

L-CUTS FOR GENUS TWO TRANSLATION SURFACES

by

Andrew Kevin Bouwman

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Mathematics

MONTANA STATE UNIVERSITY
Bozeman, Montana

July 2013

©COPYRIGHT

by

Andrew Kevin Bouwman

2013

All Rights Reserved

APPROVAL

of a dissertation submitted by

Andrew Kevin Bouwman

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to The Graduate School.

Dr. Jaroslaw Kwapisz

Approved for the Department of Mathematics

Dr. Kenneth L. Bowers

Approved for The Graduate School

Dr. Ronald W. Larsen

STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with “fair use” as prescribed in the U. S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to Bell & Howell Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted “the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part.”

Andrew Kevin Bouwman

July 2013

TABLE OF CONTENTS

1. INTRODUCTION.....	1
Main Results.....	1
Motivation and Application	4
Overview	8
2. PRELIMINARIES	10
Translation Surfaces	10
Holomorphic One-Forms	15
Hyperelliptic Involution	16
Connected Sums.....	18
3. POLYBAND CONSTRUCTION AND MCMULLEN'S THEOREM.....	21
Polyband Construction	21
Hyperellipticity	28
McMullen's Theorem.....	32
4. CONNECTED SUMS USING L-CUTS	39
Definitions	39
Extending McMullen's Theorem.....	43
Splitting along L-Cuts	45
Preliminaries for the Proof of Theorem 4.7.....	49
Interval Exchange Map	50
Zippered Rectangles and Polygons	57
Rauzy Operations and Rauzy Diagrams	63
Hyperelliptic Involution on Polygons.....	71
Non-Minimal Vertical Flow	75
Ribbon Graphs.....	77
Classifying the M_i	86
Proof of Theorem 4.17	88
Proof of Theorem 4.7.....	94
Rectangles.....	94
$M \in \mathcal{H}^\circ(2)$	95
$M \in \mathcal{H}^\circ(1, 1)$	98
$M \in \mathcal{H}_{\text{VM}}^*(2)$	103
$M \in \mathcal{H}_{\text{VM}}^*(1, 1)$	105

TABLE OF CONTENTS - CONTINUED

5. SMALE'S QUESTION AND THE FRANKS MAP	111
Smale's Question	111
Hyperbolic Toral Automorphisms	111
Pseudo-Anosov Maps	112
Application to Smale's Question	114
Shadowing	115
The Homology Cover $\hat{M}$	116
Global Shadowing	116
Non-Injectivity of Pseudo-Anosov Embeddings	121
Passing to $\check{M}$	121
Pseudo-Anosov Maps with Orientable Foliations in Genus Two	124
Characterization of Embeddings	126
Franks Map	127
REFERENCES CITED	134

LIST OF FIGURES

Figure	Page
1.1. A splitting performed using a homologous pair of straight segments.	3
1.2. A splitting performed using a parallel pair of L-cuts.....	4
2.1. Polygon with antenna and adjacent parallel edges.	11
2.2. Identifying opposite edges of a parallelogram to get a torus	12
2.3. Elements of $\mathcal{H}(2)$	14
2.4. Elements of $\mathcal{H}(1, 1)$	14
2.5. Foliations of $w^k dw$ for $k = 0, 1, 2$	16
2.6. Hyperelliptic involution on a genus two surface.	18
2.7. Connected sum of tori.	19
3.1. $M \in \mathcal{H}(2)$ arises from gluing sides of a polygon P or a polyband $\mathcal{A}$	22
3.2. Smillie's construction of a (regular) closed geodesic.	24
3.3. $M \in \mathcal{H}(2)$ results in a $[3, 2, 1]$ polyband.	26
3.4. $M \in \mathcal{H}(1, 1)$ results in a $[4, 3, 2, 1]$ polyband.	27
3.5. Level lines of a 1-form	29
3.6. Stamped torus	32
3.7. The construction of J for $M \in \mathcal{H}(2)$	33
3.8. <i>up, up, up, down</i> and <i>up, up, down, down</i> (non-degenerate cases).....	35
3.9. <i>up, down, up, down</i> (non-degenerate case), bad J	36
3.10. <i>up, down, up, down</i> (non-degenerate case), good J	36

