Local maxima of the systole function

Maxime Fortier Bourque Kasra Rafi

December 18, 2019

Abstract

We construct a sequence of closed hyperbolic surfaces that are local maxima for the systole function in their respective moduli spaces. Their systole is arbitrarily large and the number of examples grows rapidly with the genus. More precisely, for every $n \geq 3$ there is some positive number L_n (growing roughly linearly in n) such that the number of local maxima of the systole function in genus g with systole equal to L_n grows super-exponentially in g along an arithmetic sequence of step size n. Many of these surfaces have no orientation-preserving isometries other than the identity and are the first examples of local maxima with this property.

1 Introduction

The systole of a hyperbolic surface is the length of any of its shortest closed geodesics. For any $g \geq 2$, this defines a continuous function $\mathrm{sys} : \mathcal{T}_g \to \mathbb{R}_+$ on the Teichmüller space of closed hyperbolic surfaces of genus g which is invariant under the action of the mapping class group, hence descends to a continuous function on the moduli space $\mathcal{M}_g$.

By Mumford's compactness criterion [42], the thick part $\{x \in \mathcal{M}_g \,\big|\, \mathrm{sys}(x) \geq \varepsilon\}$ of moduli space is compact for any $\varepsilon > 0$. Therefore, the systole function attains a global maximum on each moduli space. The precise value of the maximum is unknown in general; the best bounds known to date are

$$\frac{4}{3} \leq \limsup_{g\to\infty} \frac{\max\{\mathrm{sys}(x) \,\big|\, x \in \mathcal{M}_g\}}{\log g} \leq 2 \tag{1.1}$$

M. Fortier Bourque: School of Mathematics and Statistics, University of Glasgow, University Place, Glasgow, United Kingdom, G12 8QQ; e-mail: maxime.fortier-bourque@glasgow.ac.uk

K. Rafi: Department of Mathematics, University of Toronto, 40 St. George Street, Toronto, ON, Canada M5S 2E4; e-mail: rafi@math.toronto.edu

Mathematics Subject Classification (2010): 30F60, 32G15

where the upper bound is a standard area argument, while the lower bound comes from a construction of Buser and Sarnak [13, p.45].

For closed orientable surfaces, the only known global maximizer of the systole is the Bolza curve in genus 2, as determined by Jenni [36]. More maximizers are known if we allow punctures [50, 1] (principal congruence covers of the modular curve are maximizers) or surfaces that have non-empty boundary or that are non-orientable [26]. To attack the problem in general, Schmutz Schaller initiated a systematic study of the systole function and its local maxima in [49], where he found necessary and sufficient conditions for a surface to be a local maximizer and constructed several examples. The criterion in question is analogous to a theorem of Voronoi on the Hermite invariant of Euclidean lattices. Bavard cast both of these results into a more general framework [4, 5]. Further progress on the systole of hyperbolic surfaces and its singularities was made by Akrout, Casamayou-Boucau, Gendulphe, and others [2, 14, 15, 27]. We refer the reader to [52] and [43] for surveys on the systole of hyperbolic surfaces and related topics.

The goal of this paper is to show that the number of local maxima of the systole function grows super-exponentially with the genus. The local maximizers we construct have arbitrarily large systole and many of them have no orientation-preserving isometries besides the identity. Prior to this work, there was only one infinite sequence of local maximizers known among closed surfaces, with systole bounded above by 5.634 [49]. A finite number of additional examples were discovered in [51, 29, 32, 15]. All the previously known examples of local or global maxima have large isometry groups: the quotient of any of these surfaces by its group of isometries is a hyperbolic polygon with a small number of sides. Schmutz Schaller anticipated the existence of local maximizers with no symmetries [49, p.565] but he was not able to find any [51, p.437]. The reason why so few examples of local maxima were known before is that proving that a surface is extremal is quite delicate. Having a large isometry group simplifies the problem considerably. For instance, if the quotient of a surface by its group of isometries is a triangle, then one of the two conditions to be a local maximum (namely, eutaxy) comes for free [5, Corollary 1.3]. We do not manage to get rid of all the symmetries in our construction as each of our surfaces admits an orientation-reversing involution, but we obtain many examples with isometry group of size 2. Our main result can be summarized as follows (see Theorem 6.5 for a more precise version):

Theorem 1.1. *For every $n \geq 3$, there exists some $L_n > 0$ such that the number of local maxima of the systole function in $\mathcal{M}_g$ without any non-trivial orientation-preserving isometry and systole equal to L_n grows super-exponentially along an arithmetic sequence of genera g.*

We also obtain some local maxima with more symmetries: the Bolza curve fits naturally in the sequence (corresponding to L_1) and we get one or two local maxima with non-

trivial automorphism group at height $L_2 \approx 5.909$ in every genus $g \geq 13$. The value L_n (defined implicitly in Lemma 2.5 and explicitly in Remark 2.18) that the systole function takes at the other local maxima tends to infinity roughly linearly with n (see Lemma 6.1). For $n \geq 3$, we do not get local maxima in every genus because $2(g-1)$ has to be divisible by n, and has to be sufficiently large (see Theorem 6.5). Thus, we miss all the genera that are equal to a prime number plus one, among others.

Motivation

Akrout [2] proved that sys is a topological Morse function on $\mathcal{T}_g$, following partial results by Schmutz Schaller [53]. This implies that in theory, one could compute topological invariants of $\mathcal{M}_g$ by finding the Morse singularities of sys and their indices. For example, the orbifold Euler characteristic of $\mathcal{M}_g$ is given by the formula

$$\chi(\mathcal{M}_g) = \sum_{x \in \mathcal{C}} \frac{(-1)^{\mathrm{ind}(x)}}{|\operatorname{Aut}(x)|}$$

where $\mathcal{C}$ is a set of representatives of the critical points of sys in $\mathcal{T}_g$ under the action of the mapping class group, $\mathrm{ind}(x)$ is the Morse index of sys at x, and $\operatorname{Aut}(x)$ is the group of automorphisms of x [2, 53].

One could further try to compute—or say something about—the rational homology groups of $\mathcal{M}_g$ using the systole function. In this direction, Bestvina expressed in [6] the hope that the systole function should be "simplest possible" in the sense that the number critical points of index j for the function $-\mathrm{sys} : \mathcal{M}_g \to (-\infty, 0)$ should not depend on g once the latter is large enough. This would give an alternate proof that the jth rational homology group of $\mathcal{M}_g$ becomes independent of g once the latter is large enough [35]. Theorem 1.1 shows that Bestvina's hope is spectacularly false for $j = 0$. And this is probably just the tip of the iceberg—there are likely many more local maxima than the particular ones constructed here.

Luckily, better approaches to study the topology of moduli space are known: the orbifold Euler characteristic of $\mathcal{M}_g$ was computed by Harer and Zagier [33], its virtual cohomological dimension by Harer [34], and its rational homology in the stable range by Madsen and Weiss [40].

Another reason to study local maxima of the systole function is that the global maximum is among them. One could hope to increase the lower bound in (1.1) by stumbling upon local maxima that are near the top of moduli space. Unfortunately, the height of our local maxima grows at most like $\log \log g$ rather than $\log g$. Perhaps one could reduce their genus by using a similar trick as in [44], although we have not explored this yet.

A general goal we have is to understand the shape of moduli space from the point of view of the systole function. For example, Mirzakhani asked whether moduli space has

"long fingers". We may define the *finger* associated with a local maximum x to be the component F of the superlevel set $\{y \in \mathcal{M}_g \,\big|\, \mathrm{sys}(y) > L\}$ containing x, where L is the smallest positive number such that F does not contain any other critical points than x. The *length* of the finger F is then $\mathrm{sys}(x) - L$. In other words, how large can the total variation of the systole function be between a local maximum and the nearest critical point? We do not know the answer, but the examples from Theorem 1.1 provide a place to start. Other works studying the shape of moduli space in relation to the systole include [23, 21, 39, 16, 46, 24, 3, 19, 41].

Proof outline

As explained in [49] and [4], showing that a surface x is a local maximum of the systole function consists in three steps:

1. finding the set $\mathcal{S}$ of systoles of x;
2. showing that x is *perfect*, i.e., that the differential of the vector of lengths of the curves in $\mathcal{S}$—a function on Teichmüller space—is injective at x;
3. proving that x is *eutactic*, i.e., that under every non-trivial infinitesimal deformation of x, at least one of the curves in $\mathcal{S}$ shrinks, meaning that its length has strictly negative derivative in that direction.

The idea of our construction is to glue surfaces out of basic blocks that we call *rings* according to the combinatorics of a graph. If the graph has sufficiently large girth, then the systoles are constrained in pairs of transverse rings (Theorem 2.16). The rings themselves have a large number of symmetries which we use to identify the systoles (Proposition 2.3). The idea of modelling surfaces on graphs is of course not new. It goes back to at least [11] where Buser constructed surfaces with arbitrarily large systole by gluing pairs of pants along cubic graphs. See also [10, 17, 18, 16, 46, 3, 44, 45, 25] for other applications of this idea.

The main tool we need for step (2) is the famous cosine formula of Wolpert and Kerckhoff for the variation of length along twist deformations [54, 38]. We then use the Gershgorin circle theorem to prove that the differential of lengths has full rank, after estimating its entries with respect to a particular basis. The question of which finite sets F of curves are such that their lengths define a global embedding of Teichmüller space into $\mathbb{R}_+^F$ is closely related and classical [48, 30, 31][22, p.287].

The proof of step (3) is easier, but still novel. Whereas Schmutz Schaller relied heavily on symmetries to prove this step in [49] and [51], we manage with only partial symmetries. That is, all we need is that our surfaces are covered by copies of a well-understood chunk, which is itself a union of rings. The rings we use to assemble the surfaces need to

be carefully chosen for this part to work; we went through a few iterations before finding the right candidate.

We then show that any orientation-preserving isometry between the resulting surfaces is induced by a graph isomorphism (Theorem 5.3). Since the number of graphs with large girth and trivial automorphism group grows super-exponentially (see Section 6), we obtain the desired lower bound for the number of local maxima of the systole function without non-trivial orientation-preserving isometries. The fact that our surfaces all admit an orientation-reversing involution is due to the symmetry of the rings and how we glue them. In principle, similar techniques applied with different building blocks might yield totally asymmetric local maxima, but we have not found a suitable building block yet.

Organization

The paper is organized as follows. Section 2 is devoted to the construction of the surfaces and finding their systoles. Steps (2) and (3) of the above program are carried out in Sections 3 and 4 respectively. In Section 5, we show that any orientation-preserving isometry between our surfaces is induced by a graph isomorphism. Finally, in Section 6 we put together estimates on the number of asymmetric graphs of large girth, thereby completing the proof of Theorem 1.1.

Acknowledgements. We thank Robert Young for suggesting the use of the Gershgorin circle theorem and Dmitri Gekhtman for pointing out Wolpert's length-twist duality, which together lead to the proof of step (2) (Theorem 3.9). We also thank Curt McMullen for comments on an earlier draft as well as the anonymous referee for their useful suggestions. MFB and KR were partially supported by Discovery Grants from the Natural Sciences and Engineering Research Council of Canada (RGPIN 06768 and 06486 respectively).

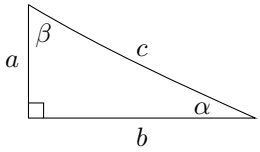
2 The construction

In this section, we construct a highly symmetric surface $R(n,t)$ of genus 1 with $2n$ boundary components of length $4t$ each, for any $n \geq 1$ and any $t > 0$. We then fix a specific value of t for each n and build closed surfaces out of pieces isometric to $R(n,t_n)$.

2.1 Trigonometry

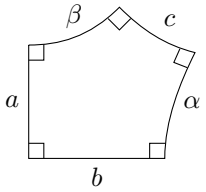
We first gather some trigonometric formulas here for use throughout the paper. See e.g. [12, p.454] for reference.

Right triangles



$$\cosh c = \cosh a \cosh b \tag{2.1}$$
$$\cos \beta = \cosh b \sin \alpha \tag{2.2}$$

Right-angled pentagons



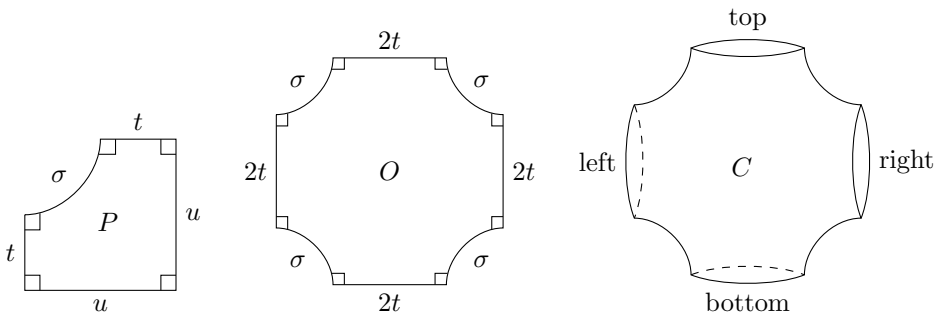
$$\cosh c = \sinh a \sinh b \tag{2.3}$$
$$\cosh c = \coth \alpha \coth \beta \tag{2.4}$$

2.2 The cross

We start with a right-angled pentagon $P = P(t)$ with two non-adjacent sides of length $t > 0$. Let σ be the side between those of length t and let u be the length of each of the other two sides. We have

$$\cosh \sigma(t) = \coth^2 t = \sinh^2 u(t) \tag{2.5}$$

by Equations (2.3) and (2.4). Reflect P across the two sides of length u and the vertex opposite to σ to obtain a right-angled octagon $O = O(t)$ with side lengths alternating between $2t$ and σ. Double O across the sides of length σ to form a four-holed sphere $C = C(t)$ that we call a *cross*. Each of the four boundary geodesics of C has length $4t$. We refer to them as the left, right, top and bottom boundaries of C following Figure 1. Similarly, the cross has a front and a back.



(a) The pentagon (b) The octagon (c) The cross

Figure 1: The cross C is made with two octagons, each assembled from four pentagons

We note in passing that C is an orbifold cover of a quadrilateral $Q = Q(t)$ with three right angles, one angle equal to $\pi/4$, one side of length t and one side of length $\sigma/2$ obtained by cutting P along the median between σ and the opposite vertex. The closed surfaces we construct in the end are also orbifold covers of Q, although not regular covers in general.

2.3 The ring

Let $n \geq 1$ be an integer. We take a string of n crosses $C_1, C_2, \ldots, C_n$ where the right boundary of C_j is glued to the left boundary of C_{j+1} without twist for $j = 1, \ldots, (n-1)$. Finally, the left boundary of C_1 is glued to the right boundary of C_n with a half twist (see Figure 2). The resulting surface $R = R(n, t)$ is called a *ring*. It is a surface of genus one with $2n$ boundary components.

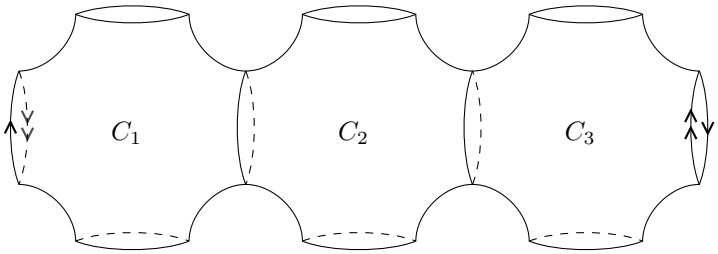


Figure 2: The ring R is a string of n crosses with its ends glued by a half twist

There is an alternative description of the ring which is useful for drawing pictures so that no part of the ring is hidden. Take a strip of $2n$ octagons $O_1, \ldots O_{2n}$ with the right side of each glued to the left side of the next and the right side of O_{2n} glued to the left side of O_1, forming a topological annulus $A = A(n, t)$. Then the top left and top right sides of O_j are glued to the bottom left and bottom right sides of O_{n+j} respectively for $j = 1, \ldots, 2n$ in order to form R, where indices are taken modulo $2n$ (see Figure 3).

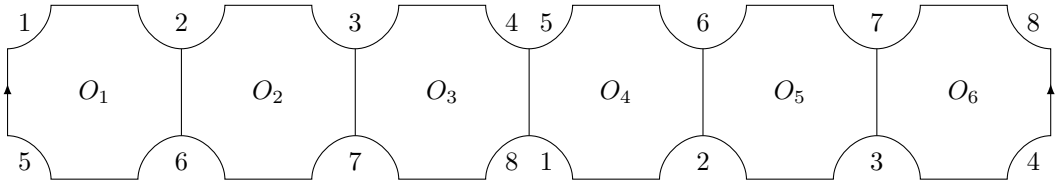


Figure 3: The ring R is also a strip of $2n$ octagons with its left and right sides glued and the segments labelled σ identified in pairs in the pattern shown

In other words, the sides of the annulus A labelled σ are glued in pairs by a glide reflection that reflects across the core geodesic e of A (the horizontal axis of symmetry

in Figure 3) and translates halfway around e. The union of the octagons O_j and O_{n+j} is equal to the cross C_j from the previous description.

2.4 Geodesics in the ring

Following Schmutz Schaller, we will often use the same symbol for the name of a curve and its length. The closed geodesics separating adjacent crosses in the ring are called *f-curves*. More precisely, for each j from 1 to n, we let f_j be the left boundary of C_j. Each f-curve has length $4t$. The geodesic that runs along the horizontal axis of symmetry of all the crosses is called e, which has length $4n \cdot u$ or

$$e = 4n \operatorname{arcsinh}(\coth t). \tag{2.6}$$

The next geodesics of interest are called *a-curves* and *b-curves*. For each j from 1 to $2n$, let a_j be the geodesic joining the bottom of the left side of O_j and the top of the left side of O_{n+j} (these two points are identified in R) and is otherwise disjoint from the seams and the octagons $O_{n+j}, O_{n+j+1}, ..., O_{n+j+(n-1)}$, where indices are taken modulo $2n$ (see Figure 4). Similarly, we let $b_j = \rho_{f_j}(a_j)$ where $\rho_{f_j} : R \to R$ is the reflection across the geodesic f_j. By symmetry, all the a-curves and b-curves have the same length which we denote by a.

