free. The interval scaling of $Q_R, R^{1/3}$, was chosen to come as close as possible to $\mathbb{Z}_2$-systolic freedom among similar metrics. $\square$

Next we turn to the construction of $\mathbb{Z}_2 - (1,2)$-free metrics using mapping cylinders of arithmetic surfaces $\Sigma_g = \Gamma_{-1,p}(N)$ from Section 12.0. It seems reasonable at this point to define a variant *weak* c-(p,q) freedom, where the coefficient $c \cong \mathbb{Z}$ or $\mathbb{Z}_2$. We say a dimension $d = p+q$ has weak c-(p,q) freedom if there is a family of closed Riemannian d-manifolds M_i with nonvanishing $H_j(M_i; c)$, $j = p, q$ so that the systolic ratio is

$$\frac{d - c\text{-systole } (M_i)}{p - c\text{-systole } (M_i) \cdot q - c\text{-systole } (M_j)} \longrightarrow 0.$$

Thus, in the definition of weak freedom, we allow variable topology. The negation of weak freedom is *strong rigidity*. In Section 12.0, we commented that dimension 2 is strongly (1,1) rigid. The reason for introducing the notion of weak freedom is that when we consider (Section 12.3) the quantitative *freedom function* and *weak freedom function*, our examples exhibit substantially less freedom than weak freedom and it is weak freedom that has the more direct relevance to the construction of local quantum codes.

We will use the genus $g = g(N)$ of the surface $\Sigma_g = \mathbb{H}^2/\Gamma_{-1,p}(N)$ as our parameter. We record the important geometric properties of Σ_g:

$$1\text{-systole}(\Sigma_g) > C_1 \log g.$$

Also by a variant of Selberg's theorem (see [6]), the first eigenvalue of the Laplacian (on functions) exceeds some constant $C_2 > 0$, with

$$\lambda_1(\Sigma_g) > C_2. \tag{12.2}$$

By an analysis of the Dirichlet integral [7], (12.2) implies a uniform linear isoperimetric inequality for a null bounding 1-manifold $\gamma \subset \Sigma_g, \gamma = \partial A = \partial B, A \cup_\partial B = \Sigma_g$, area $A \leq$ area B:

$$\text{area}(A) \leq C_3 \text{ length}(\gamma), \tag{12.3}$$

C_3 independent of g.

Now (12.3) leads to a differential inequality on the Morse theory of "distance from a base point $*_g \subset \Sigma_g$":

$$\frac{\mathrm{d}}{\mathrm{d}t} \text{ area}\big(\text{Ball}_*(t)\big) = \text{length}\,(\partial_t) \geq \frac{1}{C_3} \text{ area}\big(\text{Ball}_*(t)\big),$$

as long as area$\big(\text{Ball}_*(t)\big) \leq \frac{1}{2}\text{area}\,(\Sigma_g)$. It follows that

$$\text{diameter}\,(\Sigma_g) < C_4 \log g, \tag{12.4}$$

C_4 independent of g. From (12.4) and the Morse theory of the distance function from a base point, we find an embedded wedge of circles W^*, which is the union of descending 1-manifolds spanning $H_1(\Sigma_g;\mathbb{Z}_2)$ with a length estimate (12.5). By the embedded property, we may find $W \subset W^* \subset \Sigma_g$ and using Alexander duality, the complement $\Sigma_g \setminus W$ is a disk, so $H_1(W;\mathbb{Z}_2) \longrightarrow H_1(\Sigma_g;\mathbb{Z}_2)$ is an isomorphism, and

$$\text{radius}_*(W) \leq C_4 \log g. \tag{12.5}$$

Three properties of Σ_g (for an infinite set of gs) will serve as input to our construction:

1. Σ_g has a uniform linear isoperimetric inequality (12.3).
2. There exists an isometry $\tau : \Sigma_g \to \Sigma_g$, with order$(\tau) \geq C_5(\log g)^{1/2}$.
3. The map $\Sigma_g \to \Sigma_g/\langle \tau(\sigma) \equiv \sigma\rangle \;=:\; {}_gS$ is a covering projection to the surface ${}_gS$ (of genus $< g$) and ${}_gS$ has injectivity radius$({}_gS) \geq C_6(\log g)^{1/2}$

where C_4, C_5 and C_6 are positive constants independent of g.

In Lemma 2 of [20] it is proved that:

$$\text{inj. rad.}\big(\mathbb{H}^2/\Gamma_{-1,p}(N)\big) = \mathcal{O}(\log N)$$

and in the proof of Theorem 6 of [20] that genus$\big(\mathbb{H}^2/\Gamma_{-1,p}(N)\big) =: g(\Sigma_g) =: g(N)$ satisfies:

$$\mathcal{O}(N^2) \leq g(N) \leq \mathcal{O}(N^3)$$

so

$$\text{inj. rad.}(\Sigma_g) = \mathcal{O}(\log g). \tag{12.6}$$

Now choose a sequence of h and g to satisfy $\log g = \mathcal{O}(\log h)^2$ and so that $N(h)$ divides $N(g)$. Thus, we have a covering projection $\Sigma_g \longrightarrow \Sigma_h$. Let α be the shortest essential loop in Σ_h. By (12.6) length$(\alpha) = \mathcal{O}(\log h)$. Choosing a base point on α, $[\alpha] \in \Gamma_{-1,p}\big(N(h)\big)/\Gamma_{-1,p}\big(N(g)\big)$ satisfies:

$$\text{order}[\alpha] \geq \mathcal{O}(\log h) = \mathcal{O}(\log g)^{1/2},$$

since the translation length of $\alpha = \mathcal{O}(\log g)^{1/2}$ must be multiplied by $\mathcal{O}(\log g)^{1/2}$ before it reaches length $\mathcal{O}(\log g)$, a necessary condition to be an element in the subgroup $\Gamma_{-1,p}\big(N(g)\big)$.

Let τ be the translation determined by $[\alpha]$. We have just checked condition (2) above: $\mathrm{order}(\tau) > \mathcal{O}(\log g)^{1/2}$. Factor the previous covering as:

$$\Sigma_g \longrightarrow \Sigma_g/\langle\tau\rangle \longrightarrow \Sigma_h$$

and set $\Sigma_g/\langle\tau\rangle =: {}_gS$. Since ${}_gS$ covers Σ_h, we conclude condition (3):

$$\text{inj. rad.}({}_gS) \geq \text{inj. rad.}(\Sigma_h) \geq \mathcal{O}(\log h) = \mathcal{O}(\log g)^{1/2}.$$

Let $M_g = (\Sigma_g \times \mathbb{R})/\langle (x,t) \equiv (\tau x, t+1)\rangle$ be the Riemannian "mapping torus" of τ. We can also think of $M_g = \Sigma_g \times [0,1]/\langle (x,0) \equiv (\tau x, 1)\rangle$. Our first objective is to describe, with quantitative estimates, a sequence of Dehn surgeries that transforms M_g into a topological $S^2 \times S^1$. Since the number and length of these surgeries is so extravagant, we will describe as an alternative $\mathcal{O}(g)$ surgeries of length $\mathcal{O}(\log g)$ that transform M_g into a $\mathbb{Z}_2$-homology $S^2 \times S^1$. The second sequence is sufficient to exemplify weak freedom. By two theorems of Lickorish [17], we may first write out τ^{-1} in the mapping class group of Σ_g as a product of d_g Dehn twists σ_i along simple loops $\gamma_i \subset \Sigma_g$:

$$\tau^{-1} = \sigma_{d_g} \circ \cdots \circ \sigma_2 \ \circ \ \sigma_1$$

and second, performing Dehn surgeries along pushed-in copies of $\{\gamma_i\}$,

$$\left\{\gamma_1 \times \left(\frac{1}{2} + \frac{1}{3d_g}\right), \gamma_2 \times \left(\frac{1}{2} + \frac{2}{3d_g}\right), \dots, \gamma_i \times \left(\frac{1}{2} + \frac{i}{3d_g}\right), \dots, \gamma_{d_g} \times \left(\frac{1}{2} + \frac{1}{3}\right)\right\}$$

obtain a diffeomorphic copy of $\Sigma_g \times [0,1]$ whose product structure, when compared to the original, gives $[\tau^{-1}] : \Sigma_g \times 1 \longrightarrow \Sigma_g \times 1$.