LIST OF FIGURES - CONTINUED

Figure	Page
3.11. J and $\eta(J)$ cut M into two tori	37
3.12. Degenerate cases with 2, 3, and all 4 vertices of L^+ in ∂C	38
4.1. L-cuts on $M \in \mathcal{H}(2)$ and $M \in \mathcal{H}(1, 1)$	41
4.2. A singular L-cut	42
4.3. A parallel pair of non-singular L-cuts	43
4.4. Parallel L-cuts with a third, non-parallel one.	44
4.5. Two L-cuts bounding a rectangle R in M	46
4.6. A split via a parallel pair of (non-singular) L-cuts.	47
4.7. Split via a parallel pair of singular L-cuts	48
4.8. Vertical flow from I for $M \in \mathcal{H}^\circ(2)$	51
4.9. Vertical flow from I for $M \in \mathcal{H}_{\text{VM}}^*(2)$	53
4.10. A zippered rectangle for $M \in \mathcal{H}^\circ(2)$	58
4.11. A zippered rectangle for $M \in \mathcal{H}_{\text{VM}}^*(2)$	59
4.12. An octagon representing $M \in \mathcal{H}^\circ(2)$ formed from a zippered rectangle. ...	61
4.13. Four octagons representing a single $M \in \mathcal{H}_{\text{VM}}^*(2)$	62
4.14. A zippered rectangle producing a polygon with a fishtail.	63
4.15. Performing a Rauzy operation “0” on $M \in \mathcal{H}^\circ(2)$	65
4.16. Performing a Rauzy operation “0” on $M \in \mathcal{H}_{\text{VM}}^*(2)$	66
4.17. Rauzy diagram for the stratum $\mathcal{H}(2)$	68

LIST OF FIGURES - CONTINUED

Figure	Page
4.18. Rauzy diagram for the stratum $\mathcal{H}(1, 1)$	69
4.19. Behavior of the hyperelliptic involution in $\mathcal{H}(2)$	73
4.20. Behavior of the hyperelliptic involution in $\mathcal{H}(1, 1)$	74
4.21. Orientation of ribbon graphs	81
4.22. Hyperelliptic measured ribbon graphs for $M \in \mathcal{H}^*(1, 1)$	82
4.23. Hyperelliptic measured ribbon graphs for $M \in \mathcal{H}^*(1, 1)$	83
4.24. Graphs from MRG(2) not realizable by a surface in $\mathcal{H}^*(2)$	83
4.25. Graphs from MRG(1, 1) not realizable by a surface in $\mathcal{H}^*(1, 1)$	83
4.26. Detail of the last measured ribbon graph from Figure 4.25.	85
4.27. Geometric and conceptual depictions of the M cases.	90
4.28. Geometric and conceptual depictions of the P cases.	92
4.29. Geometric and conceptual depictions of the N cases.	93
4.30. Depiction of invasion of a candidate rectangle by a vertex	96
4.31. Candidate rectangles R_1 and R_2 for $\mathcal{H}^\circ(2)$	96
4.32. Representative diagrams for $M \in \mathcal{H}^\circ(2)$	97
4.33. Candidate rectangles R_1 through R_5 for $\mathcal{H}^\circ(1, 1)$	99
4.34. Representative diagrams for $\mathcal{H}^\circ(1, 1)$, cases 1 and 2	101
4.35. Representative diagrams for $\mathcal{H}^\circ(1, 1)$, case 3.	102
4.36. Representative diagrams for $\mathcal{H}_{\text{VM}}^*(2)$	104