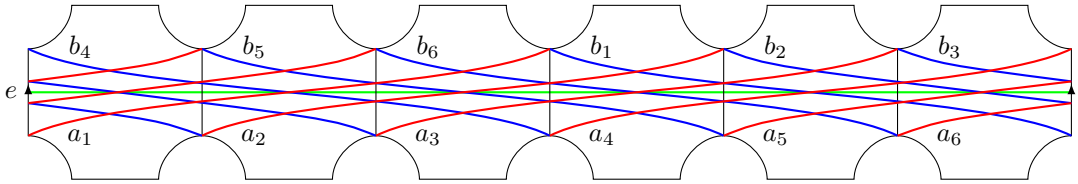


Figure 4: The geodesic e (in green), the a-curves (in red), and the b-curves (in blue)

Observe that $i(a_j, e) = i(b_j, e) = i(a_j, b_j) = 1$ for every j and that the curves a_j, b_j and e bound two triangles with the same interior angles. These two triangles are therefore congruent, so that their side lengths are $a_j/2$, $b_j/2$ and $e/2$. In particular, they are isoceles since $\ell(a_j) = \ell(b_j)$. The altitude of each triangle has length t and bisects the base, which yields the formula

$$\cosh(a/2) = \cosh(t)\cosh(e/4) \tag{2.7}$$

by Equation (2.1) for right triangles. One such pair of triangles is illustrated in Figure 15.

Remark 2.1. Surfaces of genus 1 with m boundary components are studied extensively in [48] where it is shown that the lengths of the boundary geodesics and the a-, b- and e-curves in such a surface define an injective function on Teichmüller space. Actually, the length of any boundary geodesic can be recovered from the remaining ones. This detailed analysis is pursued in [49, Section 4] where these rings serve as building blocks for

constructing maximal surfaces. We combine rings differently, resulting in a more flexible construction.

2.5 Symmetries of the ring

There is an orientation-reversing isometric involution $\rho_{\text{seams}} : R \to R$ that has the union of the σ-segments (the *seams* of the crosses) as its set of fixed points. The map ρ_{seams} exchanges the front and back octagons in each cross $C_j \subset R$. It acts as a glide reflection along e by half its length.

Another obvious isometry is the reflection $\rho_e : R \to R$ across the geodesic e. This isometry permutes the top and bottom of each cross.

For each $j \in \{1, \dots, n\}$, there is a reflection ρ_{f_j} across the geodesic f_j.

Another useful isometry $\eta : R \to R$ simply shifts each O_j to O_{j+1}, where indices are taken modulo $2n$. That is, η is a hyperbolic translation along e to the right by distance $e/2n = 2u$.

Lastly, for each $j \in \{1, \dots, n\}$ the composition $\nu_j := \eta \circ \rho_{f_j}$ is the reflection of R across the vertical axis of symmetry of C_j.

2.6 Systoles in the ring

Recall that the systole of a hyperbolic surface is the length of any of its shortest closed geodesics, also called systoles. Contrary to some authors, we allow boundary geodesics to be systoles.

The following well-known criterion is very useful for finding systoles (this is just mentioned in passing in [49] and [51]). Note that the analogous statement and proof is false for surfaces with punctures. See [20, Theorem 1.3] for the correct replacement.

Lemma 2.2. *If two closed geodesics α and β on a compact hyperbolic surface with geodesic boundary intersect at least twice transversely, then there exists a closed geodesic γ of length strictly less than $(\ell(\alpha) + \ell(\beta))/2$. In particular, two distinct systoles can intersect at most once.*

Proof. Let p and q be two intersection points of α and β. Construct a curve δ by taking the shorter subarc of α between p and q and similary for β. Since geodesic bigons are non-contractible, δ is homotopic to a closed geodesic γ that is strictly shorter. □

We will apply the contrapositive of the last sentence in Lemma 2.2 repeatedly: *if two systoles intersect at least twice, then they coincide*. We use this fact in combination with the various symmetries of the ring to determine its systoles. We proceed by elimination, arguing that any geodesic—save for a few exceptions—intersects some of its translates at least twice transversally, hence cannot be a systole in view of the above.

Proposition 2.3. *Let $n \geq 1$ and $t > 0$. Assume that $a(t) < 4t$ and $a(t) < e(t)$. Then the systoles in $R(n,t)$ are exactly the a-curves and the b-curves.*

Proof. Let γ be a systole of R. We claim that γ intersects the seams, e, and each f-curve at most once. Otherwise, γ and its image γ^* by one of the reflections ρ_{seams}, ρ_e, or ρ_{f_j} intersect at least twice. In that case $\gamma = \gamma^*$ by Lemma 2.2. We rule out the possibility that γ coincides with $\rho_{\text{seams}}(\gamma)$, $\rho_e(\gamma)$, or $\rho_{f_j}(\gamma)$ one by one below, thereby proving the claim.

Suppose that $\rho_{\text{seams}}(\gamma) = \gamma$ and that γ is disjoint from e. Then either γ is a boundary component of R in which case $\ell(\gamma) = 4t > a$ and γ is not a systole, or else γ intersects its shift $\eta(\gamma)$ twice transversely, contradicting Lemma 2.2. We conclude that γ intersects e, and it does so at least twice by ρ_{seams}-symmetry. Therefore $\rho_e(\gamma)$ and γ intersect at least twice as well so that they coincide. Then either $\gamma = e$ or $\gamma \perp e$. In the first case $\ell(\gamma) > a$ by hypothesis so that γ is not a systole. In the second case γ has to be equal to some f-curve, so that $\ell(\gamma) = 4t > a$. We conclude that γ intersects the seams at most once. Actually, γ intersects the seams exactly once. Indeed, the complement of the seams is a topological annulus whose only simple closed geodesic is e, which is not a systole. Thus γ cannot be disjoint from the seams.

Now suppose that γ intersects e at least twice so that $\rho_e(\gamma) = \gamma$. Since ρ_e does not fix any point on the seams, the number of intersection points between γ and the seams is even, which contradicts the previous paragraph. Therefore, γ intersects e at most once. In fact, γ cannot be disjoint from e either. This is because the seams disconnect $R \setminus e$, yet γ intersects them only once. This shows that γ intersects e exactly once.

Lastly, suppose that γ intersects some f_j at least twice so that $\rho_{f_j}(\gamma) = \gamma$. Since γ cannot be equal to f_j, it is orthogonal to it. Moreover, γ must intersect the seams and e at one of the places where f_j does, for otherwise there would be a second intersection point by ρ_{f_j}-symmetry. The only closed curve that is orthogonal to f_j at one of these four points is e, which is too long. Hence γ intersects each f-curve at most once.

Now that the claim is proved, it is not hard to show that γ is either an a-curve or a b-curve. If we cut R along the seams, we get an annulus A. The curve γ gets cut into an arc ω in A joining a pair of points that get identified by the gluing pattern. The arc ω must join a point on the bottom boundary of A to a point on the top boundary since it intersects e. Moreover, ω cannot wrap around A more than once, for otherwise it would intersect some f-curve twice. Thus ω wraps exactly halfway around A (remember, the seams are glued via a glide reflection along e by distance $e/2$). It follows that γ is homotopic to—hence equal to—one of the a-curves or b-curves. □

2.7 Transverse rings

Since the crosses used to build the ring R have diagonal symmetry, we can make two rings overlap along a shared cross. We call this configuration a *pair of transverse rings*.

We can think of one ring as being horizontal and the other vertical, as in Figure 5. The e-curves in the two rings intersect twice, bisecting each other perpendicularly. There are four different ways to apply the surgery procedure from the proof of Lemma 2.2 to this pair of curves, yielding four geodesics shorter than e that we call *c-curves*. One of them is depicted in Figure 5.

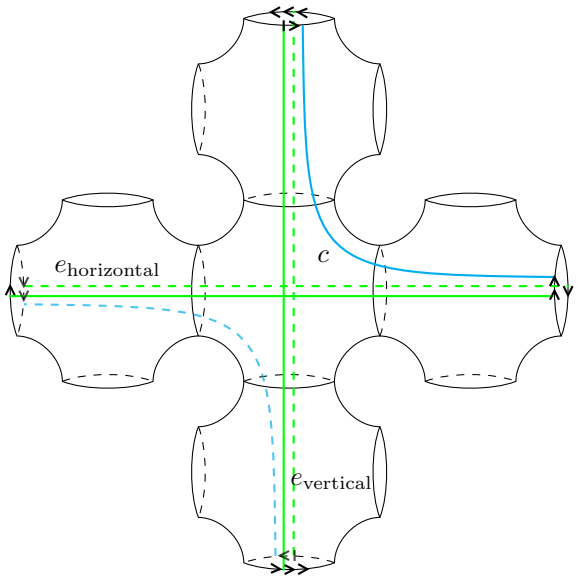


Figure 5: One of the four c-curves in a pair of transverse rings, obtained by surgery on the e-curves

The four c-curves have equal length since they are related by symmetries. Furthermore, there is a right-angled pentagon with two adjacent sides of length $e/4$ and the opposite side of length $c/2$ (see Figure 5). Equation (2.3) gives the formula

$$\cosh(c/2) = \sinh^2(e/4) \tag{2.8}$$

for the length of any c-curve.

When $n = 1$ the pair of transverse rings is reduced to a single cross and there are actually only two c-curves because some surgeries on the e-curves coincide. In this case, each c-curve is equal to the union of two opposite seams and Equation (2.8) is really the same as Equation (2.5). We will analyze this case more carefully in the next subsection.

The next step is to fix the parameter t in such a way that the curves a, b and c all have the same length. A first useful observation is that c is a decreasing function of t.

Lemma 2.4. *For every $n \geq 1$, the functions $e(t)$ and $c(t)$ are decreasing in t.*

Proof. Recall that $e(t) = 4n \operatorname{arcsinh}(\coth(t))$. Since $\coth$ is decreasing and $\operatorname{arcsinh}$ is increasing, e is decreasing. Therefore $c(t) = 2 \operatorname{arccosh}(\sinh^2(e(t)/4))$ is decreasing as well, being the composition of a decreasing function with an increasing one. □

We use this to prove the existence and uniqueness of a parameter t_n such that the curves a, b and c in the pair of transverse rings all have the same length.

Lemma 2.5. *For every $n \geq 1$, there exists a unique $t_n > 0$ such that $a(t_n) = c(t_n)$.*

Proof. We have

$$\frac{\cosh(a(t)/2)}{\cosh(e(t)/4)} = \cosh(t)$$

and

$$\frac{\cosh(c(t)/2)}{\cosh(e(t)/4)} = \frac{\sinh^2(e(t)/4)}{\cosh(e(t)/4)} = \tanh(e(t)/4)\sinh(e(t)/4)$$

by Equations (2.7) and (2.8). Therefore, the equation $a(t) = c(t)$ is equivalent to

$$\cosh(t) = \tanh(e(t)/4)\sinh(e(t)/4). \tag{2.9}$$

The left-hand side of (2.9) is an increasing function of t which diverges as $t \to \infty$. The right-hand side is decreasing in t since it is the product of two positive decreasing functions. Moreover, it diverges as $t \to 0$ since $e(t)$ does. The existence and uniqueness of t_n follows. □

From now on, we will only work with the rings $R(n, t_n)$ with t_n as in Lemma 2.5. In order to determine the systoles in that ring, we need to check that the hypotheses of Proposition 2.3 are satisfied, but this is only true when $n \geq 2$. The case $n = 1$ is treated separately in the next subsection.

Lemma 2.6. *We have $a(t_n) < 4t_n$ and $a(t_n) < e(t_n)$ for every $n \geq 2$.*

Proof. The inequality $a(t_n) = c(t_n) < e(t_n)$ follows from the fact that c is obtained by surgery on two e-curves, or can be deduced from Equation (2.8).

To show that $a(t_n) < 4t_n$ we consider the time $s_n > 0$ such that $a(s_n) = 4s_n$ and prove that $c(s_n) > 4s_n$. This implies that $s_n < t_n$ since c is decreasing whereas $a(t)$ diverges as $t \to \infty$. The inequality $a(t_n) < 4t_n$ then follows from the fact that

$$\frac{\cosh(a(t)/2)}{\cosh(2t)} = \frac{\cosh(t)}{\cosh(2t)} \cdot \cosh(e(t)/4)$$

is decreasing, being the product of two positive decreasing functions. Figure 6 illustrates this phenomenon for $n = 3$.

Hence let $s_n > 0$ be the unique parameter such that $a(s_n) = 4s_n$. Then

$$\cosh(s_n)\cosh(e(s_n)/4) = \cosh(a(s_n)/2) = \cosh(2s_n) = 2\cosh^2(s_n) - 1$$

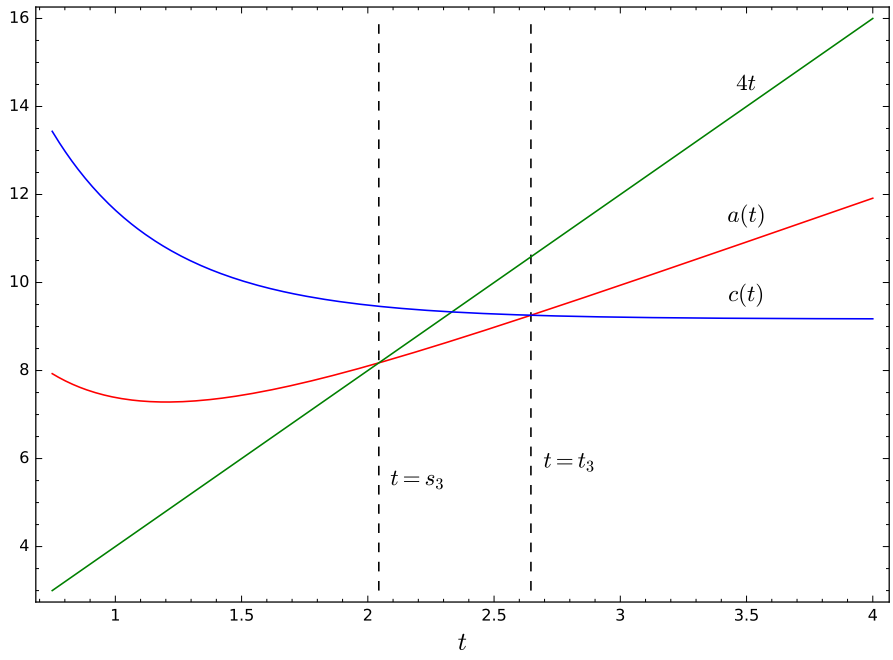


Figure 6: A plot of the functions $a(t)$, $c(t)$ and $4t$ for $n = 3$

and

$$\begin{aligned}\cosh(c(s_n)/2) = \sinh^2(e(s_n)/4) &= \cosh^2(e(s_n)/4) - 1 \\ &= \left(\frac{2\cosh^2(s_n) - 1}{\cosh(s_n)}\right)^2 - 1.\end{aligned}$$

Let $x = \cosh^2(s_n)$ so that $\cosh(2s_n) = 2x - 1$ and

$$\cosh(c(s_n)/2) = \frac{(2x-1)^2}{x} - 1.$$

The inequality we want to prove is $\cosh(c(s_n)/2) > \cosh(2s_n)$, which is equivalent to $(2x-1)^2 > 2x^2$ or $x > 1 + \frac{1}{\sqrt{2}}$ after simplification. Therefore, it suffices to show that

$$s_n > \operatorname{arccosh}\left(\sqrt{1 + \frac{1}{\sqrt{2}}}\right) \approx 0.764.$$

But at $t = 1$ we get

$$\begin{aligned}\frac{\cosh(a(1)/2)}{\cosh(2\cdot 1)} &= \frac{\cosh(1)}{\cosh(2\cdot 1)} \cdot \cosh(e(1)/4) \\ &= \frac{\cosh(1)}{\cosh(2)} \cdot \cosh(n \operatorname{arcsinh}(\coth(1))) \\ &\geq \frac{\cosh(1)}{\cosh(2)} \cdot \cosh(2 \cdot 1.086) > \cosh(1) > 1\end{aligned}$$

which implies that $s_n > 1$ and finishes the proof. □

We conclude that the systoles in the ring $R(n, t_n)$ are the a-curves and the b-curves.

Corollary 2.7. *For every $n \geq 2$, the systoles in $R(n, t_n)$ are the a-curves and the b-curves.*

Proof. This follows from Lemma 2.6 and Proposition 2.3. □

We observed earlier that c is a decreasing function of t. The function a is not monotone but we can show it is increasing at t_n. These two facts will play a key role in Section 4.

Lemma 2.8. *We have $a'(t_n) > 0$ for every $n \geq 2$.*

Proof. From $\cosh(a(t)/2) = \cosh(t)\cosh(e(t)/4)$ we compute

$$\begin{aligned}\sinh(a(t)/2)\, a'(t)/2 &= \sinh(t)\cosh(e(t)/4) + \cosh(t)\sinh(e(t)/4)\, e'(t)/4 \\ &> \sinh(e(t)/4)\,[\sinh(t) + \cosh(t)\, e'(t)/4]\,.\end{aligned}$$

Thus it suffices to show that $-e'(t_n)/4 < \tanh(t_n)$. Since $e(t)/4 = n\,\mathrm{arcsinh}(\coth t)$ we get

$$-e'(t)/4 = \frac{n}{\sinh^2(t)\sqrt{\coth^2(t)+1}} < \frac{n}{\sinh^2(t)\,\sqrt{2}}$$

so that the required inequality becomes $n < \sqrt{2}\,\tanh(t_n)\sinh^2(t_n)$.