Thus, d_g Dehn surgeries on M_g produce the mapping torus for $\tau^{-1} \circ \tau$, i.e., $\Sigma_g \times S^1$. To allow us to estimate the freedom function in this family Luo has computed (see Appendix) an upper bound $C(g)$ to both the number and length of closed geodesics $\{\gamma_i\}$. Unfortunately, $F(g)$ has the form $F(g) = g^{g^{g^{\cdots g}}}$ ($3g-3$ many exponents).

To convert $\Sigma_g \times S^1$ to $S^2 \times S^1$ an additional $2g$ Dehn surgeries are needed. Do half (a "sub-kernel") of these surgeries at level $\frac{1}{2} + \frac{1}{6d_g}$ and the dual half at level $\frac{1}{2}$. The result of all $d_g + 2g$ Dehn surgeries is topologically $S^2 \times S^1$, and once these surgeries are metrically specified, we obtain a sequence of Riemannian 3-manifolds $(S^2 \times S^1)_g$. To merely establish $\mathbb{Z}_2$-freedom, we do not need Luo's estimates; the estimate is used in Section 12.3 to quantify the amount of $\mathbb{Z}_2$-freedom.

We now take up the topologically easier problem of simply converting M_g to a $\mathbb{Z}_2$-homology $S^2 \times S^1$. Applying the Gysin sequence [22] we have:

$$H_1(\Sigma_g; \mathbb{Z}_2) \xrightarrow{1-\tau_*} H_1(\Sigma_g; \mathbb{Z}_2) \to H_1(M_g; \mathbb{Z}_2) \to \mathbb{Z}_2 \to 0$$

or

$$0 \to \operatorname{coker}(1 - \tau_*) \to H_1(M_g; \mathbb{Z}_2) \to \mathbb{Z}_2 \to 0.$$

Let $r = \text{rank image}(1 - \tau_*)$. Take from the petals of W (12.5) a basis of loops $W_1, \ldots, W_{2g}$ for $H_1(\Sigma_g; \mathbb{Z}_2)$; these have length $\mathcal{O}(\log g)$. At most r of these loops lie in image$(1 - \tau_*)$ so there is a subset reordering the index $\{W_1, \ldots, W_{2g-r}\}$ spanning cokernel $(1 - \tau_*)$. It is elementary (see, e.g., [6]) that surgeries (with any framings) applied to $\left\{W_1 \times \frac{1}{2} + \frac{1}{2(2g-r)}, W_2 \times \frac{1}{2} + \frac{2}{2(2g-r)}, \ldots, W_{2g-r} \times 1\right\}$ kill coker $(1-\tau_*)$ and produce a $\mathbb{Z}_2$-homology, $S^2 \times S^1$, which we denote P_g.

In Section 12.2 where $\mathbb{Z}_2$-freedom is established, four metrical properties of these surgeries will be required. They are:

1. *The core curves for the Dehn surgeries are taken to be geodesics in $\Sigma_g \times [0, 1]$ so that the boundaries $\partial T_{i,\epsilon}$ of their ϵ neighborhoods are Euclidean flat. (Flatness follows from translational symmetry.)*

2. *The replacement solid tori $T'_{i,2\epsilon}$ have $\partial T'_{i,\epsilon}$ isometric to $\partial T_{i,\epsilon}$ and are defined as twisted products $D^2 \times [0, 2\pi\epsilon]/\beta$ (the meridians in $T_{i,\epsilon}$ have length $2\pi\epsilon$) where β is an isometric rotation of the disk D^2 adjusted to equal the holonomy obtained by traveling orthogonal to the surgery slopes in $\partial T_{i,\epsilon}$ from $\partial D^2 \times pt$ back to itself.*

3. *The geometry on the disk D^2 above is rotationally symmetric and has a product collar on its boundary as long as the boundary itself, yet area$(D^2) \leq \mathcal{O}\big((\text{length}\, \partial D^2)^2\big)$.*

4. *Finally, $\epsilon > 0$ is so small that the total volume of all the replacement solid tori, $\cup_i T'_{i,\epsilon}$, is $o(g)$.*

With specifications (1) . . . (4), Dehn surgery yields a piece-wise smooth Riemannian manifold for which all the relevant notions of p-area are defined. We could work in this category but there is no need to do so since perturbing to a smooth metric will not effect the status of $\mathbb{Z}_2$-systolic freedom (rigidity) in either strong or weak forms.

12.2 Verification of freedom and curvature estimates

We regard the Riemannian manifolds $(S^2 \times S^1)_g$ and P_g as essentially specified in Section 12.1. Technically, there is the parameter ϵ which controls the "thickness" of the Dehn surgeries. On two occasions (Propositions 12.1 and 12.2), we demand this to be sufficiently small at the cost of an increase in the maximum absolute value of the Riemann curvature tensor as a function of g.

To obtain the family P_g, we performed $\mathcal{O}(g)$ surgeries on loops of length $\mathcal{O}(\log g)$. By the "collar theorem" (see Section 3 of [6]), if each loop is represented by a simple geodesic in a Σ_g level it has a collar of length $e^{-\mathcal{O}(\log g)}$ in that level. In the interval direction the surgeries are separated by a distance $\mathcal{O}(g^{-1})$; so together we may find disjoint tubular neighborhoods $T_{i,\epsilon}$ of radius:

$$\epsilon = \left(\frac{1}{g}\right)^{\alpha}, \tag{12.7}$$

for some $\alpha > 1$.

To ensure that Dehn surgery does not substantially increase $\mathrm{vol}(M_g) = \mathrm{vol}(\Sigma_g \times [0,1]) = \mathrm{area}(\Sigma_g) = 2\pi\chi(\Sigma_g) = \mathcal{O}(g)$, we need:

$$\mathcal{O}(g) \cdot \text{area } D_i^2 \cdot 2\pi\epsilon \leq \mathcal{O}(g),$$

or from the long collar property (C) of D_i^2:

$$\mathcal{O}(g)(\log g)^2 \cdot \epsilon \leq \mathcal{O}(g)$$

so

$$\epsilon \leq \mathcal{O}(\log g)^{-2}. \tag{12.8}$$

Comparing with line (12.7), we see that this condition (12.8) is less stringent so we may pick $\epsilon = \left(\frac{1}{g}\right)^{\alpha}$, some $\alpha > 1$.

To obtain $(S^2 \times S^1)_g$ with volume $\mathcal{O}(g)$, ϵ must be chosen fantastically small. Again, the collar theorem gives the controlling restriction on ϵ: $\epsilon < e^{-F(g)}$. The volume condition:

$$F(g) \cdot F(g)^2 \cdot 2\pi\epsilon \leq \mathcal{O}(g)$$

is less stringent. We have proved:

PROPOSITION 12.1

Provided $\epsilon < \left(\frac{1}{g}\right)^\alpha$, for some $\alpha > 1$ the family P_g has $\mathrm{vol}(P_g) = \mathbb{Z}_2 - 3$-systole$(P_g) = \mathcal{O}(g)$. *Provided $\epsilon < e^{-F(g)}$, the family $(S^2 \times S^1)_g$ has* $\mathrm{vol}(S^2 \times S^1)_g = \mathbb{Z}_2 - 3$-systole$(S^2 \times S^1)_g = \mathcal{O}(g)$.

The next proposition is more subtle.

PROPOSITION 12.2

Provided $\epsilon < \mathcal{O}\left(\frac{1}{\log g}\right)$, the family P_g has $\mathbb{Z}_2 - 2$-systole $(P_g) = \mathcal{O}(g)$. *Provided $\epsilon < g \cdot F(g)^{-2}$, the family $(S^2 \times S^1)_g$ has* $\mathbb{Z}_2 - 2$-systole $(S^2 \times S^1)_g = \mathcal{O}(g)$.

Note: The corresponding bounds on ϵ are less strict in Proposition 12.2 than in Proposition 12.1; so, in practice, we must choose ϵ to satisfy the stricter bounds in Proposition 12.1.