LIST OF FIGURES - CONTINUED

Figure	Page
4.37. Representative diagrams for $\mathcal{H}_{\text{VM}}^*(1, 1)$, special cases.	106
4.38. Representative diagrams for $\mathcal{H}_{\text{VM}}^*(1, 1)$, second batch.	108
4.39. Representative diagrams for $\mathcal{H}_{\text{VM}}^*(1, 1)$, third batch.	109
4.40. Representative diagrams for $\mathcal{H}_{\text{VM}}^*(1, 1)$, case D	110
5.1. Homologically co-heteroclinic points lift to heteroclinic points.	119
5.2. Heteroclinic points shadow each other in $\hat{M}$	119
5.3. Relationship of M , $\hat{M}$, and $\check{M}$	123

ABSTRACT

A connected sum is a topological way of joining two Riemann surfaces which results in another surface. It is used in the classification of all connected closed orientable surfaces as being homeomorphic to either the sphere, or a connected sum of tori. The reverse operation, here referred to as a splitting, decomposes a surface as a connected sum.

It was recently shown by Curtis McMullen that any translation surface of genus two can be written in infinitely many ways as a connected sum of two flat tori. His method was to find a certain straight saddle connection J , and perform a splitting along $J \cup \eta(J)$, where η is the hyperelliptic involution (the unique degree-two automorphism on the surface which fixes exactly six points). In this dissertation, we give an elementary argument for existence of such J and show that for all surfaces of genus two on which the vertical flow is minimal, the same kind of splitting is possible along a parallel pair of paths with the straight saddle connection replaced by an L-cut: a broken line with one horizontal and one vertical segment.

As a direct consequence of this L-cut splitting, it is shown that a homeomorphism on a genus-two surface which is conjugated to a hyperbolic toral automorphism restricted to an invariant subset (if any such situation even exists) can only be pseudo-Anosov with non-orientable foliations. This makes progress toward addressing an old question of Stephen Smale about the existence of an invariant set of a hyperbolic toral automorphism which is itself a compact surface.

CHAPTER 1

INTRODUCTION

Main Results

In [1], Curtis McMullen proved the following theorem:

Theorem 1.1 (McMullen). *(cf. **Theorem 3.7**). Any translation surface M of genus $g = 2$ contains a saddle connection J such that $\eta(J) \neq J$ (where η is the hyperelliptic involution, the unique involution on M fixing $2g + 2 = 6$ points). The surface M splits along $J \cup \eta(J)$ into a connected sum of two tori. Moreover, given a (maximal) open cylinder C such J can be found so that neither J nor $\eta(J)$ cross ∂C (although they may be contained in ∂C or have endpoints in ∂C).*

McMullen was able to establish infinitely many such splittings, allowing him to describe the behavior of the $\mathrm{SL}_2(\mathbb{R})$ action on the bundle of holomorphic one-forms over the moduli space of Riemann surfaces of genus two, in particular classifying all possible orbit closures.

One of the contributions of this thesis is an elementary proof of McMullen's theorem, seen in Chapter 3. (We learned that a similar proof of existence of J can be found in the survey [2], but there the result is weaker as it establishes only one such splitting.) In the process of proving McMullen's theorem, we develop a novel alternative approach to understanding all surfaces of genus two by something we call

polyband construction. This is a variation of Veech's zippered rectangle construction (see [3]) that yields a particularly simple combinatorial presentation of such surfaces. In this presentation, many features of genus two surfaces, such as the hyperelliptic involution, become quite apparent. The proof of McMullen's theorem is also made rather simple.