We know that $t_n > 1$ from the proof of Lemma 2.6. Furthermore, one can show that

$$\sqrt{2}\,\tanh(x)\sinh^2(x) > 0.963 \cdot \cosh(x)$$

for every $x \geq 1$. Indeed, $\sinh^3(x)/\cosh^2(x)$ is increasing and the inequality can be verified numerically at $x = 1$. Recall that $\cosh(t_n) = \tanh(e(t_n)/4)\sinh(e(t_n)/4)$ by definition of t_n. We thus obtain

$$\begin{aligned}\sqrt{2}\,\tanh(t_n)\sinh^2(t_n) &> 0.963 \cdot \cosh(t_n) \\ &= 0.963 \cdot \tanh(e(t_n)/4)\sinh(e(t_n)/4) \\ &> 0.963 \cdot \tanh(n\lambda)\sinh(n\lambda) \geq n\end{aligned}$$

for every $n \geq 2$, where $\lambda = \mathrm{arcsinh}(1)$. The last inequality holds because the function $\tanh(\lambda x)\sinh(\lambda x)/x$ is increasing in x and larger than $1/0.963$ at $x = 2$. This implies the desired result. □

In addition to knowing the systoles in the ring, we also need an estimate on the lengths of arcs that enter and exit the ring from a given cross. Since the arcs going vertically across any cross $C_j \subset R$ are fairly short, we need to exclude them.

Lemma 2.9. *Let $n \geq 2$. Any non-trivial arc in $R(n, t_n)$ from one boundary component to itself is longer than $a(t_n)/2$. Any geodesic arc that joins the top and bottom of a cross $C_j \subset R(n, t_n)$ but is not contained in C_j is longer than $a(t_n)/2$.*

Proof. Let γ be a shortest non-trivial arc from one boundary B of $R(n,t_n)$ to itself. In particular, γ is geodesic and orthogonal to the boundary.

If γ intersects some f_j twice, then we can reflect a subarc $\omega \subset \gamma$ from f_j to itself across f_j to obtain a non-trivial closed curve of length $2\ell(\omega)$ in $R(n,t_n)$. By Corollary 2.7 we get that $2\ell(\gamma) > 2\ell(\omega) \geq a(t_n)$.

If γ intersects the seams, then we can perform a surgery on γ and $\rho_{\text{seams}}(\gamma)$ to obtain a strictly shorter essential arc from B to itself, unless $\gamma = \rho_{\text{seams}}(\gamma)$. One way to see this is to double the ring $R(n,t_n)$ across its boundary and apply Lemma 2.2 to the doubled arcs. Thus distinct non-trivial arcs of minimal length from B to itself are disjoint. But if $\gamma = \rho_{\text{seams}}(\gamma)$, then γ intersects some f-curve at least twice, hence is longer than $a(t_n)/2$ by the previous paragraph. The only exception is if γ is contained in a single cross C_j. But in that case, if we double C_j across B we obtain a pair of crosses and a closed geodesic of length $2\ell(\gamma)$ in it. This pair embeds isometrically in $R(n,t_n)$, showing that $2\ell(\gamma) > a(t_n)$. The inequality is strict because no systole in $R(n,t_n)$ is symmetric about any f-curve.

We can therefore assume that γ is disjoint from the seams and intersects each f-curve at most once. Up to the symmetries of $R(n,t_n)$, this leaves two possibilities for γ depending whether it intersects e or not.

Recall that the complement of the seams in $R(n,t_n)$ is an annulus A. As such, there is a well-defined orthogonal projection $A \to e$. If γ does not intersect e, then it intersects all the f-curves, and its orthogonal projection onto e is longer than

$$\frac{(2n-1)}{2n} e(t_n) > \frac{1}{2} e(t_n) > \frac{1}{2} a(t_n).$$

See Figure 7. Since the orthogonal projection does not increase distances, we get that $\ell(\gamma) > a(t_n)/2$.

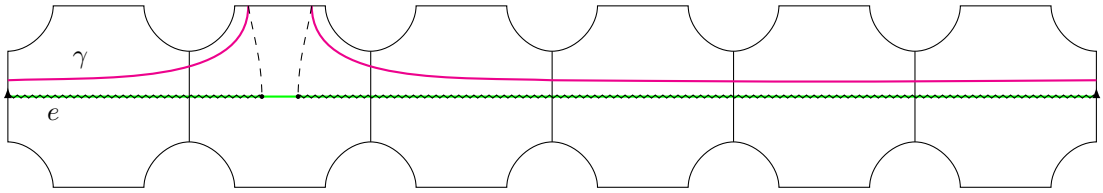


Figure 7: If γ does not intersect e, then its projection onto e covers most of e

If γ intersects e, then γ and $\rho_e(\gamma)$ intersect. One of the two possible surgeries on $\gamma \cup \rho_e(\gamma)$ yields a pair of arcs α and β, each joining the top and bottom boundaries of some cross C_j in $R(n,t_n)$, neither of which can be homotoped into C_j (see Figure 8). This gives $\ell(\gamma) > \ell(\alpha) = \ell(\beta)$, so it suffices to show that $\ell(\alpha) > a(t_n)/2$. We have reduced the first part of the statement of the lemma to the second part.

Let τ be an arc of minimal length in $R(n,t_n)$ that joins the top and bottom boundaries of some cross C_j and cannot be homotoped into C_j. By the same argument as above, we

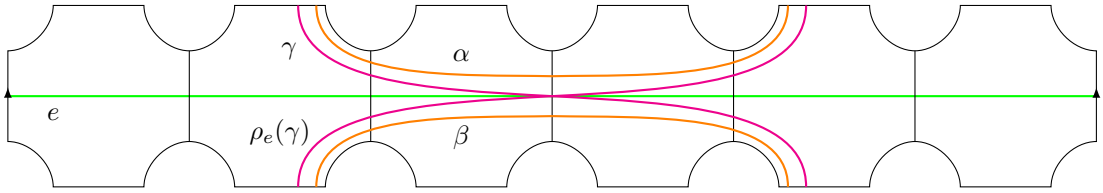


Figure 8: If γ intersects e there is a surgery on $\gamma \cup \rho_e(\gamma)$ producing a pair of arcs α and β joining two opposite boundaries of a cross

may assume that τ intersects each f-curve at most once and is disjoint from the seams. If τ intersects e, then it wraps most of the way around the annulus A so that its orthogonal projection onto e is longer than $e(t_n)/2 > a(t_n)/2$ similarly as above. Otherwise, τ is equal to the arc α from the preious paragraph or one of its images by the group $\langle \rho_{\text{seams}}, \rho_e, \nu_j \rangle$ where ν_j is the reflection swapping the left and right sides of C_j. In any case, there is a right-angled pentagon with two adjacent sides of lengths $e/4$ and $e/4n$, and the opposite side of length $\tau/2$ (see Figure 9). Equations (2.3) and (2.6) give

$$\cosh(\tau/2) = \sinh(e/4)\sinh(e/4n) = \sinh(e/4)\coth(t) > \sinh(e/4).$$

Squaring yields

$$\frac{\cosh(\tau)+1}{2} = \cosh^2(\tau/2) > \sinh^2(e/4) = \cosh(c/2)$$

hence

$$\cosh(\tau) > 2\cosh(c/2) - 1 > \cosh(c/2).$$

This shows that $\ell(\tau) > c(t_n)/2 = a(t_n)/2$, which concludes the proof. □

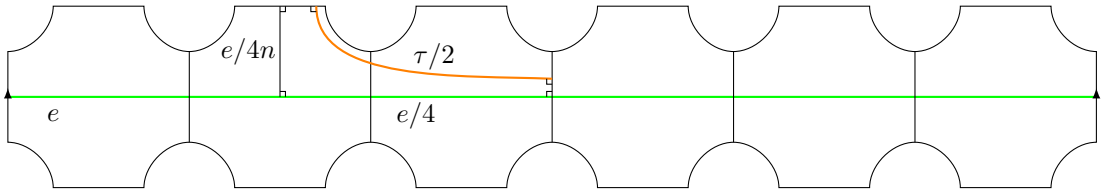


Figure 9: The right-angled pentagon allowing us to compute the length of the shortest arc in Lemma 2.9

2.8 The Bolza curve

When $n = 1$, the pair of transverse rings is a closed surface of genus 2 obtained by gluing the opposite sides of the cross $C(t_1)$ with half twists. We now show that this surface—denoted $\Sigma(1)$—is the *Bolza curve*, which is the surface of genus 2 with the largest systole

and the largest automorphism group [36]. In fact, it is the only local maximum of the systole in genus 2 [49].

Proposition 2.10. $\Sigma(1)$ *is the Bolza curve.*

Proof. Let s be the side length of a regular hyperbolic triangle with interior angles $\pi/4$. Eight such triangles fit together at a point to form a regular right-angled octagon O. Glue two such octagons together to form a cross isometric to $C(s/2)$, then glue opposite ends of $C(s/2)$ with half twists. The a- and b-curves in the resulting closed surface are main diagonals of O, hence have length $2s$. Similarly, each c-curve is equal to the union of two opposite sides of the octagon, hence has length $2s$. This shows that $a(s/2) = 2s = c(s/2)$ so that $t_1 = s/2$. In particular, the f-curves in $\Sigma(1)$ have the same length as the curves of type a, b and c.

Now cut the front octagon of $\Sigma(1)$ into 8 equilateral triangles and attach them to the corresponding sides of the back octagon. The result is a regular octagon with interior angles $\pi/4$. The sides of the latter are identified in opposite pairs to form $\Sigma(1)$ (see Figure 10). This is a standard representation of the Bolza curve [37, Section 3]. □

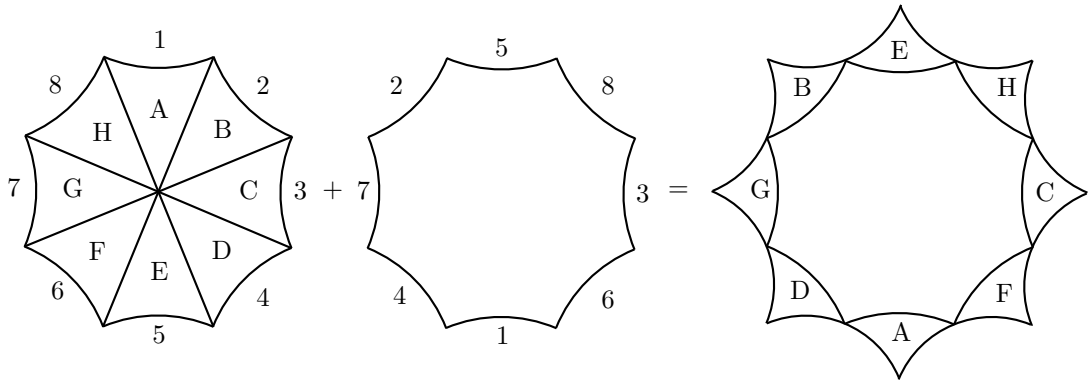


Figure 10: $\Sigma(1)$ is the Bolza surface

Remark 2.11. The above proof shows that $a(t_1) = c(t_1) = 4t_1$. After some elementary algebraic manipulations[3], one arrives at the exact formula $t_1 = \operatorname{arccosh}(1+\sqrt{2})/2$.

It is interesting to note that in genus 3, there are at least three local maxima, and the most symmetric surface, the Klein quartic, is not the global maximum [49].

2.9 The tree of rings

For $n \geq 1$, let $T(n)$ be the n-regular tree. We build a hyperbolic surface $\Sigma(n)$ called the *tree of rings* by replacing each vertex $v \in T(n)$ with a copy R_v of the ring $R(n, t_n)$ such

[3]We have $\cosh(2t_1) = \cosh(c(t_1)/2) = \sinh^2(e(t_1)/4) = \coth^2(t_1)$ by definition, which implies that $(2\cosh^2(t_1)-1)(\cosh^2(t_1)-1) = \cosh^2(t_1)$. This is a quadratic equation in $\cosh^2(t_1)$ whose only solution larger than 1 is given by $\cosh^2(t_1) = 1 + 1/\sqrt{2}$. Thus $\cosh(2t_1) = 2\cosh^2(t_1) - 1 = 1 + \sqrt{2}$.

that two rings R_v and R_w are transverse if and only if the vertices v and w are adjacent in $T(n)$. In other words, each edge of $T(n)$ is replaced by a cross $C(t_n)$ and the crosses are glued in such a way that those corresponding to the n edges adjacent to any vertex in $T(n)$ form a ring isometric to $R(n, t_n)$.

The resulting surface $\Sigma(1)$ is closed of genus two, $\Sigma(2)$ has two ends accumulated by genus and $\Sigma(n)$ has a Cantor set of ends accumulated by genus when $n \geq 3$ (see Figure 11).

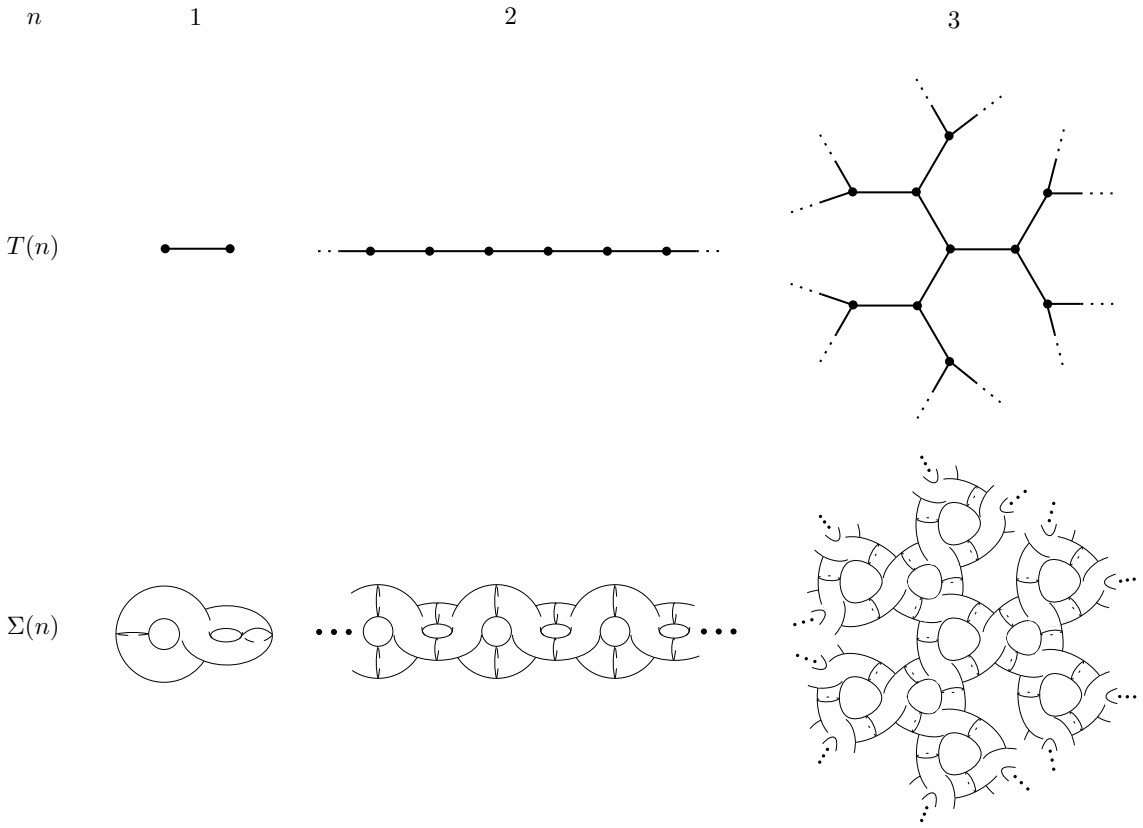


Figure 11: The tree of rings for $n = 1, 2, 3$

We now determine the systoles in $\Sigma(n)$, starting with $\Sigma(1)$ as a warm-up.

Proposition 2.12. *The systoles in* $\Sigma(1)$ *are the* a*- and* b*-curves in the horizontal and vertical rings, the two* c*-curves and the two* f*-curves. The total number of systoles is* 12 *and their length is* $2 \operatorname{arccosh}\left(1 + \sqrt{2}\right) \approx 3.057$.

Proof. The e-curves are longer than the c-curves by construction, hence longer than the a- and b-curves. The proof of Proposition 2.3 applies almost verbatim to show that the shortest curves disjoint from the horizontal (resp. vertical) f-curve are the a- and b-curves in the horizontal (resp. vertical) ring together with the vertical (resp. horizontal) f-curve. The only difference is that the f-curves were ruled out in Proposition 2.3 for being too long by hypothesis.

Let $\gamma \subset \Sigma(1)$ be a systole that intersects both f-curves. Consider the two diagonal axes of symmetry of the cross $C(t_1)$. These curves divide $\Sigma(1)$ into a union of two congruent annuli with piecewise geodesic boundary, each containing one of the f-curves as its core geodesic. By hypothesis γ traverses each annulus at least once. It is easy to see that the shortest arc across either annulus has length $c/2$. Therefore $\ell(\gamma) \geq c$ with equality if and only if γ is a concatenation of two seams, i.e., a c-curve.