PROOF of Proposition 12.2 For this proof only let N_g denote either P_g or $(S^2 \times S^1)_g$. According to [12] a nonoriented minimizer among all nonzero codimension 1 cycles always exists and is smooth provided the ambient dimension is no more than 7. Let $X_g \subset N_g$ denote these minimizers. The argument in the two cases is parallel so no confusion should result from the double use of the symbol X_g. For a contradiction, assume $\mathrm{area}(X_g) < \mathcal{O}(g)$.

The Dehn surgeries in Section 12.1 were confined to $\Sigma_g \times \left[\frac{1}{2}, 1\right]$; so the surfaces $\Sigma_g \times t$, $t \in \left(0, \frac{1}{2}\right)$ persist as submanifolds of N_g. By Sard's theorem, for almost all $t_0 \in \left(0, \frac{1}{2}\right)$, $\Sigma_g \times t_0$ intersects X_g transversely. Let W_t, $t \in \left(0, \frac{1}{2}\right)$ denote the intersection. By the co-area formula,

$$\mathcal{O}(g) > \text{area } (X_g) \geq \int_{t=0}^{1/2} \text{length}(W_t) \mathrm{d}t.$$

Consequently, for some transverse $t_0 \in \left(0, \frac{1}{2}\right)$,

$$\text{length}\,(W_{t_0}) < \mathcal{O}(g). \tag{12.9}$$

Since both $\Sigma_g \times t_0$ and X_g represent the nonzero element of $H_2(N_g; \mathbb{Z}_2)$, the complement $N_g \setminus (\Sigma_g \times t_0 \cup X_g)$ can be two-colored into black and white regions (change colors when crossing either surface) and the closure B of the black points, say, is a piecewise smooth $\mathbb{Z}_2$-homology between $\Sigma_g \times t_0$ and X_g.

For homological reasons, the reverse Dehn surgeries $N_g \rightsquigarrow M_g$ to the τ-mapping torus have cores with zero (mod 2) intersection with X_g. This means that the tori $\partial T_{i,\epsilon} = \partial T'_{i,\epsilon}$ each meet X_g in a null homologous, possibly disconnected, 1-manifold $X_g \cap \partial T_{i,\epsilon} \subset \partial T_{i,\epsilon}$. Again, if ϵ is a sufficiently small function of g, we may "cut off" X_g along these tori to form $X'_g = (X_g \setminus \cup_i T_{i,\epsilon}) \cup \delta_i$, where δ_i denotes a bounding surface for $X_g \cap \partial T_{i,\epsilon}$ in $\partial T_{i,\epsilon}$, with negligible increase in area. Note that $X'_g \subset M_g$. In particular, we still have:

$$\text{area}(X'_g) < \mathcal{O}(g), \tag{12.10}$$

provided

$$(\#\text{ surgeries}) \cdot (\text{max length surgery}) \cdot \epsilon < \mathcal{O}(g). \tag{12.11}$$

In the cases $N_g = P_g$ and $(S^2 \times S^1)_g$, (12.11) reduces, respectively, to

$$\mathcal{O}(g) \cdot \mathcal{O}(\log g) \cdot \epsilon < \mathcal{O}(g),$$

and

$$F(g) \cdot F(g) \cdot \epsilon < \mathcal{O}(g).$$

Specifically, we choose δ_i to be the "black" piece of $\partial T_{i,\epsilon}$, i.e., $\delta_i \subset B$. If we set

$$B' = \text{closure}(B \setminus \bigcup_i T_{i,\epsilon})$$

and recall

$$\bigcup_i T_{i,\epsilon} \cap \Sigma'_g \times t_0 = \emptyset,$$

we see that B' is a $\mathbb{Z}_2$-homology from X'_g to $\Sigma_g \times t_0$.

It is time to use property (i), i.e., (12.3): W_{t_0} separates $\Sigma_g \times t_0$ into two subsurfaces meeting along their boundaries: One subsurface sees black on the positive side, the other on its negative side. Thus, the smaller of these two subsurfaces, call it $Y \subset \Sigma_g \times t_0$, must satisfy:

$$\text{area}(Y) \leq C_3 \text{length}(W_{t_0}),$$

where C_3 is independent of g. Combining with (12.9), we have:

$$\text{area}(Y) < \mathcal{O}(g). \tag{12.12}$$

Now modify X'_g to Z_g by cutting along W_{t_0} and inserting two parallel copies of Y. This may be done so that the result is disjoint from $\Sigma_g \times t_0$ but bordant to it by a slight modification B'' of B', with B'' still disjoint from $\Sigma_g \times t_0$. See Figure 12.2 below.

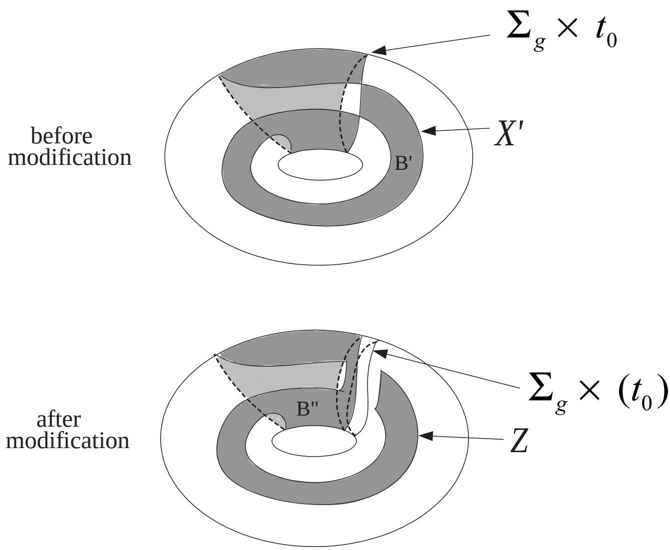

FIGURE 12.2
Modification of X'_g to Z_g.

Combining (12.10) and (12.12),

$$\text{area}(Z_g) < 3 \cdot \mathcal{O}(g) = \mathcal{O}(g).$$

Now reverse the Dehn surgeries and consider:

$$B'' \subset M_g \setminus \Sigma_g \times (t_0) \subset M_g. \tag{12.13}$$

The middle term of (12.13) is diffeomorphic to $\Sigma_g \times \mathbb{R}$, which is a codimension 0 submanifold of $\mathbb{R}^3$. This proves that B'', and in particular Z, is orientable. But this looks absurd; we have constructed an oriented surface Z oriented-homologous to the fiber $\Sigma_g \times t_0$ of M_g with smaller area.

As in Example 12.1, let $\frac{\partial}{\partial t}$ be the divergenceless flow in the interval direction. Lift Z to $\widetilde{Z}$ in the infinite cyclic cover $\Sigma_g \times \mathbb{R}$ and consider the flow through the lift $\widetilde{B}''$, the lift of B''. The divergence theorem states that the flux through $\widetilde{Z}$ is equal to the flux through $\Sigma_g \times t_0$. Since $\frac{\partial}{\partial t}$ is orthogonal to $\Sigma_g \times t_0$,

$$\text{area}\,(\Sigma_g \times t_0) \leq \text{area}\,(\widetilde{Z}) = \text{area}\,(Z) \tag{12.14}$$

completing the contradiction. ∎

PROPOSITION 12.3
$\mathbb{Z}_2$ − 1-systole(P_g) $\geq$ $\mathcal{O}(\log g)^{1/2}$ and $\mathbb{Z}_2$ − 1-systole$(S^2 \times S^1_g)$ $\geq \mathcal{O}(\log g)^{1/2}$.

PROOF of Proposition 12.3 We actually show that any homotopically essential loop obeys this estimate. The long collar condition, (C) in Section 12.1, implies that any arc in $T'_{i,\epsilon}$ with end points on $\partial T'_{i,\epsilon}$ can be replaced with a shorter arc with the same end points lying entirely within $\partial T'_{i,\epsilon}$. It follows that any essential loop in P_g or $(S^2 \times S^1)_g$ can be homotoped to a shorter loop lying in the complement of the Dehn surgeries.