McMullen's splitting has the undesirable property that the direction of the straight saddle connection forming the split cannot be prescribed. The primary result of this dissertation is that such a prescription is possible if one relaxes the requirement that the saddle connection be straight and allows it to have a single turning point. Specifically, an L-cut is a saddle connection which begins in a prescribed direction (horizontal, upon orienting the surface appropriately), and turns once, proceeding in a direction perpendicular to the first (vertical). First, the existence of two such paths is established on surfaces with vertical minimal flow:

Theorem 1.2 (Existence of L-cuts). (*cf.* **Theorem 4.7**). *Suppose M is a genus two translation surface with vertical flow which is minimal. Then there exists a parallel pair of L-cuts (K, K') in M . Moreover, if M has no vertical and no horizontal saddle connections then the L-cuts K and K' are non-singular.*

In this context, *parallel* has a particular meaning (which will be defined precisely in Chapter 4) analogous to the relationship between J and $\eta(J)$. (Chiefly, they have the same endpoints, their horizontal and vertical segments have the same lengths, and they are homologous.) Later, it is established that when a genus two translation

surface has a parallel pair of L-cuts, it is possible to express it as a connected sum of two flat tori using this pair:

Theorem 1.3 (L-cut Splitting). *(cf. Theorem 4.8). Given a genus 2 translation surface M with a parallel pair of L-cuts (K, K') as in Theorem 4.7, M splits along $K \cup K'$ as a connected sum of two flat tori.*

Compare the splitting using straight segments depicted in Figure 1.1 with the splitting using L-cuts in Figure 1.2.

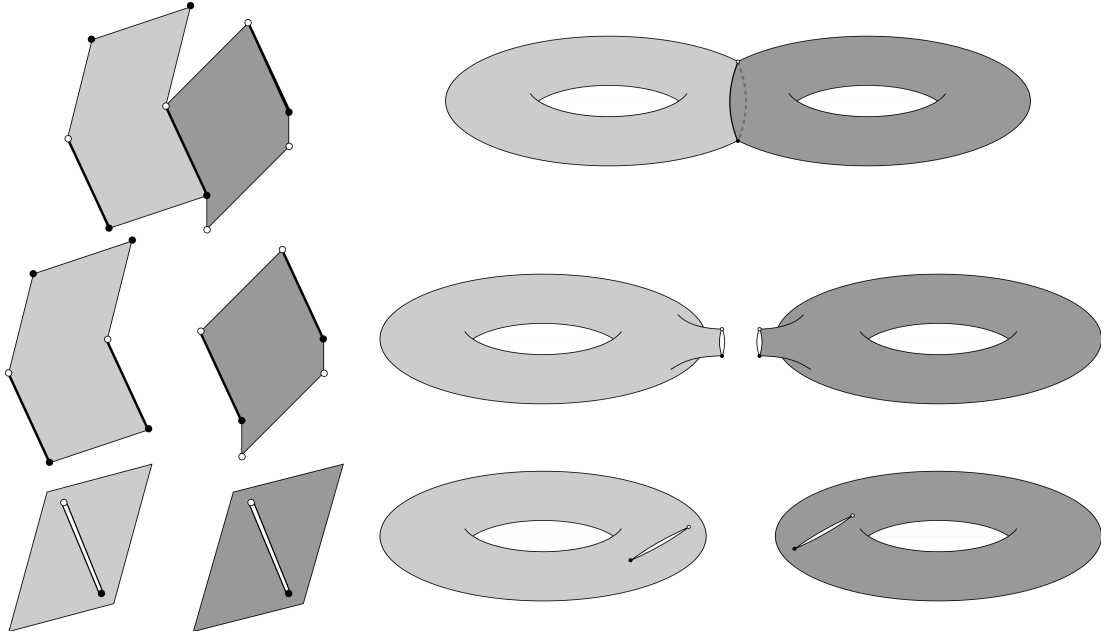


Figure 1.1: A splitting performed using a homologous pair of straight segments, as in McMullen's result.