Since the a-, b-, c- and f-curves all have the same length equal to $2\operatorname{arccosh}\left(1+\sqrt{2}\right)$ (see Remark 2.11), they are the systoles. □

Proposition 2.13. *For every $n \geq 2$, the systoles in the tree of rings $\Sigma(n)$ are the a- and b-curves contained in rings, together with the c-curves contained in pairs of transverse rings.*

Proof. Let γ be a systole of $\Sigma(n)$. We define the *shadow* of γ in $T(n)$ as follows. First we cut γ along the f-curves into subarcs $\gamma_1, \ldots, \gamma_k$ labelled in cyclic order along γ. For each subarc γ_j that joins two boundaries of a cross C which are not opposite of each other (i.e., each subarc that "turns" from one ring to another), its shadow $s(\gamma_j)$ is the edge in $T(n)$ corresponding to the pair of transverse rings that intersect along C. The shadow $s(\gamma_j)$ of each subarc γ_j that does not turn is defined to be the vertex $v \in T(n)$ corresponding to the ring R_v containing γ_j in its interior. The shadow $s(\gamma)$ is defined as the concatenation of the shadows $s(\gamma_1), \ldots, s(\gamma_k)$. This forms a loop in $T(n)$.

The shadow $s(\gamma)$ is not well-defined if γ is disjoint from the f-curves or is equal to one of them. But in that case γ is contained in a ring so that it is either an a-curve or a b-curve by Corollary 2.7.

Being a loop in a tree, $s(\gamma)$ has at least two places where it backtracks, that is, an edge which it traverses twice in a row in opposite directions. By definition of the shadow, a backtrack corresponds to an arc entering and leaving a ring through the same cross, turning at the beginning and at the end. By Lemma 2.9, such an arc is longer than $a(t_n)/2$. In particular, if $s(\gamma)$ has two backtracks happening along two distinct edges, then γ has two disjoint subarcs longer that $a(t_n)/2$ each, so that it is not a systole.

This leaves the possibility that $s(\gamma)$ is just a loop formed by traversing one edge $\{v, w\}$ of $T(n)$ twice in opposite directions. In that case, γ is contained in a pair of transverse rings $R_v \cup R_w$ and turns exactly twice in the cross $C = R_v \cap R_w$.

We can write γ as the concatenation of two arcs γ_v and γ_w where $\gamma_v = \gamma \cap R_v$ and $\gamma_w = \gamma \setminus \gamma_v$. This means that γ_v contains both turns of γ. In particular, γ_v is not contained in C so that $\ell(\gamma_v) > a(t_n)/2$ by Lemma 2.9.

Suppose that the two endpoints of γ_v belong to the same boundary component of C. Then γ_w—which is contained in R_w—can be reflected across that f-curve to form a non-trivial closed curve in R_w. That curve is longer than $a(t_n)$ by Corollary 2.7, hence $\ell(\gamma_w) > a(t_n)/2$. This gives $\ell(\gamma) = \ell(\gamma_v) + \ell(\gamma_w) > a(t_n)$.

By exchanging the roles of R_v and R_w, the previous argument shows that the two turning subarcs of γ have endpoints in all four boundary components of C. This implies that γ intersects one of the two diagonal axes of symmetry of C—call it d—twice. But the reflection of C in the curve d extends to a global isometry ρ_d of $\Sigma(n)$. By Lemma 2.2, we have $\rho_d(\gamma) = \gamma$. If γ also intersects the seams, then it intersects them twice by symmetry across d. In that case, γ is invariant under ρ_{seams} as well. But then the two turning subarcs of γ in C are mirror images across the seams, hence have endpoints in only two boundary components of C. That contradicts the first sentence of this paragraph.

We know that $\rho_d(\gamma) = \gamma$ and that γ is disjoint from the seams. Consider the subarc $\alpha \subset \gamma$ contained in R_v with two endpoints on d and let $\beta = \rho_d(\alpha)$ so that $\gamma = \alpha \cup \beta$. If α intersects any f-curve twice, then $\ell(\alpha) > a(t_n)/2$ by an argument above so that $\ell(\gamma) = \ell(\alpha) + \ell(\beta) = 2\ell(\alpha) > a(t_n)$. Thus α intersects each f-curve at most once. This determines the homotopy class of α up to moving the endpoints along d since the complement of the seams in R_u is an annulus. That is, α wraps once around R_u intersecting each f-curve once along the way, while staying disjoint from the seams. We conclude that γ is homotopic to a c-curve, hence equal to one of them. □

2.10 Signed graphs

Let $n \geq 3$. In order to get a closed surface, we glue copies of the cross $C(t_n)$ along a finite n-regular graph Γ instead of the tree $T(n)$. In order to determine the gluings precisely, we need a bit more structure on Γ, namely,

- a cyclic ordering of the edges adjacent to any vertex;
- a sign $\varepsilon(e_1, e_2) \in \{+, -\}$ attributed to any two consecutive edges e_1, e_2 around a vertex, subject to the condition that the product of the signs around any vertex is negative.

We call a graph equipped with this additional structure a *signed graph*. Note that a choice of cyclic ordering around each vertex is known as an (oriented) ribbon structure. However, we will now define when two signed graphs are isomorphic, and such isomorphisms need not preserve the ribbon structure.

Given a vertex x in a signed graph Γ, we define the *vertex flip* around x to be the operation that reverses the cyclic ordering around x and changes the signs between each edge e containing x and its two immediate neighbors in the cycling ordering around the vertex $e \setminus x$ (see Figure 12). Clearly, any two vertex flips commute. We say that two signed graphs are *isomorphic* if one can be obtained from the other by a set of vertex flips.

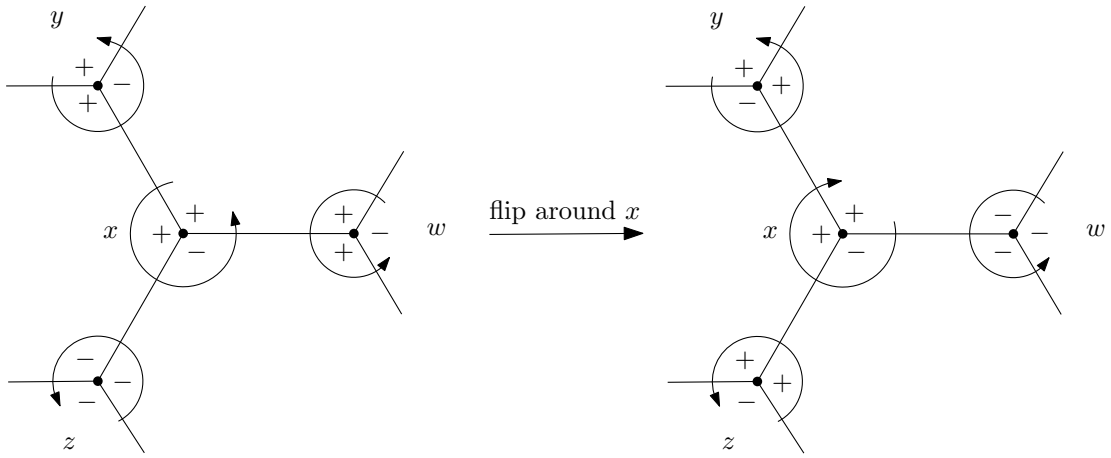


Figure 12: A vertex flip on a signed graph

2.11 Gluing crosses according to a signed graph

Let $n \geq 3$ and let Γ be a connected, n-regular, signed graph. We construct a surface $X(\Gamma)$ modelled on Γ as follows. To each edge e in Γ corresponds a cross C_e isometric to $C(t_n)$. The edge $e = \{u, v\}$ has two neighboring edges (which coincide when $n = 2$) around each of u and v. We glue

- the predecessor of e around u to the left of C_e;
- the successor of e around u to the right of C_e;
- the predecessor of e around v to the bottom of C_e;
- the successor of e around v to the top of C_e.

Each of these gluings is done as to make the seams match. This still leaves two possibilities for each gluing: either with a half twist or not. This is determined using the signs between consecutive edges: the "+" signs mean no twist and the "−" signs call for half twists.

Note that for a string of crosses, the half twists do not affect the isometry type. However, when we close up the string to form a loop, they do. For example, with an even number of half twists the seams separate, but with an odd number of half twists they do not. In fact, one can show that a chain of n crosses isometric to $C(t_n)$ glued end to end is isometric to the ring $R(n, t_n)$ if and only if the number of half twists is odd. This is why we require the product of the signs around each vertex in Γ to be negative.

We also remark that rotating each cross by angle π around one of its diagonals exchanges left and bottom as well as right and top. Thus changing the order between u and v above merely switches the horizontal and vertical axes but not the gluings themselves.

Each ring can be seen as either horizontal or vertical interchangeably; this notion need not be globally defined.

The surface $X(\Gamma)$ is defined as

$$X(\Gamma) = \left(\bigsqcup_{e \in E} C_e \right) / \sim$$

where E is the set of edges of Γ and the equivalence $\sim$ identifies boundary points of different crosses as described above.

The sign structure of Γ induces a cyclic ordering of the crosses in each ring. For any ring R in $X(\Gamma)$, there are exactly n other rings transverse to it. When two of these transverse rings pass through adjacent crosses of R, let us say that they are *parallel*. Whether the cyclic orderings in parallel rings passing through adjacent crosses C_{e_1} and C_{e_2} agree or not is indicated by the sign $\varepsilon(e_1, e_2)$. If we reverse the cyclic ordering of the crosses in R, then the comparison between parallel rings transverse to R is unaffected. However, for every ring parallel to R, there is a change of sign: if orders agreed before, they do not anymore and vice versa. In other words, if Γ_1 and Γ_2 are isomorphic signed graphs, then there is an orientation-preserving isometry between $X(\Gamma_1)$ and $X(\Gamma_2)$.

As an example, if Γ is the n-regular tree $T(n)$ equipped with an arbitrary sign structure, then $X(\Gamma)$ is isomorphic to the tree of rings $\Sigma(n)$. This is because any two sign structures on $T(n)$ are isomorphic, a fact left as an exercise[4] to the reader.

2.12 The ribbon graph induced by a signed graph

There is a useful combinatorial object $\widehat{\Gamma}$ that comes between the signed graph Γ and the surface $X(\Gamma)$ which makes the correspondence more transparent. This object is a (non-orientable) 4-regular ribbon graph, and is obtained from Γ as follows:

- to each edge $e = \{u, v\}$ of Γ corresponds a vertex $\widehat{e}$ in $\widehat{\Gamma}$;
- each vertex in $\widehat{\Gamma}$ is 4-valent, and its adjacent edges are given a cyclic order;
- the vertices in $\widehat{\Gamma}$ that correspond to the predecessor and successor of e around u in Γ share edges with $\widehat{e}$, and these edges are to be opposite in the cyclic order;
- similarly for the vertices corresponding to the two immediate neighbors of e in the cyclic order around v;
- the ribbon edge between two vertices in $\widehat{\Gamma}$ is given a half twist if the sign between the corresponding edges of Γ is negative, and no twist otherwise.

[4] Hint: First show that any sign pattern (with negative product) around a vertex v can be changed into any other (with negative product) by doing some vertex flips around the neighbors of v. Furthermore, this can be done even if one neighbor of v is required to be left intact.

In this way, the n edges adjacent to any vertex in Γ become a cycle of length n in $\widehat{\Gamma}$ which is homeomorphic to a Möbius band, because there is an odd number of half twists. Adjacent vertices in Γ correspond to Möbius bands that intersect transversely in $\widehat{\Gamma}$.

To go from $\widehat{\Gamma}$ to $X(\Gamma)$, simply inflate each 4-valent vertex to a cross $C(t_n)$. Associate the edges around the vertex to the four boundary components of $C(t_n)$ so that the cyclic order goes: left, bottom, right, top. Then glue crosses with or without half twist according to whether the edges of $\widehat{\Gamma}$ have a half twist or not.

From the surface $X(\Gamma)$, we can go back to $\widehat{\Gamma}$ by collapsing the front and back of each cross (i.e., taking the quotient of $X(\Gamma)$ by the reflection across the seams) then taking the graph dual to the decomposition of the resulting surface into octagons. Note that in this way, the seams of $X(\Gamma)$ correspond to the boundary components of $\widehat{\Gamma}$.

2.13 The $n = 2$ case

A 2-regular signed graph Γ does not appear to carry enough information to prescribe how to glue crosses together. For instance, there is only one cyclic ordering on two elements, whereas there are two distinct directions of travel along a ring made with two crosses.

For $n = 2$, we start directly with a graph G playing the role of $\widehat{\Gamma}$ instead. That is, let G be a finite, connected, 4-regular, ribbon graph such that any path in G which does not turn (i.e., goes to the opposite edge in the cyclic order at each vertex) is closed of length 2, and has a neighborhood homeomorphic to a Möbius band. Given such a graph G, we obtain a surface $X(G)$ by replacing each vertex of G with a cross $C(t_2)$ and gluing them in the prescribed way as in the previous subsection. The resulting surface $X(G)$ is such that each of its crosses belongs to two rings isometric to $R(2, t_2)$.

We claim that there are two isomorphism classes of such graphs G with V vertices if $V \geq 2$ is a multiple of 3, and only one isomorphism class otherwise.

Pick any Möbius band B of length two in G and cut G along the two edges of B. The resulting object H has two vertices that have two opposite half-edges not connected to anything. Pick either of these vertices, start on one side of it, and start walking along an uncut edge. At the next encountered vertex, turn left, and so on, until you reach a dead end. In this way, the path traced is a boundary component of H which passes through each vertex only once.

We can draw the ribbon graph H in the plane as a tubular neighborhood of a regular V-gon with its sides extended a little bit, one side cut open, and the ends of each uncut side glued via a half twist (see Figure 13). The left-turning path traced above corresponds to the inner boundary component of this cut V-gon.

The graph G is obtained from H by pairing up the two free half-edges of the first vertex with the two free half-edges of the last vertex, and giving one pair a half twist. There are two ways to pair them, and two choices for which pair gets a half twist, for a

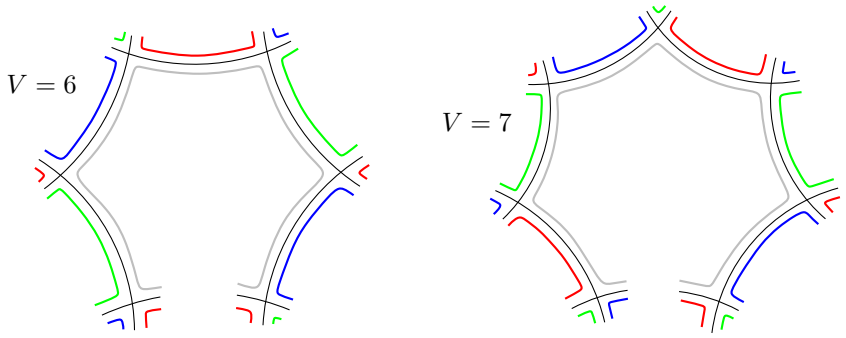


Figure 13: A representation of the ribbon graph H with 6 and 7 vertices. The ends of each long segment are glued with a half twist. This leaves four half-edges that need to be paired up.

total of four choices (see Figure 14). However, some of these choices yield isomorphic objects. To see this, color the four boundary components of H gray, red, green and blue. In the planar representation, H has $2V + 2$ ends and $2V + 2$ gaps between these ends, one of which is on the inside. Each outer gap is connected (via half twists at the ends of extended sides) to the third next gap. This is why the residue of V modulo 3 is relevant.

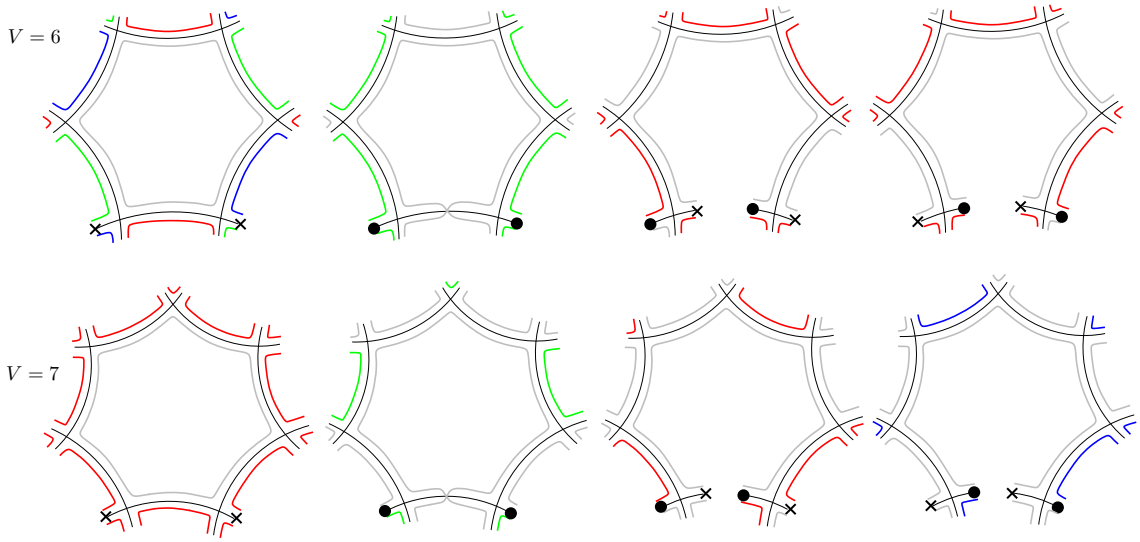


Figure 14: The four admissible pairings with 6 and 7 vertices. The crosses indicate a half twist whereas the dots indicate a lack thereof. Sides with the same color belong to the same boundary component.

To fix ideas, color the inner gap gray and the first outer gap, as well as those that it is connected to, in red. Similarly, color the other two boundary components green and blue. The last outer gap gets colored red if and only if $2V$ (hence V) is a multiple of 3. Assume this is the case. Then at each of the first and last vertices there is one free half-edge with one side gray and one side red, and one free half-edge with one side green and one side

blue.