Thus, it is sufficient to show that any homotopically essential loop γ in M_g has length $\gamma \geq \mathcal{O}(\log g)^{1/2}$. For a contradiction, suppose the opposite. Since the bundle projection $\pi : M_g \to [0,1]/\langle 0 = 1\rangle$ is length nonincreasing, degree$(\pi \circ \gamma) < \mathcal{O}(\log g)^{1/2}$. Lift $\gamma \setminus pt.$ to an arc $\widetilde{\gamma}$ in $\Sigma_g \times \mathbb{R}$. The lift $\widetilde{\gamma}$ joins some point (p, t) to $(\tau^d p, t + d)$ where $d =$ degree$(\pi \circ \gamma)$. Since $d < \mathcal{O}(\log g)^{1/2}$ and since condition (2) from Section 12.1 requires order$(\tau) \geq \mathcal{O}(\log g)^{1/2}$, we see that p and $\tau^d p$ differ by a nontrivial covering translation of the cover $\Sigma_g \to {}_gS$. But any non-trivial covering translation moves each point of the total space at least twice the injectivity radius of the base, a quantity guaranteed by (3) of Section 12.1 to be at least $\mathcal{O}(\log g)^{1/2}$. Now using that the projection $\Sigma_g \times \mathbb{R} \to \Sigma_g$ is also length nonincreasing, we see that length$(\widetilde{\gamma}) \geq \mathcal{O}(\log g)^{1/2}$. Since length$(\widetilde{\gamma})$ = length(γ), the same estimate applies to γ. ∎

THEOREM 12.5
Given the restrictions on $\epsilon > 0$ imposed in Proposition 12.1, the family P_g exhibits weak $\mathbb{Z}_2 - (2,1)$-systolic freedom. The family $(S^2 \times S^1)_g$ exhibits $\mathbb{Z}_2 - (2,1)$-systolic freedom.

PROOF of Theorem 12.5 From Propositions 12.1, 12.2 and 12.3, we have:

$$\frac{\mathbb{Z}_2 - 3\text{-systole}(P_g)}{\mathbb{Z}_2 - 2\text{-systole}(P_g) \cdot \mathbb{Z}_2 - 1\text{-systole}(P_g)} \leq \frac{\mathcal{O}(g)}{\mathcal{O}(g)\ \mathcal{O}(\log g)^{1/2}} \longrightarrow 0,$$

and the same statement holds replacing P_g by $(S^2 \times S^1)_g$. ▮

Let us now estimate the maximum absolute value of curvatures and their first covariant derivatives, maximum $\{|\nabla_h R^\ell_{ijk}|, |R^\ell_{ijk}|\} =: R(\gamma)$ for the two families P_g and $(S^2 \times S^1)_g$. Of course piecewise smooth constructions may have "infinite" curvature at the gluing locus, but after rounding the corners, the residual curvature is dominated by $\mathcal{O}\big(\frac{1}{\ell}\big)^2$ where ℓ is the smallest length scale of the construction, and the first derivatives are dominated by $\mathcal{O}\big(\frac{1}{\ell}\big)^3$. In our case $\ell = \epsilon$. Thus, for the family P_g,

$$R(g) \leq \mathcal{O}(g^{3\alpha}), \tag{12.15}$$

for some $\alpha > 1$, and for $(S^2 \times S^1)_g$,

$$R(g) \leq e^{3F(g)}. \tag{12.16}$$

As in [11], let us consider the $\mathbb{Z}_2$-freedom function of both families. This function quantifies the amount of freedom present in a family. To do this, we homothetically rescale each member of the family to be as small as possible and yet have $R \leq 1$, i.e., all its curvatures and first derivatives of curvature lying in the interval $[-1, 1]$.

This is accomplished by rescaling P_g by $\mathcal{O}(g^{\frac{3}{2}\alpha})$ as suggested by (12.15) and (12.16). We find:

$$\frac{\mathbb{Z}_2 - 3\text{-systole}}{\mathbb{Z}_2 - 2\text{-systole} \cdot \mathbb{Z}_2 - 1\text{-systole}} \leq \frac{\mathcal{O}(g^{3\alpha})\ \mathcal{O}(g)}{\mathcal{O}(g^{2\alpha})\ \mathcal{O}(g) \cdot \mathcal{O}(g^{\alpha})\ \mathcal{O}(\log^{1/2} g)}. \tag{12.17}$$

Now we ask for a function which will serve as a lower bound to $\frac{d(n)}{n}$ where n and d are the numerator and denominator (as a function of the numerator) in the right hand side of (12.17). This, by definition, is a lower bound to the $\mathbb{Z}_2$-freedom of the family. An easy computation yields:

$$\frac{d(n)}{n} = \mathcal{O}(\log^{1/2} n),$$

so $\{P_g\}$ has (at least) $\log^{1/2}$ $\mathbb{Z}_2$-systolic freedom.

Rescaling $(S^2 \times S^1)_g$ (according to (12.16)) by $e^{F(g)}$ we find:

$$\frac{\mathbb{Z}_2 - 3\text{-systole}}{\mathbb{Z}_2 - 2\text{-systole} \cdot \mathbb{Z}_2 - 1\text{-systole}} \leq \frac{e^{3F(g)}\ \mathcal{O}(g)}{e^{2F(g)}\ \mathcal{O}(g) \cdot e^{F(g)}\ \mathcal{O}(\log^{1/2} g)}$$

or

$$\frac{d(n)}{n} = \mathcal{O}\bigg(\log G\Big(\frac{1}{3}(\log n)\Big)\bigg)^{1/2}, \tag{12.18}$$

where G: $\mathbb{R}^+ \longrightarrow \mathbb{R}^+$ is some rapidly decaying monotone function that is an inverse to Luo's function F for integral g, $G\big(F(g)\big) = g$. Thus the family $(S^2 \times S^1)_g$ has at least this (12.18) absurdly small $\mathbb{Z}_2$-freedom function.

Many examples of $\mathbb{Z}_2$-freedom in all higher dimensions can be obtained from the families P_g and $(S^2 \times S^1)_g$ by taking products with spheres of appropriate radii. Additional surgeries may be done after taking products to adjust the topology of the families if desired. For an application to quantum codes, the simplest construction is to form $\overline{P}_g = P_g \times S^1_{r(g)}$ where S^1_r is the circle of radius $r(g) = g(\log g)^{-1/2}$.

PROPOSITION 12.4

$\mathbb{Z}_2 - 2$-systole$(\overline{P}_g) = \mathcal{O}(g)$.

PROOF of Proposition 12.4 Let $[\alpha] \in H_2(\overline{P}_g; \mathbb{Z}_2) \cong \big(H_1(P_g; \mathbb{Z}_2) \otimes H_1(S^1_r; \mathbb{Z}_2)\big) \oplus H_2(P_g; \mathbb{Z}_2)$. Since the projection $\overline{P}_g \longrightarrow P_g$ is area nonincreasing, a 2-cycle α must have area $\mathcal{O}(g)$ if its projection in P_g is essential. Suppose for a contradiction that area$(\alpha) < \mathcal{O}(g)$. Then by the above, $[\alpha] \in H_1(P_g; \mathbb{Z}_2) \otimes H_1(S^1_r; \mathbb{Z}_2)$ the first factor. For generic $\theta \in S^1_r, \alpha \cap \text{proj}^{-1}(\theta)$ defines the essential element of $H_1(P_g; \mathbb{Z}_2)$. By Fubini's theorem, area$(\alpha) \geq \mathbb{Z}_2 - 1$-systole $(P_g) \cdot \mathbb{Z}_2 - 1$-systole $(S^1_r) = \mathcal{O}(\log^{1/2} g) \cdot (g \log^{-1/2} g) = \mathcal{O}(g)$, a contradiction. ∎

THEOREM 12.6

The family $\overline{P}_g$ exhibits weak $\mathbb{Z}_2$-systolic $(2,2)$-freedom with freedom function $\mathcal{O}(\log^{1/2})$.