For M on which vertical flow is not minimal, the following decomposition is established:

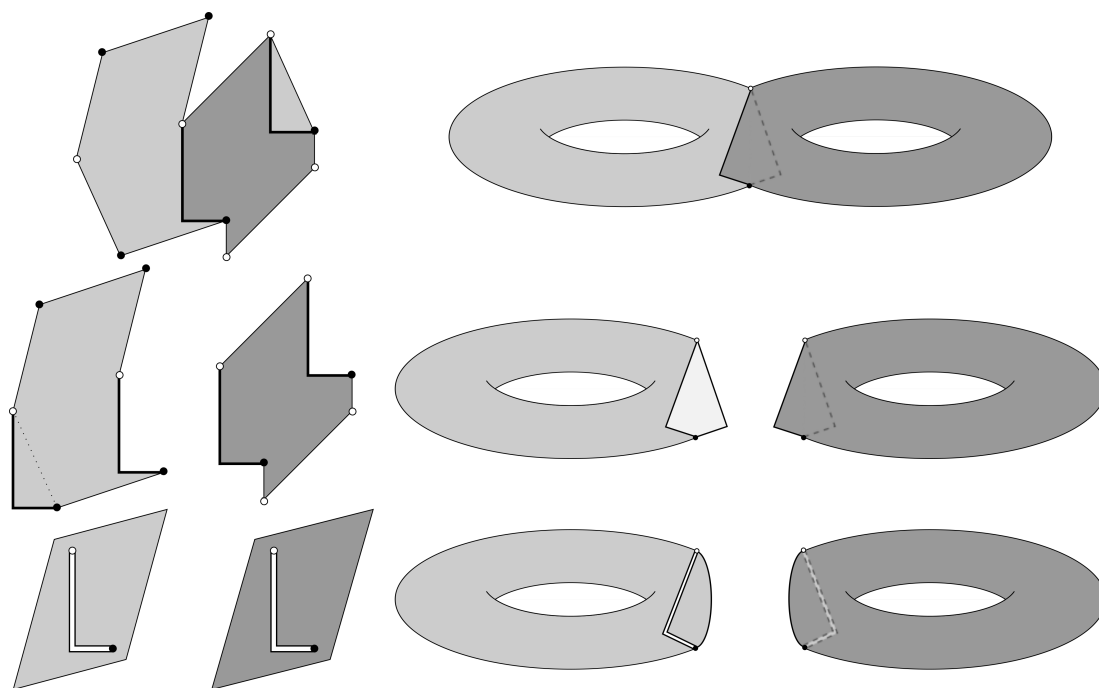


Figure 1.2: A splitting performed using a parallel pair of L-cuts.

Theorem 1.4 (Decomposition of Surfaces with Non-Minimal Flow). (*cf.* **Theorem 4.17**). *Suppose M is a genus two translation surface on which the vertical flow is not minimal. Let $\mathcal{S}$ be the union of all vertical saddle connections and singular points of M . Then M can be written as $\bigcup_{i=1}^m M_i$, with $m \leq 3$, where each M_i satisfies either that $M_i \setminus \mathcal{S}$ is an open periodic cylinder, or M_i is a surface homeomorphic to a torus with one or two open discs removed and the vertical flow on M_i is minimal.*

Motivation and Application

One interesting consequence of the L-cut splitting is a partial resolution of a question posed by Stephen Smale regarding the nature of compact invariant sets of

a hyperbolic toral automorphism [4]. There are uncountably many such invariant sets of fractal nature [5, 6]. However, in the face of even minimal smoothness conditions, the only invariant submanifolds are necessarily made up of a finite union of subtori [7]. The question of whether there are any other types of invariant topological submanifolds remains open, though some progress has been made.

Albert Fathi demonstrated in [8] using a construction of John Franks in [9] that there is a subset of the torus which is invariant under a hyperbolic toral automorphism and is an image of a closed orientable surface of genus $g > 1$ under a continuous map which is locally injective on the complement of a finite set. Marcy Barge and Jaroslaw Kwapisz demonstrated in [10] that this map is in many cases almost a conjugacy, i.e., it is almost everywhere injective. As a consequence of Fathi's result, Barge and Kwapisz show that hyperbolic pseudo-Anosov homeomorphisms with orientable foliations can be almost conjugated to a restriction of a hyperbolic toral automorphism to an invariant subset.