As indicated earlier, there are four ways to close up H:

- If we glue gray to gray and red to red, then green gets glued to green and blue to blue. The resulting ribbon graph G has four boundary components of length V each.
- If we glue gray to red, then green gets glued to blue. The resulting ribbon graph G has two boundary components of length $2V$ each.
- If we glue gray to blue, then red gets glued to green. The resulting ribbon graph G has two boundary components of length $2V$ each.
- If we glue gray to green, then red gets glued to blue. The resulting ribbon graph G has two boundary components of length $2V$ each.

The four possibilities are depicted on the first row of Figure 14 for $V = 6$. One can check that the last three ribbon graphs are all isomorphic via cut-and-paste, so we indeed get two distinct isomorphism classes.

Suppose that V is not a multiple of 3. Then if two colors are on two sides of the same free half-edge of the first vertex in H, they are on different free half-edges of the last vertex and vice versa. In this case, it is not possible to glue each color to itself, nor is it possible to connect the colors in two pairs. Whichever of the four admissible gluings we pick, one color closes up while the three other colors connect together (see the second row of Figure 14 for $V = 7$). That is to say, any ribbon graph G as above with $V \neq 0 \mod 3$ vertices has one boundary component of length V and one boundary component of length $3V$. This implies that we can represent G as a tubular neighborhood of a regular V-gon in the plane with sides extended and all half twists on the outside (as in the first column of Figure 14). In other words, there is only one isomorphism class.

Remark 2.14. In the sequel, we will not distinguish between the case $n = 2$ and $n \geq 3$. That is, we will abuse notation and speak of the surface $X(\Gamma)$ for a 2-regular signed graph Γ. In those instances, one should take $X(\Gamma)$ to be any of the surfaces $X(G)$ for graphs G as above with the same number of vertices as Γ.

2.14 Systoles

We will show that the systoles of the surface $X(\Gamma)$ defined above are the a-, b- and c-curves, provided that Γ has sufficiently large girth. The *girth* of a graph is defined as the length of its shortest non-trivial loop. The problem with graphs with small girth is that the seams of the crosses in $X(\Gamma)$ can close up to form short geodesics. The following lemma shows that the seams are indeed the main thing to worry about.

Lemma 2.15. *Let $t > 0$. The shortest non-trivial arcs in the cross $C(t)$ (defined in Section 2.2) with endpoints in the boundary are the seams.*

Proof. The only potential candidates for shortest arcs are the seams or the other axes of symmetry of $C(t)$. Indeed, any arc that intersects one of the loci of reflection can be shortened by surgery with its reflection unless it coincides with the latter. Moreover, these loci cut the cross into topological disks. Since each seam is homotopic to a surgery on one horizontal and one vertical axis, the seams are the shortest arcs. □

Recall that the length of the seams in $C(t)$ is $\sigma(t) = \operatorname{arccosh}(\coth^2 t)$. We will also refer to the number t_n defined by Lemma 2.5 as well as $a(t_n)$, where

$$a(t) = 2\operatorname{arccosh}(\cosh(t)\cosh(n\operatorname{arcsinh}(\coth t)))$$

according to Equations (2.7) and (2.6). We will estimate these quantities in Section 6.1. We can now prove the main result of this section.

Theorem 2.16. *Let $n \geq 2$ and let Γ be a connected, n-regular, signed graph of girth larger than $a(t_n)/\sigma(t_n)$. Then the systoles in the surface $X(\Gamma)$ are the a-, b- and c-curves, which have length $a(t_n)$. If Γ is finite, then the genus g of $X(\Gamma)$ is equal to $E+1$ where E is the number of edges in Γ and there are $(12g-12)$ systoles in $X(\Gamma)$.*

Remark 2.17. The girth of a tree is infinite by convention, hence Theorem 2.16 generalizes Proposition 2.13.

Proof. Let γ be a systole of $X(\Gamma)$. We define the shadow $s(\gamma)$ in the graph Γ in the same way as in the proof of Proposition 2.13. If $s(\gamma)$ is non-contractible in Γ, then it traverses more than $a(t_n)/\sigma(t_n)$ edges by hypothesis. This means that γ traverses as many crosses, hence is longer than

$$(a(t_n)/\sigma(t_n)) \cdot \sigma(t_n) = a(t_n)$$

by Lemma 2.15, contradiction. It follows that $s(\gamma)$ is a contractible loop, so that it lifts to the universal cover of Γ, the n-regular tree $T(n)$. The tree of rings $\Sigma(n)$ similarly covers $X(\Gamma)$ and γ lifts to $\Sigma(n)$. By Proposition 2.13, any lift $\widetilde{\gamma}$ is one of the curves of type a, b or c in a ring or a pair of transverse rings of $\Sigma(n)$. Since the covering map $\Sigma(n) \to X(\Gamma)$ is injective on each ring and each pair of transverse rings, γ itself is an a-, b- or c-curve.

Let g be the genus of $X(\Gamma)$. There are $4n$ curves of type a or b per ring, n crosses per ring, and 2 rings per cross, hence 8 such curves per cross. Since each cross has Euler characteristic -2, there are $(g-1)$ crosses in $X(\Gamma)$, hence $(8g-8)$ curves of type a or b in total. Since each cross is central to exactly one pair of transverse rings and there are four c-curves per pair, the number of c-curves is equal to $(4g-4)$. Thus, the total number of systoles is $(12g-12)$. By construction, the number of crosses is equal to the number E of edges in Γ so that $g = E+1$. Note that the number V of vertices in Γ satisfies $nV = 2E$ since Γ is regular of degree n, so an alternative formulation is $g = 1 + nV/2$. □

Remark 2.18. The number L_n in Theorem 1.1 from the introduction is defined as $a(t_n)$.

As we will see in subsection 6.1, the quantity $a(t_n)/\sigma(t_n)$ grows exponentially with n. Therefore the girth of Γ—and hence the genus of $X(\Gamma)$—has to be very large for the above result to hold. The first order of business, however, is to show that the surfaces obtained are local maxima of the systole function. This is proved in the next two sections.

3 The lengths of the systoles determine the surface locally

In this section, we show that the systoles in $X(\Gamma)$ can detect any infinitesimal movement, that is, the derivative of their length is injective on the tangent space to Teichmüller space.

3.1 Twist deformations

Given a simple closed geodesic β in a hyperbolic surface X, we denote by τ_β the infinitesimal Fenchel-Nielsen twist deformation along β. More precisely,

$$\tau_\beta = \left.\frac{d}{dt}\right|_{t=0} X_t$$

where X_t is the surface obtained by cutting X along β, twisting distance t to the left, then regluing. Given any closed geodesic $\alpha \subset X$, the *cosine formula* says that

$$\frac{\partial \ell_\alpha}{\partial \tau_\beta} = \sum_{p \in \alpha \cap \beta} \cos \angle_p(\alpha, \beta) \tag{3.1}$$

where $\angle_p(\alpha, \beta)$ is the counter-clockwise angle from α to β at the point p [54, 38].

For every $n \geq 1$, the systoles in the ring $R(n, t_n)$ include the curves $a_1, \ldots, a_{2n}$ and $b_1, \ldots, b_{2n}$ by Proposition 2.3 and Proposition 2.12. We want to compute the effect of twisting around any of these curves on the length of any of them. To this end, let M be the $4n \times 4n$ matrix whose (i, j)-th entry is the derivative of the length of the i-th curve in the set $S = \{a_1, \ldots, a_{2n}, b_1, \ldots, b_{2n}\}$ with respect to the twist deformation along the $(2n + j)$-th curve (modulo $4n$) in S. Recall that the a-curves are pairwise disjoint, as are the b-curves, and that each a_i intersects each b_j exactly once (see Figure 4). The cosine formula (3.1) thus gives

$$M_{i,j} = \begin{cases} \cos \angle(a_i, b_j) & \text{if } i, j \in \{1, \ldots, 2n\} \\ \cos \angle(b_i, a_j) & \text{if } i, j \in \{2n+1, \ldots, 4n\} \\ 0 & \text{otherwise.} \end{cases}$$

In other words, M is block diagonal of the form

$$M = \begin{pmatrix} A & 0 \\ 0 & -A^\intercal \end{pmatrix}.$$

In particular, M is invertible if and only if A is. In the following two subsections we will show that A (and hence M) is indeed invertible.

Proposition 3.1. *For any $n \geq 1$, the matrix M of derivatives of lengths of a- and b-curves in the ring $R(n, t_n)$ with respect to the twist deformations around these curves has full rank.*

An immediate consequence is that the twists deformations around the a- and b-curves form a basis of the tangent space to the Teichmüller space of the ring.

Corollary 3.2. *The twist deformations around the a- and b-curves in the ring $R(n, t_n)$ form a basis of the tangent space to the Teichmüller space of $R(n, t_n)$ with fixed boundary lengths for any $n \geq 1$.*

Proof. The ring R is a surface of genus 1 with $2n$ boundary components. As such, it admits a pants decomposition with $2n$ interior curves. The Fenchel–Nielsen coordinates for these interior curves parametrize the Teichmüller space with fixed boundary lengths. Hence the latter has dimension $4n$, as does its tangent space at the point R. By Proposition 3.1, the twist deformations about the a- and b-curves in R are linearly independent. Since there are $4n$ such curves, these tangent vectors form a basis of the tangent space. □

In order to prove that the matrix A of cosines of angles has full rank, we need to estimate these angles. It turns out that each column in A has one entry close to 1 and the other entries fairly close to -1. The intuition for this is that since the a- and b-curves are each a union of two hypothenuses of right triangles with very long sides, they fellow travel the f-curves and the e-curve in the notation from Section 2.4. From this pattern, we will deduce that 0 is not an eigenvalue of A.

3.2 Angle estimate

Let $\theta = \theta(n)$ be the angle from e to any of the curves a_j in the ring $R(n, t_n)$. Then the angle from any b_j to e is also equal to θ. Also let ϕ_j be the counter-clockwise angle from a_j to b_1. Recall that there are $2n$ curves a_j that are images of one another by the shift $\eta : R(n, t_n) \to R(n, t_n)$ which translates by distance $e/2n$ along the curve e. In particular, the curves a_j intersect e at regularly spaced intervals of length $e/2n$ each. Therefore b_1, e and a_j together bound an isoceles triangle whose base has length $|n+1-j| \cdot e/2n$, whose angles at the base are equal to θ and whose third angle is equal to ϕ_j (see Figure 15). This holds for every $j \in \{1, \dots, 2n\}$ except for $j = n + 1$, where we get a triple intersection between b_1, e and a_{j+1}.

By dropping the altitude in each isoceles triangle and applying Equation (2.2) we obtain

$$\cos \frac{\phi_j}{2} = \sin \theta \cosh \left((n + 1 - j) \frac{e}{4n} \right).$$

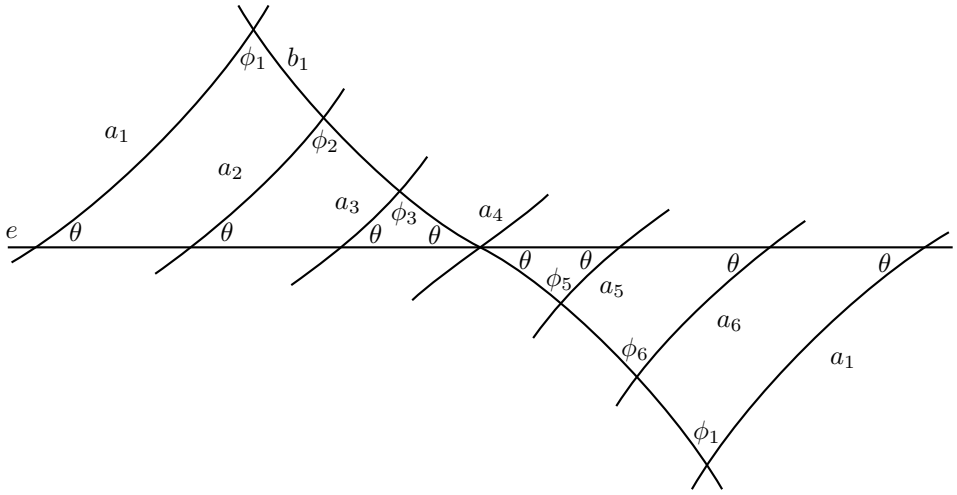


Figure 15: The isoceles triangles bounded by the curves b_1, e and a_j in the ring $R(n, t_n)$ for $n = 3$

The double angle formula for cosine yields

$$\cos\phi_j = 2\sin^2\theta\cosh^2\left((n+1-j)\frac{e}{4n}\right) - 1. \tag{3.2}$$

Observe that the formula holds for $j = n+1$ as well since $\phi_{n+1} + 2\theta = \pi$ so that

$$\cos\phi_{n+1} = \cos(\pi - 2\theta) = -\cos(2\theta) = 2\sin^2\theta - 1.$$

We will show that the first angle ϕ_1 is very small whereas the following angles ϕ_j are close to π. We first need an elementary inequality involving sums of hyperbolic cosines.

Lemma 3.3. *For any $n \geq 1$ and any $x > \operatorname{arccosh}(\sqrt{2})$ we have*

$$2\sum_{k=0}^{n-1}\cosh^2(kx) < \cosh^2(nx).$$

Proof. We proceed by induction on n. For $n = 1$, the inequality reduces to $2 < \cosh^2(x)$ which is true by hypothesis. Now suppose that

$$2\sum_{k=0}^{n-1}\cosh^2(kx) < \cosh^2(nx)$$

for some $n \geq 1$. Then

$$2\sum_{k=0}^{n}\cosh^2(kx) < 3\cosh^2(nx) \tag{3.3}$$

which we want to show is less than $\cosh^2((n+1)x)$.

The addition formula for hyperbolic cosine gives

$$\begin{aligned}\cosh((n+1)x) &= \cosh(nx)\cosh(x) + \sinh(nx)\sinh(x)\\ &> \sqrt{2}\,\cosh(nx) + \sinh(nx)\\ &= \left\{\sqrt{2} + \tanh(nx)\right\}\cosh(nx)\\ &> \left\{\sqrt{2} + \tanh(\operatorname{arcsinh}(1))\right\}\cosh(nx)\\ &= \left(\sqrt{2} + \frac{1}{\sqrt{2}}\right)\cosh(nx)\\ &> \sqrt{3}\,\cosh(nx)\end{aligned}$$

where we used the fact that $nx \geq x > \operatorname{arccosh}\left(\sqrt{2}\right) = \operatorname{arcsinh}(1)$. Putting this back in Equation (3.3) gives

$$2\sum_{k=0}^{n}\cosh^2(kx) < 3\cosh^2(nx) < \cosh^2((n+1)x).$$

By induction, the inequality holds for any $n \geq 1$. □

We can now show that the first angle ϕ_1 is much closer to 0 than any of the other angles, which are all close to π. The precise statement is expressed in terms of the cosines of the angles.

Lemma 3.4. *For any $n \geq 1$, the angles ϕ_j from a_j to b_1 satisfy*

$$\sum_{j=2}^{2n}(\cos\phi_j + 1) < (\cos\phi_1 + 1).$$

Proof. By Equation (3.2) this inequality is equivalent to

$$\sum_{j=2}^{2n}\cosh^2\left((n+1-j)\frac{e}{4n}\right) < \cosh^2(e/4). \tag{3.4}$$

Each summand on the left appears twice except for $j = n+1$ so that

$$\sum_{j=2}^{2n}\cosh^2\left((n+1-j)\frac{e}{4n}\right) < 2\sum_{k=0}^{n-1}\cosh^2\left(k\frac{e}{4n}\right).$$

Recall that $e = 4n\operatorname{arcsinh}(\coth(t_n)) > 4n\operatorname{arcsinh}(1)$ and hence

$$\frac{e}{4n} > \operatorname{arcsinh}(1) = \operatorname{arccosh}\left(\sqrt{2}\right).$$

We can therefore apply Lemma 3.3 with $x = e/4n$ to obtain the desired inequality

$$\sum_{j=2}^{2n}\cosh^2\left((n+1-j)\frac{e}{4n}\right) < 2\sum_{k=0}^{n-1}\cosh^2\left(k\frac{e}{4n}\right) < \cosh^2(e/4).$$

□

Corollary 3.5. *For any $n \geq 1$, the angles ϕ_j from a_j to b_1 satisfy*

$$\sum_{j=1}^{2n} \cos \phi_j \neq 0.$$

Proof. First assume that $n \geq 2$. The above statement is equivalent to

$$\sum_{j=1}^{2n} (\cos \phi_j + 1) \neq 2n.$$

By the previous lemma we have $\sum_{j=1}^{2n}(\cos \phi_j + 1) < 2(\cos \phi_1 + 1) < 4 \leq 2n$.

If $n = 1$, then a_1 meets b_1 at right angle since both of them intersect the f-curve with angle $\pi/4$. Furthermore, $\phi_2 = 3\pi/4$ (see Figure 10). Therefore

$$\cos \phi_1 + \cos \phi_2 = -\sqrt{2}/2 \neq 0.$$

□

3.3 The Gershgorin circle theorem

If the diagonal entries of a matrix dominate the rest, then the matrix is invertible. More generally, one can deduce information about the location of the eigenvalues from the size of the entries [28].

Theorem 3.6 (Gershgorin)**.** *Let U be an $n \times n$ matrix with entries $u_{i,j}$. Then the eigenvalues of U are contained in the union of the closed disks with centers $u_{j,j}$ and radii $\sum_{i \neq j} |u_{i,j}|$. In particular, if $|u_{j,j}| > \sum_{i \neq j} |u_{i,j}|$ for every j, then U is invertible.*

The last sentence of the theorem is quite transparent: if $x \in \mathbb{R}^n$ is non-zero and x_j is its largest entry in absolute value, then x times the j-th column of U is non-zero by the triangle inequality.