PROOF of Theorem 12.6 This follows immediately from Proposition 12.4 and the fact that vol $(\overline{P}_g) = \mathcal{O}(g^2)(\log g)^{-1/2}$. ∎

Remark: Although the family $\overline{P}_g$ suffices to construct local quantum codes (see Section 12.3) that exceed square-root efficiency, for aesthetic reasons one may wish to achieve $\mathbb{Z}_2$-$(2,2)$-freedom on a family whose underlying manifolds are all diffeomorphic to $S^2 \times S^2$. This would remove the word "weak" in Theorem 12.6, but the price is that the quantitative level of $\mathbb{Z}_2$-freedom may slip to an absurdly small amount as in (12.18). This is what happens if we form $(S^2 \times S^2)_g$ from $(S^2 \times S^1)_g \times S^1_r$ by performing two 1-surgeries on the $\pi_1\big((S^2 \times S^1)_g \times S^1_r\big) \cong \mathbb{Z} \oplus \mathbb{Z}$ generators.

THEOREM 12.7

There is a family of Riemannian metrics on $S^2 \times S^2$, $(S^2 \times S^2)_g$, which exhibits $\mathbb{Z}_2$-$(2,2)$-freedom, with freedom function bounded below by $\big(\log G(\frac{1}{3}\log d)\big)^{1/2}$. There exists a weak family of Riemannian metrics on closed 4-dimensional manifolds $\{(S^2 \times S^2)^h_g\}$ with the $\mathbb{Z}_2$-homology of $S^2 \times S^2$ which exhibit $\mathbb{Z}_2$-$(2,2)$-freedom with freedom function bounded below by $\mathcal{O}(\log d)^{1/2}$.

PROOF of Theorem 12.7 The first statement follows by performing two framed 1-surgeries on each $(S^2 \times S^1)_g \times S^1_r$. The second family is similarly derived from $\overline{P}_g$ by two 1-surgeries on each member. ∎

12.3 Quantum codes from Riemannian manifolds

Let N be a closed Riemannian manifold of dimension d. We assume that the metric on N has been homothetically scaled so that the injectivity radius is at least 1 and the maximum absolute value of any entry in the curvature tensor satisfies $|R^\ell_{ijk}| \leq 1$. Also all first covariant derivatives of curvature are assumed bounded, $|\nabla_h R^\ell_{ijk}| \leq 1$.

THEOREM 12.8

There are constants $0 < d_1, d_2, d_3 < \infty$ depending only on the dimension d so that N admits a piecewise smooth pair (cellulation, dual triangulation) $= (\mathcal{C}, \mathcal{C}^*)$, so that:

(1) The number of cells of all dimensions in both $\mathcal{C}$ and $\mathcal{C}^*$ satisfies: $|\mathcal{C}| + |\mathcal{C}^*| < d_1$ vol(M).

(2) For every p-cell π of $\mathcal{C} \cup \mathcal{C}^*$:

$$p - vol\,(\pi) < d_2.$$

(3) For each cell of $\mathcal{C} \cup \mathcal{C}^*$, both the number of cells in its boundary and the number of cells in its coboundary are less than d_3.

PROOF of Theorem 12.8 The curvative bound implies that there is a fixed ϵ, depending only on dimension d, so that every 4ϵ-ball in M is 1-1 quasi-isometric to the 4ϵ-ball in Euclidean d-space.

Let $V \subset M$ be a maximal set of points so that the ϵ-balls centered at V are disjoint. The 2ϵ-balls centered at V cover M. Roughly, we would like to take $\mathcal{C}$ to be the V-Voronoi cellulation centered at V but there are bothersome (open?) technicalities about estimating p-volumes of p-faces. A simple alternative is to let $\mathcal{C}'$ be the V-Voronoi cellulation, which after perturbing V we may assume to be generic and piecewise smooth, and to let $\mathcal{C}^*_{\text{comb.}}$ be the combinatorial triangulation formally dual to $\mathcal{C}'$. We can construct an embedded, in M, piecewise smooth homeomorph $\mathcal{C}^*$ of $\mathcal{C}^*_{\text{comb.}}$ by inductively constructing the simplices of $\mathcal{C}^*$ by "geodesic sweep out." The vertices of $\mathcal{C}^*$ are the set V. The 1-simplices are the shortest center-connecting arcs for pairs of $\mathcal{C}'$-cells that meet along a $d-1$ face. The 2-simplices (corresponding to triple coincidences of $\mathcal{C}'$-cells of dimension $d-1$) are defined (unnaturally) by picking out one of the three vertices and sweeping a geodesic segment emanating from that vertex and ending on the opposite edge. Proceeding inductively, the totality of the simplices so constructed is $\mathcal{C}^*$. Define $\mathcal{C}_{\text{comb.}}$ to be the combinatorial dual cellulation: barycentrically subdivide $\mathcal{C}^*_{\text{comb.}}$ to obtain $\overline{\mathcal{C}^*}_{\text{comb.}}$. Then the p-cells of $\mathcal{C}_{\text{comb.}}$ are the closed stars of the vertices of $\overline{\mathcal{C}^*}_{\text{comb.}}$.

Again using geodesic sweep-out, $\mathcal{C}_{\text{comb.}}$ may be piecewise smoothly embedded in M; call the result $\mathcal{C}$. Because of the explicit inductive construction of all the cells of $\mathcal{C}^*$ and $\mathcal{C}$, d_2 can be found so that the estimate (2) of Theorem 12.8 holds. Notice that the bound on covariant derivatives of curvature already comes into showing that the 2-simplices, constructed as a family of geodesic arcs, have bounded area.

Now the local combinatorics of $\mathcal{C}$ and $\mathcal{C}^*$ is dictated by that of $\mathcal{C}'$, which is bounded by volume considerations. For v, $v' \in V$, each d-cell D_v of $\mathcal{C}'$ has volume at least $(\frac{1}{1.1})^d$ vol(Euclidean ϵ-ball) and can only share a $(d-1)$-face with another $D_{v'}$ if $\text{dist}(v, v') < 4\epsilon$. The number of such potential $(d-1)$-faces per d-cell is bounded by:

$$\frac{(1.1)^d \text{ vol(Euclidean } 4\epsilon\text{-ball)}}{\left(\frac{1}{1.1}\right)^d \text{ vol(Euclidean } \epsilon\text{-ball)}} = (1.1)^{2d} \cdot 4^d.$$

Similarly the number of $(d-k)$-cells of C is bounded by the binomial coefficient $\binom{\lceil (1.1)^{2d} \rceil \cdot 4^d}{k}$. For appropriate constants, d_1 and d_3 conditions (1) and (3) now follow. ∎

DEFINITION 12.2 *Given a cellulated, triangulated manifold* $(M; \mathcal{C}, \mathcal{C}^*)$, *we may define combinatorial* $\mathbb{Z}_2-$ systoles *and dual systoles as follows: the combinatorial* $\mathbb{Z}_2 - k$-systole *is the fewest number of* k-cells *of* $\mathcal{C}$ *that form an essential cycle in* $H_k(M; \mathbb{Z}_2)$*; and the combinatorial* $\mathbb{Z}_2 - \ell$-systole *is the fewest number of* ℓ-cells *of* $\mathcal{C}^*$ *that form an essential cycle in* $H_\ell(M; \mathbb{Z}_2)$ *(constructed from the dual chains).*

Let us translate the result of Section 12.2 into combinatorics using Theorem 12.8. Recall from Theorem 12.6 the family of Riemannian 4-manifolds $(S^2 \times S^2)_g^h$ with $\mathbb{Z}_2 - (4,2)$-systoles scaling like

$$\left(\mathcal{O}(g^2) \log^{-1/2} g, \mathcal{O}(g)\right)$$

and bounded curvatures. These may be (cellulated, triangulated) in accordance with Theorem 12.8:

$$\left((S^2 \times S^2)_g; \mathcal{C}_g;\ \mathcal{C}_g^*\right).$$

The result are (cellulations, triangulations) with bounded geometry. No cell or simplex has more than d_3 terms in its boundary or coboundary. We obtain the following scalings on $\mathcal{C}_g$ and $\mathcal{C}_g^*$: combinatorial $\mathbb{Z}_2 -$ 2-systole $(\mathcal{C}_g) = \mathcal{O}(g)$ and combinatorial $\mathbb{Z}_2 -$ 2-systole $(\mathcal{C}_g^*) = \mathcal{O}(g)$, and we may bound the number of 2−cells:

$$\mathcal{O}\big(\text{2-cells } (\mathcal{C}_g)\big) \le \mathcal{O}\big(\text{4-cells } (\mathcal{C}_g)\big) \le \mathcal{O}\big(\text{vol}(S^2 \times S^2)_g\big) \le \mathcal{O}(g^2 \log^{-1/2} g).$$