Mikhail Gromov discusses the nature of Franks' map (including its injectivity) in his survey [11] where he places it in a broader context centered on the idea of global shadowing. Franks' map can be viewed (see, e.g., [10]) as a dynamical (or, more precisely, dynamically equivariant) version of the classical Abel-Jacobi map from a Riemann surface M into its Jacobi variety (which is the torus $H_1(M, \mathbb{R})/H_1(M, \mathbb{Z})$ equipped with a complex structure). Gromov calls it the Franks-Abel map and makes a point of contrasting the injectivity of the Abel-Jacobi map with the (conjectured)

non-injectivity of the Franks-Abel variant.

Gavin Band showed in [12], for a certain class of pseudo-Anosov maps whose foliations have one singularity, that the Franks map cannot be even locally injective at the singular points of the stable and unstable foliations. In this paper, we extend Band's result to all surfaces M of genus two and thus show that a hyperbolic pseudo-Anosov map with orientable foliations on such a surface cannot be embedded into a hyperbolic toral automorphism as an invariant set:

Theorem 1.5 (Non-Embedding). *(cf. Theorem 5.12). Let M be a surface of genus two. If $\psi : M \rightarrow \mathbb{T}^N$ is a continuous map satisfying $\psi \circ f = f_A \circ \psi$, where $f : M \rightarrow M$ is a pseudo-Anosov map with orientable foliations and $f_A : \mathbb{T}^N \rightarrow \mathbb{T}^N$ is a hyperbolic toral automorphism, then ψ is not injective.*

It follows that a pseudo-Anosov map embeddable into a hyperbolic toral automorphism must have non-orientable foliations, but it is suspected that no global embedding exists for any pseudo-Anosov map. We also show failure of local injectivity, by a mechanism similar to Band's.

Furthermore, it is shown that any embedding of a pseudo-Anosov map on a surface M of any genus into a hyperbolic toral automorphism must either be an instance of a Franks map, or take M into a proper subtorus invariant under the hyperbolic toral automorphism:

Theorem 1.6 (Essential Uniqueness of Embedding). *(cf. Theorem 5.20.) If $\psi : M \rightarrow \mathbb{T}^N$ is some embedding such that $\psi \circ f = f_A \circ \psi$, where f is a pseudo-Anosov*

map with orientable foliations and $f_A : \mathbb{T}^N \rightarrow \mathbb{T}^N$ is a hyperbolic toral automorphism, and $\psi_* : H_1(M, \mathbb{R}) \rightarrow H_1(\mathbb{T}^N, \mathbb{R}) \cong \mathbb{R}^N$ is surjective, then ψ coincides with a Franks map $h : M \rightarrow \mathbb{T}^N$ constructed from a suitable set of cohomology classes. If ψ_* is not surjective, then ψ takes values in a proper f_A -invariant subtorus.

Any results about embeddings of pseudo-Anosov maps can be generalized to arbitrary embeddings of a compact surface into a hyperbolic toral automorphism. This is because the restriction of the automorphism to such an embedded surface is a pseudo-Anosov map by a deep theorem given independently by Koichi Hiraide [13] and Jorge Lewowicz [14]:

Theorem 1.7 (Hiraide, Lewowicz). *(cf. Theorem 5.15.) Every expansive homeomorphism of a compact surface is conjugate to a pseudo-Anosov map.*

In particular, since hyperbolic toral automorphisms are known to be expansive, our Non-Embedding Theorem implies for genus two that such a pseudo-Anosov map cannot have orientable foliations, yielding the following result:

Corollary 1.8. *(cf. Corollary 5.16.) Any embedding ψ of a genus two surface onto an invariant set of a hyperbolic toral automorphism f_A must be such that $g := \psi^{-1} \circ f_A \circ \psi$ is a pseudo-Anosov map with non-orientable foliations.*

We know that the combinatorial complexity of pseudo-Anosov maps with non-orientable foliations in genus two is by one magnitude higher than that of those with

orientable foliations (see the tables at the end of [15]). The situation for higher genus is even more daunting.