We apply this criterion to the matrix $A + J$ where A is the matrix of cosines of angles from the a-curves to the b-curves in the ring $R(n, t_n)$ and J is the $2n \times 2n$ matrix whose entries are all equal to one.

Lemma 3.7. *The matrix $A + J$ is invertible for any $n \geq 1$.*

Proof. Note that the entries of $A + J$ are positive so we do not need to take absolute values. Lemma 3.4 shows that the first entry of the first column of $A + J$ is larger than the sum of the other entries in that column. By symmetry of the ring, the entries of A satisfy $A_{i+1,j+1} = A_{i,j}$ where the indices are taken modulo $2n$, and similarly for $A + J$. The Gershgorin circle theorem thus implies that $A + J$ is invertible. □

It is easy to deduce that A itself is invertible.

Lemma 3.8. *The matrix A of cosines of angles from the a-curves to the b-curves in the ring $R(n, t_n)$ is invertible for any $n \geq 1$.*

Proof. Let V be the orthogonal complement of the vector $u = (1, \dots, 1)^\intercal$ in $\mathbb{R}^{2n}$. Since the restriction of J to V is equal to zero, A and $A+J$ act the same way on V. Moreover, A and $A+J$ both send the span of u onto itself since they have constant non-zero row sums. The row sums are all the same because the rows are cyclic permutations of one another. The row sums of $A + J$ are non-zero because its entries are positive and the row sums of A are non-zero by Corollary 3.5. Since $A + J$ is surjective by the previous lemma, we obtain

$$\mathbb{R}^{2n} = (A + J)(\mathbb{R}^{2n}) = (A + J)(V + \operatorname{span} u) = A(V + \operatorname{span} u) = A(\mathbb{R}^{2n}).$$

We conclude that A itself is surjective, hence invertible. □

This implies that the full matrix M of cosines of angles between all systoles in the ring $R(n, t_n)$ is invertible.

Proof of Proposition 3.1. We have $\det(M) = \det(A)^2 \neq 0$ by the previous lemma. □

3.4 From rings to closed surfaces

Let $X = X(\Gamma)$ be a closed surface obtained by gluing crosses $C(t_n)$ along a connected, finite, n-regular, signed graph Γ of sufficiently large girth as in Theorem 2.16, so that its systoles are the a-, b- and c-curves (and f-curves if $n = 1$). We now prove that the lengths of these curves determine the surface at the infinitesimal level (perfection).

Theorem 3.9 (Perfection)**.** *Let $n \geq 1$ and let $X = X(\Gamma)$ where Γ is a finite, connected, n-regular, signed graph of girth larger than $a(t_n)/\sigma(t_n)$. The map $\mathcal{T}(X) \to \mathbb{R}_+^S$ sending Y to the vector of lengths $(\ell_\gamma(Y))_{\gamma \in S}$ where S is the set of systoles in X has injective derivative at the point X.*

Proof. We need to show that the differentials $\{d\ell_\gamma\}_{\gamma \in S}$ span the cotangent space $T_X^*\mathcal{T}(X)$ over $\mathbb{R}$. Wolpert's twist-length duality [55, Theorem 2.10] states that

$$d\ell_\gamma = \sqrt{-1}\left(\frac{\partial}{\partial \tau_\gamma}\right)^*$$

for any simple closed geodesic γ, where the dual is taken with respect to the Weil–Petersson metric. Therefore, the length differentials $\{d\ell_\gamma\}_{\gamma \in S}$ span the cotangent space $T_X^*\mathcal{T}(X)$ if and only if the twist deformations $\{\partial/\partial\tau_\gamma\}_{\gamma \in S}$ span the tangent space $T_X\mathcal{T}(X)$.

By Proposition 2.12 (for $n = 1$) and Theorem 2.16 (for $n \geq 2$), the set S of systoles includes the a- and b-curves. We will show that the twist deformations about the a- and b-curves generate the tangent space. To see this, observe that there exists a pants decomposition $\mathcal{P}$ of X consisting entirely of curves that are each in the interior of some ring $R \subset X$. For instance, one can take $\mathcal{P}$ to be the set of all f-curves in X (the curves that cut X into crosses) together with one curve in each cross that separates it into two pairs of pants—call these d-curves. Each d-curve is in the interior of both rings that it intersects, while each f-curve is in the interior of a unique ring.

The lengths and twists around the curves in the pants decomposition $\mathcal{P}$ define Fenchel–Nielsen coordinates $\mathcal{T}(X) \to (\mathbb{R}_+ \times \mathbb{R})^{\mathcal{P}}$ once a convention is chosen for what zero twist means. For any curve $\alpha \in \mathcal{P}$, let $R \subset X$ be a ring that contains α in its interior. By Corollary 3.2, the twist deformations about the a- and b-curves in R generate the tangent space to the Teichmüller space of R with fixed boundary lengths. In particular, the two tangent vectors corresponding to changing the length or twist parameter of α at unit speed while keeping all the other Fenchel–Nielsen coordinates fixed are in the span of the twist deformations around the a- and b-curves in X. Since the Fenchel–Nielsen length and twist parameters define a smooth coordinate system for $\mathcal{T}(X)$, we are done. □

Remark 3.10. The proof actually shows that the derivative of the vector of lengths of all the a-curves and b-curves is injective at $X(\Gamma)$. The c-curves are not needed for this; they only play a role in the next section.

Remark 3.11. In [48], Schmutz Schaller describes a collection of $(6g - 5)$ curves such that their lengths define a topological embedding of Teichmüller space into $\mathbb{R}_+^{6g-5}$. See also [30, 31]. If g is the genus of $X(\Gamma)$, then there are $(8g - 8)$ curves of type a or b in $X(\Gamma)$. We do not know if their lengths define a global embedding of Teichmüller space, but Theorem 3.9 in conjunction with the inverse function theorem implies that they define an embedding in a neighborhood of $X(\Gamma)$.

4 The systole decreases under all deformations

Let $X = X(\Gamma)$ where Γ is a signed graph satisfying the hypotheses of Theorem 3.9. Now that we know that the systoles in X can detect any infinitesimal movement, it remains to show that at least one of them shrinks under any infinitesimal deformation (eutaxy). Even though we have proved that the twist deformations around the a-curves and b-curves in X generate the tangent space $T_X\mathcal{T}(X)$, it will be convenient to use a different basis to show this.

To define this other basis, we first explain how it acts on individual crosses. Let C be a cross with four boundary lengths equal to $4t_n$ as in Section 2.2. For each boundary component $\beta \subset C$ and $s > 0$, we define the deformed cross C_s^β to be the four-holed

sphere with β-boundary of length $4(t_n + s)$, the three other boundaries of length t_n, and with the same symmetries fixing β that C has, that is, a $\mathbb{Z}_2 \times \mathbb{Z}_2$ group generated by a front-to-back reflection and a left-and-right or top-to-bottom reflection depending on whether β is medial or lateral respectively.

For example, if β is the left boundary component of C, then C_s^β is obtained by taking a right-angled hexagon with left side $t_n + s$, top side of length $2t_n$ and right side of length t_n, then reflecting this hexagon across its bottom side to obtain a right-angled octagon, then doubling this octagon across the four sides with unspecified lengths (see Figure 16).

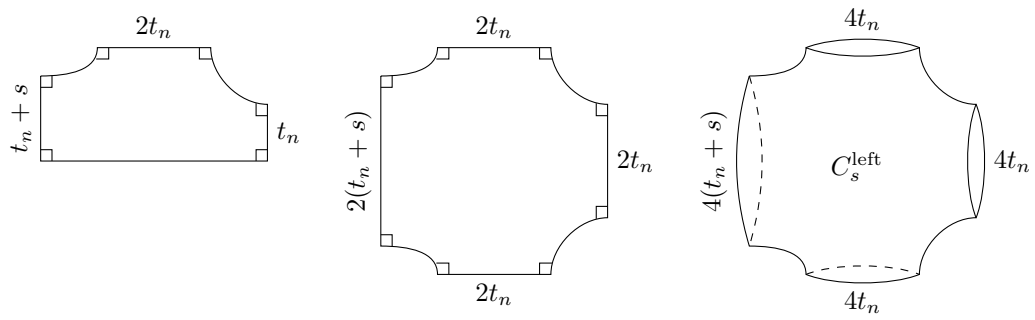


Figure 16: The length deformation of the cross about its left boundary

Now if $\beta \subset X$ is any f-curve and $s > 0$, then we define X_s^β to be the same as X but with the two crosses C and D adjacent to β replaced with C_s^β and D_s^β. These are glued together and with the other crosses in the most obvious way, without twisting. Finally, we let

$$\lambda_\beta = \left.\frac{d}{ds}\right|_{s=0} X_s^\beta$$

and call this the *symmetric length deformation about* β.

The motivation behind this construction is that the only canonical way to change the length of a curve on a surface is to flow along its gradient with respect to the Weil–Petersson metric. However, this gradient deformation is non-local in nature and its effect on the lengths of other curves (especially disjoint ones) is complicated to compute, although an explicit formula analogous to the cosine formula (3.1) exists [47, Equation (7)].

The advantage of our symmetric length deformations is that the sum

$$\Lambda = \sum_{\beta \in \{f\text{-curves}\}} \lambda_\beta$$

corresponds to expanding all the boundaries of all the crosses in X at the same rate without twisting and while preserving the symmetries of all the crosses. In other words, the effect of Λ is the same as increasing the parameter t at unit speed in the definition of the

ring $R(n, t_n)$. In particular, for any a- or b-curve α and for any c-curve γ in X we have

$$\frac{\partial \ell_\alpha}{\partial \Lambda} = a'(t_n) > 0 \quad \text{and} \quad \frac{\partial \ell_\gamma}{\partial \Lambda} = c'(t_n) < 0 \tag{4.1}$$

according to Lemma 2.8 and Lemma 2.4 respectively.

To complete our basis for the tangent space $T_X\mathcal{T}(X)$ we also include the twist deformations around the f-curves as well as two more twist deformations τ_d and τ_h per cross. In the cross C, we pick the curve d to be one of the two diagonal axes of symmetry and h to be the curve depicted in Figure 17.

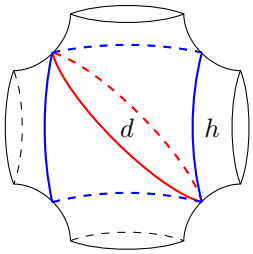


Figure 17: The curve d (in red) and the curve h (in blue) in the cross C

Observe that d and h intersect twice. Moreover, the oriented angles from d to h at the two intersection points are equal to each other since the rotation of angle π about the centers of the front and back of the cross leaves each curve invariant, preserves orientation, and exchanges the two intersection points. Finally, the angle of intersection $\psi = \angle(d, h)$ is different from $\pi/2$ since the intersections occur at the midpoints of two opposite seams, and the seams are orthogonal to d at those points.

Let L, R, T, B be the left, right, top and bottom boundaries of the cross respectively. The matrix of partial derivatives of the lengths of $\{L, R, T, B, d, h\}$ with respect to the deformations $\{\lambda_L, \lambda_R, \lambda_T, \lambda_B, \tau_h, \tau_d\}$ of the cross C has the form

$$\begin{array}{c} \\ d\ell_L \\ d\ell_R \\ d\ell_T \\ d\ell_B \\ d\ell_d \\ d\ell_h \end{array} \begin{array}{c} \begin{array}{cccccc} \lambda_L & \lambda_R & \lambda_T & \lambda_B & \tau_h & \tau_d \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \delta & \delta & \delta & \delta & 2\cos\psi & 0 \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & 0 & -2\cos\psi \end{pmatrix} \end{array}$$

for some $\delta, \varepsilon \in \mathbb{R}$. It is lower triangular with non-zero diagonal entries, hence invertible. In particular, the deformations $\{\lambda_L, \lambda_R, \lambda_T, \lambda_B, \tau_h, \tau_d\}$ form a basis of the tangent space to the Teichmüller space of C with variable boundary lengths.

Lemma 4.1. *The symmetric length deformations* $\{\lambda_\beta\}_{\beta \in \{f\text{-curves}\}}$ *together with the twist deformations* $\{\tau_\beta\}_{\beta \in \{f\text{-curves}\}}$ *and* $\{\tau_{d(C)}, \tau_{h(C)}\}_{C \in \{\text{crosses}\}}$ *form a basis of* $T_X\mathcal{T}(X)$.

Proof. It suffices to prove that these deformations span the tangent space since they are equal in number to its dimension. By the paragraph preceding the statement of this lemma, these deformations generate all the deformations of any cross in X. In particular, they generate the Fenchel–Nielsen length and twist deformations with respect to the pants decomposition of X by f-curves and d-curves. □

We can now prove that the systole function decreases under all non-trivial deformations of X, hence that X is a local maximum of sys.

Theorem 4.2 (Eutaxy)**.** *Let $n \geq 1$ and let $X = X(\Gamma)$ where Γ is a finite, connected, n-regular, signed graph of girth larger than $a(t_n)/\sigma(t_n)$. Then for every non-zero tangent vector $v \in T_X\mathcal{T}(X)$, there is at least one systole α of X such that $d\ell_\alpha(v) < 0$. That is, X is a local maximum of the systole function.*

Proof. First assume that $n \geq 2$. Let S be the set of systoles of X and suppose that $v \in T_X\mathcal{T}(X)$ is such that $d\ell_\alpha(v) \geq 0$ for every $\alpha \in S$. We will show that $v = 0$.

By the previous lemma, we can write

$$v = \sum_{\beta \in F} \kappa_\beta \cdot \lambda_\beta + \sum_{\gamma \in D} \mu_\gamma \cdot \tau_\gamma$$

for some $\kappa_\beta, \mu_\gamma \in \mathbb{R}$ where F is the set of f-curves and D is the set of all f-, d- and h-curves in X.

For every $\alpha \in S$ we have

$$0 \leq d\ell_\alpha(v) = \sum_{\beta \in F} \kappa_\beta \cdot d\ell_\alpha(\lambda_\beta) + \sum_{\gamma \in D} \mu_\gamma \cdot d\ell_\alpha(\tau_\gamma).$$

Summing over any subset $Q \subset S$ we obtain

$$0 \leq \sum_{\alpha \in Q} d\ell_\alpha(v) = \sum_{\beta \in F} \kappa_\beta \sum_{\alpha \in Q} d\ell_\alpha(\lambda_\beta) + \sum_{\gamma \in D} \mu_\gamma \sum_{\alpha \in Q} d\ell_\alpha(\tau_\gamma).$$

Let A be the set a- and b-curves in X and let C be the set of c-curves in X. The **first observation** is that if Q is equal to either A or C, then $\sum_{\alpha \in Q} d\ell_\alpha(\tau_\gamma) = 0$ for every $\gamma \in D$. Indeed, for every systole $\alpha \in Q$ intersecting γ, there is some systole $\alpha^* \in Q$ intersecting γ with the supplementary angle. To see this, observe that in the tree of rings $\Sigma(n)$ there is an orientation-reversing isometry which sends (any lift of) γ to itself and permutes the (lifts of) a- and b-curves, as well as the (lifts of) c-curves separately. For instance, if γ is an f- or d-curve then the reflection of $\Sigma(n)$ across the seams works, and if γ is an h-curve then the left-to-right reflection of the cross containing γ extends to an isometry of $\Sigma(n)$. These reflections exchange a- and b-curves and preserve the set of c-curves. Hence the statement about angles coming in supplementary pairs holds in the tree

of rings $\Sigma(n)$. Since $\Sigma(n)$ covers X and every systole in X is the image of a systole in $\Sigma(n)$, the statement holds in X as well. The cosine formula (3.1) thus implies that the total length variation of the curves in Q is nil in the direction of τ_γ. Thus

$$\sum_{\alpha\in A} d\ell_\alpha(v) = \sum_{\beta\in F}\kappa_\beta \sum_{\alpha\in A} d\ell_\alpha(\lambda_\beta) \quad \text{and} \quad \sum_{\alpha\in C} d\ell_\alpha(v) = \sum_{\beta\in F}\kappa_\beta \sum_{\alpha\in C} d\ell_\alpha(\lambda_\beta). \tag{4.2}$$

The **second observation** is that the term

$$\sum_{\alpha\in Q} d\ell_\alpha(\lambda_\beta)$$

is independent of $\beta \in F$ when Q is equal to either A or C. Indeed, the deformation λ_β only affects the lengths of systoles in pairs of transverse rings containing one of the two crosses adjacent to β. The geometry of the subsurface $Y \subset X$ containing all these pairs of transverse rings does not depend on β. This is because Y is the union of the crosses corresponding to the edges of a subgraph $H \subset \Gamma$, namely, the 2-neighborhood of a pair of consecutive edges in the cyclic order around a vertex (corresponding to the crosses meeting along β). Since Γ is assumed to have girth larger than $a(t_n)/\sigma(t_n)$, and that number is bigger than 6 (see Table 1), H is a tree isometric to the 2-neighborhood of *any* pair of consecutive edges in Γ (or in the n-regular tree). The resulting subsurface Y and the total effect of the deformation λ_β on the length of its a- and b-curves or its c-curves is therefore independent of β.