Now, as described in Section 12.0, we may build a d_3-local-parity-check-CSS code Code$_g$ by assigning qubits to the 2-cells of $\mathcal{C}_g$. Since $H_2(\overline{P}_g; \mathbb{Z}_2) \cong \mathbb{Z}_2 \oplus \mathbb{Z}_2$ these are $\left[\mathcal{O}(g^2 \log^{-1/2} g), 2, \mathcal{O}(g)\right]$ codes, i.e., quantum codes that encode 2 qubits into $\mathcal{O}\big(g^2 \log^{-1/2} g\big)$ qubits capable of recovering from $\mathcal{O}(g)$ worst case errors. Introducing the parameter $n = g^2 \log^{-1/2} g$ and $\lfloor \frac{t}{2} \rfloor$ for the number of tolerated worst case errors, we compute the asymptotics of the codes Codes_g to be

$$t = \mathcal{O}\big(n^{1/2} \log^{1/2} n\big).$$

This completes the proof of Theorem 12.3. □

Appendix

The purpose of this appendix is to prove the following result:

THEOREM 12.9
Suppose Σ_g is a closed surface with a hyperbolic metric of injectivity radius r. There exists a computable constant $C(g, r)$ so that each isometry of Σ_g is isotopic to a composition of positive and negative Dehn-twists $D_{c_1}^{\pm 1} \dots D_{c_k}^{\pm 1}$ where $k \le C(g, r)$ and the length $l(c_i)$ of c_i is at most $C(g, r)$ for each i.

As a by-product of the proof, we also obtain the following result, which may be of some independent interest in view of the recent work on symplectic 4-manifolds.

Call a self-homeomorphism of the surface *positive* if it is isotopic to a composition of positive Dehn-twists.

THEOREM 12.10
Suppose $\Sigma_{g,n}$ is a compact orientable surface of genus g with n boundary components. Let $\{a_1, \dots, a_{3g-3+2n}\}$ be a 3-holed sphere decomposition of the surface where $\partial\Sigma_{g,n} = a_{3g-2+n} \cup \dots \cup a_{3g-3+2n}$. Then each orientation preserving homeomorphism of the surface that is the identity map on $\partial\Sigma_{g,n}$ is isotopic to a composition qp where p is positive and q is a composition of negative Dehn-twists on a_i's.

The basic idea of the proof of Theorem 12.9 suggested by M. Freedman is as follows. Let f be an isometry of the surface. Choose a surface filling system of simple geodesics $\{s_1, \dots, s_k\}$ whose lengths are bounded (in terms of r and g). Since the lengths of s_i and $f(s_j)$ are bounded, the intersection numbers between any two members of $\{s_1, \ldots, s_k, f(s_1), \ldots, f(s_k)\}$ are bounded. Now the proof of Lickorish's theorem in [17] is constructive and depends only on the intersection numbers between simple loops. Thus, one produces a bounded number of simple loops of bounded lengths so that the composition of positive or negative Dehn-twists on them sends s_i to $f(s_i)$. This shows that f is isotopic to the composition.

The proof below follows Freedman's sketch. We shall choose the surface filling system to be of the form $\{a_1, \ldots, a_{3g-3}, b_1, \ldots, b_{3g-3}\}$ where $\{a_i\}$ forms a 3-holed sphere decomposition of the surface so that $l(a_i) \leq 26(g-1)$ (Ber's theorem) and the b_is have bounded lengths so that $b_i \cap a_j = \emptyset$ for $j \neq i$. Then we establish a controlled version of Lickorish's lemma (Lemma 2 in [17]) by estimating the lengths of loops involved in the Dehn-twists.

We shall use the following notations and conventions. Surfaces are oriented. If a is a simple loop on a surface, D_a denotes the positive Dehn-twist along a and $l(a)$ denotes the length of the geodesic isotopic to a. Two isotopic simple loops a and b will be denoted by $a \cong b$. Given two simple loops a, b, *their geometric intersection number*, denoted by $I(a, b)$, is $\min\{|a' \cap b'| \mid a' \cong a, b' \cong b\}$. It is well known that if a, b are two distinct simple geodesics, then $|a \cap b| = I(a, b)$. We use $|a \cap b| = 2_0$ to denote two simple loops a, b so that $I(a, b) = |a \cap b| = 2$ and their algebraic intersection number is zero.

To prove Theorem 12.9, we begin with the following.

PROPOSITION 12.5

Suppose a and b are homotopically nontrivial simple loops in a hyperbolic surface of injectivity radius r. Then,

(a) (Thurston). $I(a, b) \leq \frac{4}{\pi r^2} l(a) l(b)$.

(b) $l(D_a(b)) \leq I(a, b) l(a) + l(b)$.

(c) For each integer n, $\frac{\pi r^2 |n| I(a,b)}{4 l(b)} \leq l(D_a^n(b)) \leq |n| I(a, b) l(a) + l(b)$.

(d) If $|a \cap b| \geq 3$ or $|a \cap b| = 2$ so that the two points of intersection have the same intersection signs, then there exists a simple loop c so that $l(c) \leq l(a) + l(b)$, $|D_c(b) \cap a| < |b \cap a|$ and $l(D_c(b)) \leq 2l(a) + l(b)$.

(e) There exists a sequence of simple loops $c_1, \ldots, c_k$ so that $k \leq |a \cap b|$, $l(c_i) \leq (2i-1) l(a) + l(b)$ for each i and $D_{c_k} \ldots D_{c_1}(b)$ is either

disjoint from a, or intersects a at one point, or intersects a at two points of different signs.

PROOF of Proposition 12.5 Part (a) is essentially in [10], p. 54, Lemma 2. We produce a slightly different proof so that the coefficient is $\frac{4}{\pi r^2}$. Without loss of generality, we may assume that both a and b are simple geodesics. Construct a flat torus as the metric product of two geodesics a and b. The area of the torus is $l(a)l(b)$. Each intersection point of a with b gives a point p in the torus. Now the flat distance between any two of these points p_s is at least the injectivity radius r (otherwise there would be Whitney discs for $a \cup b$). Thus the flat disks of radius $r/2$ around these p_s are pairwise disjoint. This shows that the sum of the areas of these disks is at most $l(a)l(b)$ which is the Thurston's inequality.

To see part (b), we note that the Dehn-twisted loop $D_a(b)$ is obtained by taking $I(a,b)$ many parallel copies of a and resolving all the intersection points between b and the parallel copies (from a to b). Thus the inequality follows.

Part (c) follows from parts (a) and (b). Note that we have used the fact that $I(D_a^n(b), b) = |n| I(a,b)$ (see for instance [18] for a proof, or one also can check directly that there are no Whitney disks for $D_a^n(b) \cup b$).

Part (d) is essentially in Lemma 2 of [17]. Our minor observation is that one can always choose a positive Dehn-twist D_c to achieve the result.

We need to consider two cases.

Case 1. There exist two intersection points x, $y \in a \cap b$ adjacent along in a which have the same intersection signs (see Figure 1). Then the curve c as shown in Figure 1 (with the right-hand orientation on the surface) satisfies all conditions in the part (d). If the surface is left-hand oriented, take $D_c(b)$ to be the loop c.

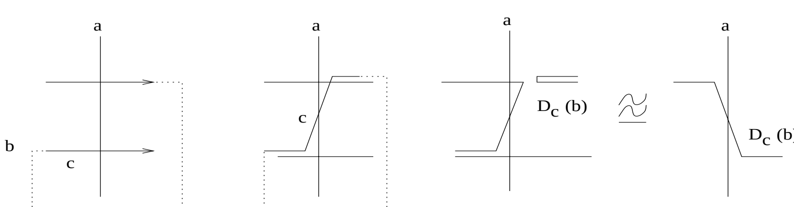

FIGURE 12.3
Right-hand orientation on the plane.