The **third and last observation** is that

$$\sum_{\beta\in F}\left(\sum_{\alpha\in A} d\ell_\alpha(\lambda_\beta)\right) = \sum_{\alpha\in A}\left(\sum_{\beta\in F} d\ell_\alpha(\lambda_\beta)\right) = \sum_{\alpha\in A} d\ell_\alpha(\Lambda) = \sum_{\alpha\in A} a'(t_n) > 0$$

by Equation (4.1), where $\Lambda = \sum_{\beta\in F}\lambda_\beta$. We deduce that $\sum_{\alpha\in A} d\ell_\alpha(\lambda_\beta) > 0$ for any $\beta \in F$ by the second observation. Similarly,

$$\sum_{\beta\in F}\left(\sum_{\alpha\in C} d\ell_\alpha(\lambda_\beta)\right) = \sum_{\alpha\in C}\left(\sum_{\beta\in F} d\ell_\alpha(\lambda_\beta)\right) = \sum_{\alpha\in C} d\ell_\alpha(\Lambda) = \sum_{\alpha\in C} c'(t_n) < 0$$

so that $\sum_{\alpha\in C} d\ell_\alpha(\lambda_\beta) < 0$ for any $\beta \in F$.

For any fixed $\gamma \in F$ we have both

$$0 \le \sum_{\alpha\in A} d\ell_\alpha(v) = \left(\sum_{\beta\in F}\kappa_\beta\right)\left(\sum_{\alpha\in A} d\ell_\alpha(\lambda_\gamma)\right)$$

and

$$0 \le \sum_{\alpha\in C} d\ell_\alpha(v) = \left(\sum_{\beta\in F}\kappa_\beta\right)\left(\sum_{\alpha\in C} d\ell_\alpha(\lambda_\gamma)\right)$$

from Equation (4.2) and the second observation. By the third observation, the sum in the rightmost parentheses is first positive then negative. We conclude that $\sum_{\beta\in F}\kappa_\beta = 0$ so that

$$\sum_{\alpha\in A} d\ell_\alpha(v) = 0 = \sum_{\alpha\in C} d\ell_\alpha(v).$$

Since each summand was assumed to be non-negative, they are all zero. By Theorem 3.9, this implies that $v = 0$.

If $n = 1$, the same argument works with A replaced by F. The point is that the f-curves are systoles in this case and their length increases under the deformation Λ.

It is easy to see that the first part of the theorem implies that X is a local maximum of the systole function. For any unit vector $v \in T_X\mathcal{T}(X)$, the above implies that there is a curve $\alpha \in S$ and an $\varepsilon_v > 0$ such that

$$\operatorname{sys}(\exp_X(tv)) \leq \ell_\alpha(\exp_X(tv)) < \ell_\alpha(X) = \operatorname{sys}(X)$$

for every $t \in (0, \varepsilon_v)$, where $\exp_X(tv)$ is the point at distance t from X along the Weil–Petersson geodesic in the direction of v. Since the length functions $(\ell_\alpha)_{\alpha\in S}$ are continuously differentiable, ε_v can be chosen locally uniformly with respect to v. As the unit sphere in $T_X\mathcal{T}(X)$ is compact, there is an $\varepsilon > 0$ which works for all v. Hence there is a neighborhood U of X in $\mathcal{T}(X)$ such that $\operatorname{sys}(Y) \leq \operatorname{sys}(X)$ for every $Y \in U$ with equality only if $Y = X$. The same holds in moduli space. □

5 Isometries are induced by graph isomorphisms

In this section, we show that distinct signed graphs Γ give rise to distinct oriented hyperbolic surfaces $X(\Gamma)$. As a byproduct, we get that if the underlying graph has no non-trivial automorphism, then the resulting surface has a trivial group of orientation-preserving isometries.

We first need to distinguish between the different kinds of systoles in $X(\Gamma)$.

Lemma 5.1. *Let $n \geq 2$ and let Γ be an n-regular signed graph such that the systoles in $X(\Gamma)$ are the a-, b- and c-curves. Then the a- and b-curves in $X(\Gamma)$ intersect a different number of systoles than the c-curves.*

Proof. Let us count the number of systoles that intersect a given c-curve γ. In the pair of transverse rings $R_1 \cup R_2$ containing γ, there is a central cross at the intersection of the two rings and $2(n-1)$ non-central crosses. For each of the latter kind, γ intersects only one side (front or back) of the cross, separating two opposite sides of that octagon. Thus for each non-central cross C, in the ring through C distinct from R_1 and R_2, exactly half of the a- and b-curves intersect γ. This is because each a- and b-curve intersects only one

side of each cross, connecting two opposite sides of that octagon. These curves contribute $2(n-1)\cdot 2n$ intersections.

As for the systoles in R_1 or R_2, again half of them intersect γ. To see this, observe that any a- or b-curve is homotopic to a union of two geodesic segments: one that travels halfway along an e-curve and one that travel halfway along an f-curve. The e-curves are disjoint from γ while each f-curve in R_j intersects γ once. Thus each f-curve in R_j contributes one a-curve and one b-curve intersecting γ. This yields a total of $2n+2n = 4n$ curves of type a or b in $R_1 \cup R_2$ that intersect γ.

How many c-curves intersect γ? We can first homotope any c-curve (including γ) to a union of two segments of e-curves. Each non-central cross in $R_1 \cup R_2$ is associated with four c-curves, half of which intersect γ. Indeed, when they are represented along the e-curves, any such c-curve ζ shares a segment I with γ. At the extremities of I, the two curves γ and ζ can turn toward either the same of different sides of I. In the first case we can homotope them to intersect only once while in the second we can homotope them to be disjoint. Thus there are $2\cdot 2(n-1)$ curves of type c that intersect γ coming from the $2(n-1)$ non-central crosses in $R_1 \cup R_2$.

Lastly, for each cross C contained in a ring that intersects $R_1 \cup R_2$ such that C is not itself contained in $R_1 \cup R_2$, we get two c-curves intersecting γ. There are $2(n-1)^2$ such crosses, accounting for $4(n-1)^2$ intersections.

Any other systole is disjoint from γ, being disjoint from $R_1 \cup R_2$. The total number of systoles intersecting γ is thus

$$4n(n-1) + 4n + 4(n-1) + 4(n-1)^2 = 8n^2 - 4n.$$

We now count the number of systoles that intersect a given a- or b-curve α. In the ring R where α lives, there are $2n$ systoles that intersect α apart from itself.

Since α intersects each cross of R in only one side (front or back) and separates two opposite sides of that octagon, it intersects exactly half of the a- and b-curves in each ring transverse to R. There are n such rings, each contributing $2n$ intersections with α.

These are all the a- and b-curves that α intersects. Now for the c-curves. By the above, α intersects half of the c-curves that intersect R. There are n crosses per ring transverse to R, n such rings, each contributing two c-curves that intersects α, for a total of $2n^2$.

The number of systoles intersecting α is equal to

$$2n + 2n^2 + 2n^2 = 4n^2 + 2n$$

which is distinct from $8n^2 - 4n$ for any $n \geq 2$ (the two real solutions are 0 and $3/2$). □

The next step is to pick out pairs of a- and b-curves that are *symmetric about the seams*, meaning that they are permuted by the reflection ρ_{seams}. In the notation of subsection 2.3, these are pairs a_j and b_j for some $j \in \{1, \ldots, 2n\}$. See Figure 4.

Lemma 5.2. *Let $n \geq 2$ and let Γ be an n-regular signed graph. Then a pair of intersecting a- and b-curves in $X(\Gamma)$ maximizes the number of intersections with other a- and b-curves if and only if it is symmetric about the seams.*

Proof. Consider a pair $\alpha \cup \beta$ of a- and b-curves such that $\beta = \rho_{\text{seams}}(\alpha)$. What is special about this pair is that for each cross it intersects, it intersects both of its sides (front and back). Let R be the ring containing $\alpha \cup \beta$. All the $4n$ systoles in R intersect $\alpha \cup \beta$. Furthermore, all the $4n$ systoles in each of the n rings transverse to R intersect the pair $\alpha \cup \beta$. The total number of intersections is $4n^2 + 4n$.

Now suppose that α and β are a- and b-curves contained in a common ring R but are not symmetric about the seams. Then there is some cross $C \subset R$ such that $\alpha \cup \beta$ intersects only one side of C. Hence in the ring transverse to R through C, only $2n$ systoles intersect $\alpha \cup \beta$, for a total of at most $4n^2 + 2n$ curves of type a or b.

Finally, suppose that α and β are not contained in a common ring. Let $R_1 \cup R_2$ be the pair of transverse rings containing them. In each ring transverse to R_j, there are $2n$ systoles that intersect $\alpha \cup \beta$ apart from α and β. Thus the number of a- and b-curves intersecting $\alpha \cup \beta$ is $2n \cdot n \cdot 2 + 2 = 4n^2 + 2$, which is less than $4n^2 + 4n$. □

We now have the required tools to prove that the map $\Gamma \mapsto X(\Gamma)$ is injective. We refer the reader back to subsection 2.10 for the definition of signed graphs and their isomorphisms, and to subsection 2.11 for the description of the map $\Gamma \mapsto X(\Gamma)$.

Theorem 5.3. *Let $n \geq 3$ and let Γ_1 and Γ_2 be n-regular signed graphs of girth larger than $a(t_n)/\sigma(t_n)$. Any orientation-preserving isometry $X(\Gamma_1) \to X(\Gamma_2)$ is induced by a unique isomorphism of signed graphs $\Gamma_1 \to \Gamma_2$.*

Proof. Let $\psi : X(\Gamma_1) \to X(\Gamma_2)$ be an orientation-preserving isometry. Then ψ sends systoles of $X(\Gamma_1)$ to systoles of $X(\Gamma_2)$. By Lemma 5.1, it sends the set of a- and b-curves on $X(\Gamma_1)$ to the set of a- and b-curves on $X(\Gamma_2)$. Furthermore, each pair of a- and b-curves in $X(\Gamma_1)$ that are symmetric about the seams in sent to a such a pair in $X(\Gamma_2)$ by Lemma 5.2.

The two angle bisectors of a symmetric pair of a- and b-curves at the intersection are along an f-curve and the seams. We may assume that the girth of Γ_1 and Γ_2 is larger than 2 (see subsection 6.1) so that the f-curves are distinguished from the seams. We conclude that ψ sends f-curves to f-curves and seams to seams. In particular, it respects the decomposition of $X(\Gamma_1)$ and $X(\Gamma_2)$ into crosses.

Let $E(\Gamma_j)$ be the set of edges of Γ_j. Since there is a bijection between the crosses in $X(\Gamma_j)$ and the edges in Γ_j, the isometry ψ induces a bijection $\phi : E(\Gamma_1) \to E(\Gamma_2)$. Since ψ maps adjacent crosses to adjacent crosses, the induced map ϕ either preserves or reverses the cyclic order around each vertex. After applying a set of vertex flips to

Γ_2 (which does not affect $X(\Gamma_2)$), we may assume that ϕ preserves cyclic orders. If two parallel rings in $X(\Gamma_1)$ have matching (resp. opposite) orderings, it is clear that ψ sends them to parallel rings with matching (resp. opposite) orderings. That is, the sign between any two consecutive edges p and q in Γ_1 is the same as the sign between $\phi(p)$ and $\phi(q)$ in Γ_2. In other words, $\Gamma_1 = \Gamma_2$ up to isomorphism. □

Remark 5.4. This statement is false for $n = 1$ for the simple reason that there is no distinction between the f-curves and the seams (the c-curves). The analogous statement for $n = 2$ is true (and the proof essentially identical) provided that we replace the signed graphs by the 4-regular ribbon graphs satisfying the conditions of subsection 2.13. By the argument in that subsection, there are two isomorphism classes of such graphs whenever the number V of vertices is a multiple of 3, and one isomorphism class otherwise. Therefore, we get two distinct corresponding points in $\mathcal{M}_g$ if $g = V + 1$ is congruent to 1 mod 3, and only one otherwise.

Corollary 5.5. *Let $n \geq 3$ and let Γ be an n-regular signed graph of girth larger than $a(t_n)/\sigma(t_n)$. If Γ has a trivial automorphism group, then $X(\Gamma)$ has a trivial group of orientation-preserving isometries.*

On the other hand, each surface $X(\Gamma)$ has at least one orientation-reversing isometry, namely the reflection across the seams.

6 Counting the number of examples in each genus

6.1 Length estimates

In this subsection, we quantify how large the girth of the signed graph Γ needs to be in terms of n for the hypothesis of Theorem 2.16 to be satisfied, that is, we estimate $a(t_n)/\sigma(t_n)$. In particular, we estimate the length $L_n = a(t_n)$ of the systoles of the resulting surface $X(\Gamma)$.

Lemma 6.1. *We have*

$$t_n = n \log\left(1 + \sqrt{2}\right) + o(1) \quad \textit{and} \quad L_n = a(t_n) = 4n \log\left(1 + \sqrt{2}\right) - 2\log 2 + o(1)$$

as $n \to \infty$.

Proof. Recall that the equality $a(t) = c(t)$ is equivalent to

$$\tanh(e(t)/4)\sinh(e(t)/4) = \cosh(t) \tag{6.1}$$

by the proof of Lemma 2.5, and that $e(t)/4 = n \operatorname{arcsinh}(\coth(t))$.

Let $\varepsilon > 0$ and let $\lambda = \log\left(1+\sqrt{2}\right) = \operatorname{arcsinh}(1)$. We will show that if n is large enough then the difference between the left-hand side (LHS) and the right-hand side (RHS) of Equation (6.1) switches sign when t is between $n\lambda - \varepsilon$ and $n\lambda + \varepsilon$.

First observe that $e(t)/4 > n \operatorname{arcsinh}(1) = n\lambda$ for every $t > 0$ and every n. Moreover, if $t \geq n\lambda - \varepsilon$, then $e(t)/4 \leq e(n\lambda - \varepsilon)/4 = n\lambda + o(1)$ as $n \to \infty$. Now

$$\tanh(x)\sinh(x) = \exp(x)/2 + o(1) \quad \text{and} \quad \cosh(x) = \exp(x)/2 + o(1)$$

as $x \to \infty$. Thus at $n\lambda - \varepsilon$ the LHS of (6.1) is at least $\exp(n\lambda)/2 + o(1)$ whereas the RHS is equal to $\exp(n\lambda - \varepsilon)/2 + o(1)$. So the RHS is smaller that the LHS at $n\lambda - \varepsilon$ if n is large enough. Similarly, the RHS is larger than the LHS at $n\lambda + \varepsilon$ if n is large enough. This shows that t_n is in the interval $(n\lambda - \varepsilon, n\lambda + \varepsilon)$ if n is large enough. Since $\varepsilon > 0$ was arbitrary, $t_n = n\lambda + o(1)$.

Recall that $a(t_n) = c(t_n)$ and $\cosh(c/2) = \sinh^2(e/4)$. Since

$$\operatorname{arccosh}(\sinh^2(x)) = 2x - \log 2 + o(1)$$

as $x \to \infty$ and $e(t_n)/4 = n\lambda + o(1)$ we obtain

$$a(t_n) = c(t_n) = 2\operatorname{arccosh}(\sinh^2(e(t_n)/4)) = 4n\lambda - 2\log 2 + o(1)$$

as $n \to \infty$. □

The next thing we need is an asymptotic lower bound for $\sigma(t_n)$.

Lemma 6.2. *We have $\sigma(t_n) \geq \left(1+\sqrt{2}\right)^{-n}$ if n is large enough.*

Proof. According to Equation (2.5) we have

$$\cosh(\sigma(t)) = \coth^2(t) = 1 + \frac{1}{\sinh^2(t)}$$

so that

$$\sinh^2(\sigma(t)/2) = \frac{\cosh(\sigma(t)) - 1}{2} = \frac{1}{2\sinh^2(t)} \geq 2\exp(-2t).$$

Now $x \geq \sinh(x/2)$ for every $x \in [0, 4.354]$. Morever $t_n \geq 1$ for every $n \geq 2$ (see the proof of Lemma 2.6), which implies that $\sigma(t_n) \leq \operatorname{arccosh}(\coth^2(1)) \approx 1.141$. We conclude that

$$\sigma(t_n) \geq \sinh(\sigma(t_n)/2) \geq \sqrt{2}\exp(-t_n) = \sqrt{2}\left(1+\sqrt{2}\right)^{-n+o(1)} \geq \left(1+\sqrt{2}\right)^{-n}$$

if n is large enough, where we used Lemma 6.1 for the equality sign. □

Remark 6.3. Actually, $\sigma(t_n)$ is closer to $2\sqrt{2}\left(1+\sqrt{2}\right)^{-n}$, but the above is all we need.

The previous two lemmata combined together yield the following estimate for the ratio of $a(t_n)$ over $\sigma(t_n)$. If the signed graph Γ has girth larger than this, then the systoles of $X(\Gamma)$ are the a-, b- and c-curves according to Theorem 2.16.

Corollary 6.4. *There is a constant $K > 0$ such that $a(t_n)/\sigma(t_n) \leq Kn\left(1+\sqrt{2}\right)^n$ for every $n \geq 2$.*

For small n, we can compute t_n, $a(t_n)$ and $\sigma(t_n)$ numerically to get a more explicit bound on the girth (see Table 1).

n	t_n	$a(t_n)$	$\sigma(t_n)$	$a(t_n)/\sigma(t_n)$
2	1.745752	5.909039	0.503760	11.729861
3	2.645975	9.256205	0.201312	45.979325
4	3.526946	12.731803	0.083188	153.048057

Table 1: Approximate values of t_n and $a(t_n)/\sigma(t_n)$ for small n

For $n = 2$, the girth of Γ needs to be at least 12, hence the genus of $X(\Gamma)$ at least 13. The minimal number V of vertices needed for a 3-regular graph Γ to have girth 46 is not known, but it is at least $2\sum_{j=0}^{22} 2^j = 16,777,214$ by the Moore bound. The corresponding surfaces $X(\Gamma)$ have genus at least $1+3V/2 = 25,165,822$. The genus required for $n = 4$ is at least $1 + 4\sum_{j=0}^{76} 3^j = 2 \cdot 3^{77} - 1$, which is astronomical.