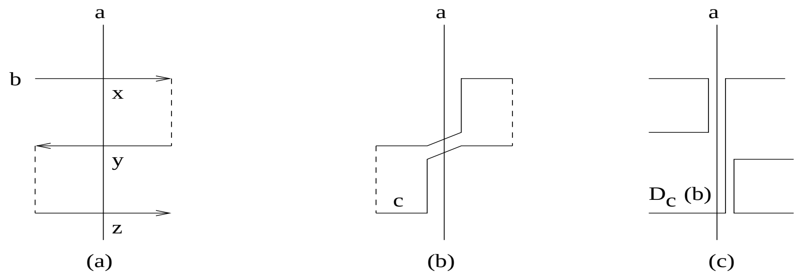

FIGURE 12.4
Right-hand orientation on the plane.

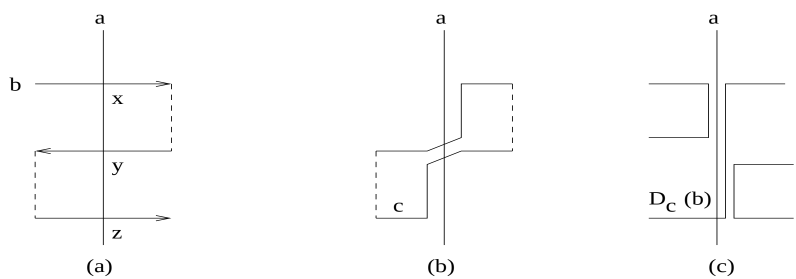

FIGURE 12.5
Left-hand orientation on the plane.

Case 2. Suppose any pair of adjacent intersection points in $a \cap b$ has different intersection signs. Then $|a \cap b| \geq 3$. Take three intersection points $x, y, z \in a \cap b$ so that x, y and y, z are adjacent in a. Their intersection signs alternate. Fix an orientation on b so that the arc from x to y in b does not contain z as shown in Figure 2. If the surface Σ is right-hand oriented as in Figure 2, take c as in Figure 2(b). Then $D_c(b)$ is shown in Figure 2(c). If the surface has the left-hand orientation, then take c as shown in Figure 3(b). The loop $D_c(b)$ is shown in Figure 3(c). One checks easily that the simple loop c satisfies all the conditions.

Part (e) follows from part (d) by induction on $|a \cap b|$. ∎

We shall also need the following well-known lemma in order to deal with disjoint loops and loops intersecting at one point.

LEMMA 12.2

Suppose a and b are two simple loops intersecting transversely at one point. Then,

(a) $D_aD_b(a) \cong a$,

(b) $(D_aD_bD_a)^2$ *sends a to a, b to b and reverses the orientations on both a and b.*

See [2] and [17] for a proof, or one can check it directly. Note that $(D_aD_bD_a)^2$ is the hyper-elliptic involution on the 1-holed torus containing both a and b.

We first give a proof of Theorem 12.10. The proof of Theorem 12.9 follows by making length estimate at each stage of the proof of Theorem 12.10.

PROOF of Theorem 12.10 Let f be an orientation-preserving homeomorphism of $\Sigma_{g,n}$, which is the identity map on the boundary. We shall show that there exists a composition p of positive Dehn-twists so that for each i, $pf^{-1}|_{a_i} = \text{id}$. It follows that pf^{-1} is a product of Dehn-twists on a_i's.

We prove the theorem by induction on the norm $|\Sigma_{g,n}| = 3g - 3 + n$ of the surface (the norm is the complex dimension of the Teichmuller space of complex structures with punctured ends on the interior of the surface). The basic property of the norm is that if Σ' is an incompressible subsurface that is not homotopic to $\Sigma_{g,n}$, then the norm of Σ' is strictly smaller than the norm of $\Sigma_{g,n}$. For simplicity, we assume that the Euler characteristic of the surface is negative (though the proof below also works for the torus).

If the norm of a surface is zero, then the surface is the 3-holed sphere. The theorem is known to hold in this case (see [9]).

If the norm of the surface is positive, we pick a nonboundary component, say a_1, of the 3-holed sphere decomposition as follows. If the genus of the surface $\Sigma_{g,n}$ is positive, a_1 is a nonseparating loop. By Proposition 12.5(e) applied to $a = a_1$ and $b = f^{-1}(a_1)$, we find a sequence of simple loops $c_1, \ldots, c_k$, $k \leq I(a,b)$ so that $a_1' = D_{c_k} \ldots D_{c_1} f^{-1}(a_1)$ satisfies: either $a_1' \cap a_1 = \emptyset$, or $|a_1' \cap a_1| = I(a_1', a_1) = 1$, or $|a_1' \cap a_1| = 2_0$. There are two cases we need to consider: (1) both a_1 and a_1' are separating loops, and (2) both of them are nonseparating.

In the first case, by the choice of a_1, the genus of the surface is zero. First, $I(a_1, a_1') = 1$ cannot occur due to homological reason. Second, since the homeomorphism $D_{c_k} \dots D_{c_1} f^{-1}$ is the identity map on the nonempty boundary $\partial\Sigma_{0,n}$, it follows that $I(a_1, a_1') = 2$ is also impossible and a_1' is actually isotopic to a_1. After composing with an isotopy, we may assume that $D_{c_k} \dots D_{c_1} f^{-1}|_{a_1}$ is the identity map. Now cut the surface open along a_1 to obtain two subsurfaces of smaller norms. Each of these subsurfaces is stablized under $D_{c_k} \dots D_{c_1} f^{-1}$. Thus the induction hypothesis applies and we conclude the proof in this case.

In the second case that both a_1 and a_1' are nonseparating, then either $|a_1' \cap a_1| = 1$, or there exists a third curve c so that c transversely intersects each of a_1 and a_1' in one point. By Lemma 12.2(a), one of the product h of positive Dehn-twists $D_{a_1'} D_{a_1}$, or $D_c D_{a_1} D_{a_1'} D_c$, will send a_1' to a_1. If the homeomorphism $h D_{c_k} \dots D_{c_1} f^{-1}$ sends a_1 to a_1 reversing the orientation, by Lemma 12.2(b), we may use six more positive Dehn-twists (on a_1, a_1', or c, a_1) to correct the orientation. Thus, we have constructed a composition of positive Dehn-twists $D_{c_m} \dots D_{c_1} f^{-1}$ so that it is the identity map on a_1 and $m \le I(a, b) + 10$. Now cut the surface open along a_1 and use the induction hypothesis. The result follows. ∎

We note that the proof fails if we do not choose a_1 to be a nonseparating simple loop in the case the surface is closed of positive genus.

Now we prove Theorem 12.9 by making length estimate on each step above.

PROOF of Theorem 12.9 Let f be an isometry of a hyperbolic closed surface $\Sigma_g = \Sigma_{g,0}$.

We begin with the following result, which gives bound on the lengths of a_i's and c used in the proof of Theorem 12.10.

PROPOSITION 12.6

Suppose Σ_g is a hyperbolic surface of injectivity radius r.

(a) (Bers) There exists a 3-holed sphere decomposition $\{a_1, \dots, a_{3g-3}\}$ of the surface so that $l(a_i) \le 26(g-1)$.

(b) If a and b are two nonseparating simple geodesics in a compact hyperbolic surface Σ which is a totally geodesic subsurface in $\Sigma_{g,n}$ so that either $I(a, b) = 0$ or $|a \cap b| = 2_0$, then there exists a simple geodesic c in Σ so that $I(c, a) = I(c, b) = 1$ and $l(c) \le \frac{8(g-1)r}{sinhr} + 8r$.

PROOF of Proposition 12.6 See Buser [6], p. 123 for a proof of part (a).

To see part (b), we first note that there are simple loops x so that $I(x,a) = I(x,b) = 1$ by the assumption on a and b. Let c be the shortest simple loop in Σ satisfying $I(c,a) = I(c,b) = 1$. We shall estimate the length of c as follows. Let $N = \lfloor \frac{l(c)}{2r} \rfloor$ be the largest integer smaller than $\frac{l(c)}{2r}$. Let $P_1 = a \cap c$, P_2, ..., P_N be N points in c so that their distances along c are $d(P_1, P_{i+1}) = 2ri$ and $d(P_i, P_{i+1}) = 2r$. Let B_i be the disc of radius r centered at P_i and B_k be the ball containing $c \cap b$. Then the shortest length property of c shows that the intersections of the interior $\mathrm{int}(B_i) \cap \mathrm{int}(B_j)$ is empty if $1 \le i < j < k$ or $k < i < j \le N$. Thus the sum of the areas of the $N-2$ balls $B_2, \dots, B_{k-1}, B_{k+1}, \dots, B_N$ is at most twice the area of the surface $\Sigma_{g,0}$. This gives the estimate required. ∎

Fix a 3-holed sphere decomposition $\{a_1, \dots, a_{3g-3}\}$ of the hyperbolic surface so that $l(a_i) \le 26(g-1)$. We may assume that the loops a_i are so labeled that $a_1, a_2, \dots, a_g$ are nonseparating loops and the rest are separating.

We now show that there exists a computable constant $C' = C'(g,r)$ so that any orientation-preserving isometry f of the hyperbolic surface Σ_g is isotopic to a product qp where q is a product of positive or negative Dehn-twists on a_is and p is a product of at most $C'(g,r)$ many positive Dehn-twists on curves of lengths at most $C'(g,r)$.

We now rerun the constructive proof of Theorem 12.10 by estimating the lengths of loops involved. To begin with, we take $a = a_1$ and $b = f^{-1}(a_1)$ of lengths at most $26g$. By Thurston's inequality, their intersection number $I(a,b)$ is at most $\frac{52^2 g^2}{\pi r^2}$. By Propositions 12.5(e), 12.6(b) and the proof of Theorem 12.10, we produce a finite set of simple loops $\{c_1, \dots, c_k\}$ so that $k \le I(a,b) + 10$, the lengths of c_i are bounded in g, r and $f_1 = D_{c_k} \dots D_{c_1} f^{-1}$ is the identity map on a_1. Now we take $a = a_2$ and $b = f_1(a_2)$ and run the same constructive proof as above in the totally geodesic subsurface $\Sigma_{g-1,2}$ obtained by cutting Σ_g open along a_1. In order for the proof to work, we need to see that the length of b is bounded. Indeed, Proposition 12.5(b) gives the estimate of $l(b)$ in terms of $l(c_i)$, $l(a_2)$, and g, r (here we estimate the intersection number $I(c_i, x)$ in terms of the lengths by Thuston's inequality). Thus, we construct a finite set of simple loops $d_1, \dots, d_m$ so that m and $l(d_i)$ are bounded in g, r, $d_i \cap a_1 = \emptyset$, and $D_{d_m} \dots D_{d_1} f_1^{-1}$ is the identity map on $a_1 \cup a_2$. Inductively, we produce the required positive homeomorphism p.

We remark that if the injectivity radius r is at least $\log 2$, then the number $C'(g,r)$ that we obtained is at least $g^{g^{g^{\cdots g}}}$ (there are $3g-3$ many exponents) in magnitude.

As a consequence, we obtain the following expression for the homeomorphism $p^{-1}f = D_{a_1}^{n_1} \dots D_{a_{3g-3}}^{n_{3g-3}}$. It remains to show that the exponents n_is are bounded. To this end, for each index i, we pick a geodesic loop b_i which is disjoint from all a_j's for $j \neq i$ and b_i intersects a_i at one point or two points of different signs. A simple calculation involving right-angled hyperbolic hexagon shows that we can choose these b_i to have lengths at most $182(g-1) - \log(r/4)$. Thus the lengths of curve $p^{-1}f(b_i)$ is bounded (in terms of g and r). By Proposition 12.5(d), the growth of the lengths of loops $D_{a_i}^n(b_i)$ is linear in $|n|$ if $|n|$ is large. Thus we obtain an estimate on the absolute value of the exponents $|n_i|$. This finishes the proof. □

References

[1] M. Berger, À l'ombre de Loewner, *Ann. Sci. École Norm. Sup. (4)* **5** (1972) 241–260.

[2] J. Birman, Mapping class groups of surfaces, in *Braids*, J. Birman and A. Libgober, eds., Proceedings of a summer research conference, Santa Cruz, CA, 13–16 July 1986, Contemp. Math. **78** (Providence, RI: AMS 1988) 13–44.

[3] I. Babenko and M. Katz, Systolic freedom of orientable manifolds, *Ann. Sci. École Norm. Sup. (4)* **31** (1998) 787–809.

[4] I. K. Babenko, M. G. Katz and A. I. Suciu, Volumes, middle-dimensional systoles, and Whitehead products, *Math. Res. Lett.* **5** (1998) 461–471.

[5] W. Browder, *Surgery on Simply-Connected Manifolds*, *Ergebnisse der Mathematik und ihrer Grenzgebiete* **65** (New York: Springer-Verlag 1972).

[6] P. Buser, *Geometry and Spectra of Compact Riemann Surfaces*, Progress in Mathematics **106** (Boston, MA: Birkhäuser 1992).

[7] P. Buser, A note on the isoperimetric constant, *Ann. Sci. École Norm. Sup. (4)* **15** (1982) 213–230.

[8] A. R. Calderbank and P. W. Shor, Good quantum error-correcting codes exist, *Phys. Rev.* A **54** (1996) 1098–1105.

[9] M. Dehn, *Papers on Group Theory and Topology*, translated and edited by J. Stillwell (New York: Springer-Verlag 1987).

[10] A. Fathi, F. Laudenbach and V. Poénaru, *Travaux de Thurston sur les surfaces*, *Astérisque* **66–67** (1979).

[11] M. H. Freedman, $\mathbb{Z}_2$-systolic-freedom, in *Proceedings of the Kirbyfest*, J. Hass and M. Scharlemann, Eds., Geometry and Topology Monographs **2** (Coventry, UK: Geometry and Topology Publications 1999) 113–123.

[12] H. Federer, *Geometric Measure Theory*, *Die Grundlehren der mathematischen Wissenschaften* **153** (New York: Springer-Verlag 1969).

[13] D. Gottesman, *Stabilizer codes and quantum error correction*, `quant-ph/9705052`; CalTech Ph.D. Thesis (1997).

[14] M. Gromov, Systoles and intersystolic inequalities, in *Actes de la Table Ronde de Géométrie Différentielle*, Luminy, 1992, *Sémin. Congr.* **1** (Paris: Soc. Math. France 1996) 291–362.

[15] A. Yu. Kitaev, Fault-tolerant quantum computation by anyons, `quant-ph/9707021`.

[16] M. Katz and A. I. Suciu, Volume of Riemannian manifolds, geometric inequalities, and homotopy theory, in *Tel Aviv Topology Conference: Rothenberg Festschrift (1998)*, Contemp. Math. **231** (Providence, RI: AMS 1999) 113–136.

[17] R. Lickorish, A representation of oriented combinatorial 3-manifolds, *Ann. Math.* **72** (1962) 531–540.

[18] F. Luo, Multiplication of simple loops on surfaces, Rutgers preprint (1999).

[19] C. Pittet, Systoles on $S^1 \times S^n$, *Diff. Geom. Appl.* **7** (1997) 139–142.

[20] P. Schmutz Schaller, Extremal Riemann surfaces with a large number of systoles, in *Extremal Riemann Surfaces*, J. R. Quine and P. Sarnak, Eds., San Francisco, CA, 1995, Contemp. Math. **201** (Providence, RI: AMS 1997) 9–19.

[21] P. W. Shor, Scheme for reducing decoherence in quantum computer memory, *Phys. Rev.* A **52** (1995) R2493–R2496.

[22] E. Spanier, *Algebraic Topology* (New York: McGraw-Hill 1966).

[23] A. M. Steane, Error correcting codes in quantum theory, *Phys. Rev. Lett.* **77** (1996) 793–797.