6.2 Counting signed graphs

In this subsection, we give a lower bound for the number of isomorphism classes of connected, n-regular, signed graphs with $(g-1)$ edges, girth larger than $a(t_n)/\sigma(t_n)$, and trivial automorphism group for g sufficiently large. This concludes the proof of Theorem 1.1 from the introduction, which we restate more precisely as follows.

Theorem 6.5. *For $n, g \geq 3$, let $N(n,g)$ be the number of local maxima x of the systole function in $\mathcal{M}_g$ with $\mathrm{sys}(x) = L_n = a(t_n)$ whose group of orientation-preserving isometries is trivial. Then there exists a constant $\beta > 0$ such that for every $n \geq 3$, there is an $\alpha_n > 0$ such that if g is large enough and $2(g-1)/n$ is an integer, then*

$$N(n,g) \geq \alpha_n \left(\beta\, g\right)^{\left(1-\frac{2}{n}\right)g}$$

where α_n satisfies

$$\log\log\log\frac{1}{\alpha_n} \sim n\log(1+\sqrt{2})$$

as $n \to \infty$.

The asymptotic notation $f(x) \sim g(x)$ used in the above statement means that

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1.$$

If the functions involved depend on several variables, we will indicate which one is sent to infinity by writing $f \sim_x g$.

We start by giving a lower bound for the number $S(n, E, w)$ of isomorphism classes of unlabelled, connected, n-regular, signed graphs with E edges and girth at least w. To simplify matters, we consider the labelled version of such graphs first.

Recall that a signed graph is a graph together with a cyclic ordering of the edges attached to every vertex and a choice of sign between consecutive edges such that the product of the signs around any vertex is negative. Also recall that the cyclic order around any vertex can be reversed (and two signs around each neighbor changed appropriately) without changing the isomorphism class of a signed graph. Thus a signed graph Γ with labelled vertices has a total of 2^V isomorphic representations with the same vertex labels, where V is the number of vertices in Γ.

Assume that a cyclic order has been chosen for the edges around each vertex (there are $(n-1)!$ cyclic orders on n elements). Then there are n signs to pick around each vertex, but since their product is required to be negative, any sign can be deduced from the remaining ones. Hence the number of admissible sign patterns around a vertex is 2^{n-1}. The total number of sign patterns on the whole graph is therefore $2^{(n-1)V}$.

Now to count the number of isomorphism classes of unlabelled signed graphs, we have to take into account the fact that some graphs admit non-trivial automorphisms, which could result in overcounting. To remedy this, we restrict ourselves to underlying graphs that are asymmetric, that is, have trivial automorphism group. Let $A(n, E, w)$ be the number of unlabelled, connected, asymmetric, n-regular graphs with E edges, and of girth at least w. Then the above reasoning shows that

$$S(n, E, w) \geq 2^{(n-2)V}((n-1)!)^V A(n, E, w). \tag{6.2}$$

Note that all the signed graphs with an asymmetric underlying graph are themselves asymmetric.

To estimate $A(n, E, w)$ we combine a few results from graph theory. Let $U(E, n, w)$ be the number of unlabelled n-regular graphs with E edges and girth at least w. In the literature, it is often assumed that the graphs are simple, namely, that $w \geq 3$ (no monogons or bigons). To emphasize this and to maintain a consistent notation, we use $U(E, n, 3)$ for the number of unlabelled n-regular simple graphs with E edges.

In [7] Bollobás showed the following:

Theorem 6.6 (Bollobás)**.** *For every $n \geq 3$, we have*

$$U(n,E,3) \sim_E \exp\left(-\sum_{i=1}^{2}\frac{(n-1)^i}{2i}\right)\cdot\frac{(2E)!}{2^E\,E!\,V!\,(n!)^V}$$

as $E \to \infty$ in such a way that $2E$ is divisible by n and where $V = 2E/n$ is the number of vertices in the graphs.

For the rest of the paper, the symbol $f \sim_E g$ means the ratio of f and g goes to one as $E \to \infty$ in such a way that $2E$ is divisible by n. Wormald [56] strengthened the above result to show:

Theorem 6.7 (Wormald)**.** *For every $n \geq 3$ and every $w \geq 3$ we have*

$$U(n,E,w) \sim_E \exp\left(-\sum_{i=3}^{w-1}\frac{(n-1)^i}{2i}\right)U(n,E,3).$$

Together, these two results imply that

$$U(n,E,w) \sim_E \exp\left(-\sum_{i=1}^{w-1}\frac{(n-1)^i}{2i}\right)\cdot\frac{(2E)!}{2^E\,E!\,V!\,(n!)^V}. \tag{6.3}$$

Bollobás also showed that regular simple graphs are generically connected [9, p.195] and asymmetric [8, Theorem 6].

Theorem 6.8 (Bollobás)**.** *For every $n \geq 3$ we have*

$$A(n,E,3) \sim_E U(n,E,3).$$

As a consequence, we have that a generic n-regular graph of girth at least $w \geq 3$ is also connected and asymmetric. Indeed,

$$\begin{aligned}\frac{U(n,E,w)-A(n,E,w)}{U(n,E,w)} &\sim_E \exp\left(\sum_{i=3}^{w-1}\frac{(n-1)^i}{2i}\right)\frac{U(n,E,w)-A(n,E,w)}{U(n,E,3)}\\ &\leq \exp\left(\sum_{i=3}^{w-1}\frac{(n-1)^i}{2i}\right)\frac{U(n,E,3)-A(n,E,3)}{U(n,E,3)} \sim_E 0\end{aligned}$$

where the first $\sim$ is Theorem 6.7, the inequality holds because graphs of girth at least w form a subset of the set of graphs of girth at least 3, and the last $\sim$ is Theorem 6.8. This shows that

$$A(n,E,w) \sim_E U(n,E,w)$$

for every $n \geq 3$ and $w \geq 3$.

Combining this with Equation (6.2) and Equation (6.3), we get (after simplification)

$$S(n,E,w) \gtrsim_E \exp\left(-\sum_{i=1}^{w-1}\frac{(n-1)^i}{2i}\right)\cdot\frac{(2E)!\,2^E}{E!\,V!\,(4n)^V}. \tag{6.4}$$

Proof of Theorem 6.5. By Theorem 4.2 and Corollary 6.4, for every finite connected n-regular signed graph Γ of girth larger than $Kn(1+\sqrt{2})^n$, the surface $X(\Gamma)$ is a local maximum of the systole function at height $L_n = a(t_n)$ in $\mathcal{M}_g$ where $g = E+1$. Also, by Theorem 5.3, non-isomorphic signed graphs correspond to distinct points in moduli space, and if the signed graph Γ is asymmetric then $X(\Gamma)$ has a trivial group of orientation-preserving isometries. In other words, the number of asymmetric local maxima of the systole function at height L_n in $\mathcal{M}_g$ is at least $S(n, g-1, \lfloor Kn(1+\sqrt{2})^n \rfloor + 1)$.

Thus, we only need to simplify Equation (6.4) and write it in terms of n and g. Set

$$\alpha_n = \exp\left(-\sum_{i=1}^{w-1} \frac{(n-1)^i}{2i}\right)$$

where $w = \lfloor Kn(1+\sqrt{2})^n \rfloor + 1$. Note that α_n depends only on n and not on g. We have

$$\frac{(n-1)^{w-1}}{2w} \leq \log \frac{1}{\alpha_n} \leq w n^w.$$

Taking the logarithm two more times, we get

$$\log\log\log \frac{1}{\alpha_n} \sim_n \log w \sim_n n \log(1+\sqrt{2}). \tag{6.5}$$

We use the Stirling's formula to simplify the remaining terms. The latter implies that there are positive constants $\beta_1, \dots, \beta_4$ such that

$$(2E)! \gtrsim_g (\beta_1\, g)^{2g}, \quad E! \lesssim_g (\beta_2\, g)^g, \quad V! \lesssim_g (\beta_3\, g/n)^{2g/n} \quad \text{and} \quad (4n)^V \lesssim_g (\beta_4\, n)^{2g/n}.$$

Hence, after collecting the constants, we can estimate the remaining terms in Equation (6.4) as

$$\frac{(2E)!\, 2^E}{E!\, V!\, (4n)^V} \gtrsim_g \beta^g \frac{g^{2g}}{g^g\, (g/n)^{2g/n}\, n^{2g/n}} \gtrsim_g (\beta\, g)^{\left(1-\frac{2}{n}\right)g},$$

for some constant $\beta > 0$ independent of n and g. This finishes the proof. □

References

[1] C. Adams. Maximal cusps, collars, and systoles in hyperbolic surfaces. *Indiana Univ. Math. J.*, 47(2):419–437, 1998.

[2] H. Akrout. Singularités topologiques des systoles généralisées. *Topology*, 42(2):291–308, 2003.

[3] J. W. Anderson, H. Parlier, and A. Pettet. Relative shapes of thick subsets of moduli space. *Amer. J. Math.*, 138(2):473–498, 2016.

[4] C. Bavard. Systole et invariant d'Hermite. *J. Reine Angew. Math.*, 482:93–120, 1997.

[5] C. Bavard. Théorie de Voronoï géométrique. Propriétés de finitude pour les familles de réseaux et analogues. *Bull. Soc. Math. France*, 133(2):205–257, 2005.

[6] M. Bestvina. Four questions about mapping class groups. In *Problems on mapping class groups and related topics*, volume 74 of *Proc. Sympos. Pure Math.*, pages 3–9. Amer. Math. Soc., Providence, RI, 2006.

[7] B. Bollobás. The asymptotic number of unlabelled regular graphs. *J. London Math. Soc. (2)*, 26(2):201–206, 1982.

[8] B. Bollobás. Distinguishing vertices of random graphs. In *Graph theory (Cambridge, 1981)*, volume 62 of *North-Holland Math. Stud.*, pages 33–49. North-Holland, Amsterdam-New York, 1982.

[9] B. Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.

[10] P. Buser. Cubic graphs and the first eigenvalue of a Riemann surface. *Math. Z.*, 162(1):87–99, 1978.

[11] P. Buser. Riemannsche Flächen mit grosser Kragenweite. *Comment. Math. Helv.*, 53(3):395–407, 1978.

[12] P. Buser. *Geometry and spectra of compact Riemann surfaces*. Modern Birkhäuser Classics. Birkhäuser Boston, Inc., Boston, MA, 2010. Reprint of the 1992 edition.

[13] P. Buser and P. Sarnak. On the period matrix of a Riemann surface of large genus. *Invent. Math.*, 117(1):27–56, 1994. With an appendix by J.H. Conway and N.J.A. Sloane.

[14] A. Casamayou-Boucau. Suites infinies de surfaces de Riemann parfaites. *Bull. Sci. Math.*, 128(9):739–748, 2004.

[15] A. Casamayou-Boucau. Surfaces de Riemann parfaites en genre 4 et 6. *Comment. Math. Helv.*, 80(3):455–482, 2005.

[16] W. Cavendish and H. Parlier. Growth of the Weil–Petersson diameter of moduli space. *Duke Math. J.*, 161(1):139–171, 2012.

[17] B. Colbois. Petites valeurs propres du laplacien sur une surface de Riemann compacte et graphes. *C. R. Acad. Sci. Paris Sér. I Math.*, 301(20):927–930, 1985.

[18] B. Colbois and Y. Colin de Verdière. Sur la multiplicité de la première valeur propre d'une surface de Riemann à courbure constante. *Comment. Math. Helv.*, 63(2):194–208, 1988.

[19] C.R.F. Dória. *How systoles grow*. PhD thesis, Instituto Nacional de Matemática Pura e Aplicada, 2018.

[20] F. Fanoni and H. Parlier. Systoles and kissing numbers of finite area hyperbolic surfaces. *Algebr. Geom. Topol.*, 15(6):3409–3433, 2015.

[21] B. Farb, C. J. Leininger, and D. Margalit. Small dilatation pseudo-Anosov homeomorphisms and 3-manifolds. *Adv. Math.*, 228(3):1466–1502, 2011.

[22] B. Farb and D. Margalit. *A primer on mapping class groups*, volume 49 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 2012.

[23] B. Farb and H. Masur. Teichmüller geometry of moduli space. II. $\mathcal{M}(S)$ seen from far away. In *In the tradition of Ahlfors-Bers. V*, volume 510 of *Contemp. Math.*, pages 71–79. Amer. Math. Soc., Providence, RI, 2010.

[24] A. Fletcher, J. Kahn, and V. Markovic. The moduli space of Riemann surfaces of large genus. *Geom. Funct. Anal.*, 23(3):867–887, 2013.

[25] M. Fortier Bourque. Hyperbolic surfaces with sublinearly many systoles that fill. To appear in *Commentarii Mathematici Helvetici*.

[26] M. Gendulphe. Paysage systolique des surfaces hyperboliques de caractéristique -1. Preprint, matthieu.gendulphe.com/Gendulphe-PaysageSystolique.pdf.

[27] M. Gendulphe. Découpages et inégalités systoliques pour les surfaces hyperboliques à bord. *Geom. Dedicata*, 142:23–35, 2009.

[28] S. A. Gershgorin. Über die Abgrenzung der Eigenwerte einer Matrix. *Bull. Acad. Sci. URSS*, 1931(6):749–754, 1931.

[29] U. Hamenstädt. New examples of maximal surfaces. *Enseign. Math. (2)*, 47(1-2):65–101, 2001.

[30] U. Hamenstädt. Length functions and parameterizations of Teichmüller space for surfaces with cusps. *Ann. Acad. Sci. Fenn. Math.*, 28(1):75–88, 2003.

[31] U. Hamenstädt. Parametrizations of Teichmüller space and its Thurston boundary. In *Geometric analysis and nonlinear partial differential equations*, pages 81–88. Springer, Berlin, 2003.

[32] U. Hamenstädt and R. Koch. Systoles of a family of triangle surfaces. *Experiment. Math.*, 11(2):249–270, 2002.

[33] J. Harer and D. Zagier. The Euler characteristic of the moduli space of curves. *Invent. Math.*, 85(3):457–485, 1986.

[34] J. L. Harer. The virtual cohomological dimension of the mapping class group of an orientable surface. *Invent. Math.*, 84(1):157–176, 1986.

[35] J. L. Harer. Stability of the homology of the moduli spaces of Riemann surfaces with spin structure. *Math. Ann.*, 287(2):323–334, 1990.

[36] F. Jenni. Über den ersten Eigenwert des Laplace-Operators auf ausgewählten Beispielen kompakter Riemannscher Flächen. *Comment. Math. Helv.*, 59(2):193–203, 1984.

[37] H. Karcher and M. Weber. The geometry of Klein's Riemann surface. In *The eightfold way*, volume 35 of *Math. Sci. Res. Inst. Publ.*, pages 9–49. Cambridge Univ. Press, Cambridge, 1999.

[38] S. P. Kerckhoff. The Nielsen realization problem. *Ann. of Math. (2)*, 117(2):235–265, 1983.

[39] C .J. Leininger and D. Margalit. On the number and location of short geodesics in moduli space. *J. Topol.*, 6(1):30–48, 2013.

[40] I. Madsen and M. Weiss. The stable moduli space of Riemann surfaces: Mumford's conjecture. *Ann. of Math. (2)*, 165(3):843–941, 2007.

[41] M. Mirzakhani and B. Petri. Lengths of closed geodesics on random surfaces of large genus. To appear in *Commentarii Mathematici Helvetici*.

[42] D. Mumford. A remark on Mahler's compactness theorem. *Proc. Amer. Math. Soc.*, 28:289–294, 1971.

[43] H. Parlier. Simple closed geodesics and the study of Teichmüller spaces. In *Handbook of Teichmüller theory. Vol. IV*, volume 19 of *IRMA Lect. Math. Theor. Phys.*, pages 113–134. Eur. Math. Soc., Zürich, 2014.

[44] B. Petri. Hyperbolic surfaces with long systoles that form a pants decomposition. *Proc. Amer. Math. Soc.*, 146(3):1069–1081, 2018.

[45] B. Petri and A. Walker. Graphs of large girth and surfaces of large systole. *Math. Res. Lett.*, 25(6):1937–1956, 2018.

[46] K. Rafi and J. Tao. The diameter of the thick part of moduli space and simultaneous Whitehead moves. *Duke Math. J.*, 162(10):1833–1876, 2013.

[47] G. Riera. A formula for the Weil–Petersson product of quadratic differentials. *J. Anal. Math.*, 95:105–120, 2005.

[48] P. Schmutz. Die Parametrisierung des Teichmüllerraumes durch geodätische Längenfunktionen. *Comment. Math. Helv.*, 68(2):278–288, 1993.

[49] P. Schmutz. Riemann surfaces with shortest geodesic of maximal length. *Geom. Funct. Anal.*, 3(6):564–631, 1993.

[50] P. Schmutz. Congruence subgroups and maximal Riemann surfaces. *J. Geom. Anal.*, 4(2):207–218, 1994.

[51] P. Schmutz. Systoles on Riemann surfaces. *Manuscripta Math.*, 85(3-4):429–447, 1994.

[52] P. Schmutz Schaller. Geometry of Riemann surfaces based on closed geodesics. *Bull. Amer. Math. Soc. (N.S.)*, 35(3):193–214, 1998.

[53] P. Schmutz Schaller. Systoles and topological Morse functions for Riemann surfaces. *J. Differential Geom.*, 52(3):407–452, 1999.

[54] S. Wolpert. An elementary formula for the Fenchel–Nielsen twist. *Comment. Math. Helv.*, 56(1):132–135, 1981.

[55] S. Wolpert. The Fenchel–Nielsen deformation. *Ann. of Math. (2)*, 115(3):501–528, 1982.

[56] N. C. Wormald. The asymptotic distribution of short cycles in random regular graphs. *J. Combin. Theory Ser. B*, 31(2):168–182, 1981.