Overview

Before diving into the details of the results and their proofs, Chapter 2 containing preliminary material is included. Here formal definitions of the basic elements of translation surface and holomorphic one-form are presented, as well as the hyperelliptic involution, and the connected sum, which are more specific to the task at hand.

Chapter 3 describes the polyband construction, an alternative to the common polygonal presentation of translation surfaces, and discrete datum, a way of classifying the associated combinatorial data. Using these tools, the existence of the hyperelliptic involution in genus two is established, followed by an alternative proof of McMullen's connected sum theorem for genus two.

The bulk of the dissertation is the introduction of L-cuts in Chapter 4, and the proof of Theorem 4.7 establishing for the above described class of genus two surfaces the existence of a pair of L-cuts along which the surface splits into two flat tori. The method of proof is very geometric in nature, relying on establishing the existence of a certain special rectangle on a polygonal presentation of the surface. It is also necessary to consider degenerate cases, where singular points of the translation surface lie in the path of the L-cuts.

Genus two has a few key advantages over higher genera. First, as the configuration of singular points is simpler, genus two has far simpler combinatorial complexity than higher genera [15, 16]. A polygon representing a genus two translation surface can be either an octagon or decagon; the number of sides increases rapidly not only with genus, but also the configuration of the singular points. Additionally, the key tool of hyperelliptic involution is not universally present in higher genus, although the moduli space of each higher genus has a connected component containing exactly the hyperelliptic surfaces; see, for example, [16].

The most significant results appear in Chapter 5. Here, Theorem 4.7 is used to address Smale's question by showing that there does not exist a global embedding of a pseudo-Anosov map with orientable foliations on a genus two surface into a hyperbolic toral automorphism. Thus, Corollary 5.16 shows that any embeddable pseudo-Anosov map must have non-orientable foliations (if indeed any such map even exists). Finally, it is shown that Franks' construction is the only way to create an embedding of a pseudo-Anosov map with orientable foliations into a hyperbolic toral automorphism.

CHAPTER 2

PRELIMINARIES

Translation Surfaces

Definition 2.1. A polygonal chain $\mathcal{C}$ is a finite collection $\{e_1, e_2, \dots, e_n\}$ of oriented line segments in $\mathbb{R}^2$ such that the end of e_i coincides with the beginning of e_{i+1} for $i = 1, \dots, n-1$. These line segments are called the edges of $\mathcal{C}$ and the points $v_0, v_1, v_2, \dots, v_n$, where v_{i-1}, v_i are the beginning and the end, respectively, of e_i ($i = 1, \dots, n$), are called the vertices of $\mathcal{C}$. $\mathcal{C}$ is said to be closed if $v_0 = v_n$ and simple if no edges intersect one another (except for e_i and e_{i+1} at their common vertex v_i , as well as e_1 and e_n at their common vertex v_0 if $\mathcal{C}$ is closed).

Definition 2.2. A closed polygonal chain $\mathcal{C}$ separates $\mathbb{R}^2$ into connected components. The union of $\mathcal{C}$ and all such bounded components forms a polygon $\mathcal{P}$. The edges and vertices of $\mathcal{C}$ are also called edges and vertices of $\mathcal{P}$.

Ideally, we would not have to be so formal about the definition of a polygon, as most polygons we will consider are associated with simple, closed polygonal chains (truly, polygons in the usual, expected sense). However, it will be necessary at times to consider a polygon with consecutive edges having the same direction, a feature usually disallowed in a polygon by removal of the common vertex. Even worse, some considered polygons will be made from a closed polygonal chain which is not quite

simple, but has one or more instances of an “antenna,” a consecutive pair of edges e_i and e_{i+1} having $v_{i-1} = v_{i+1}$. See Figure 2.1.

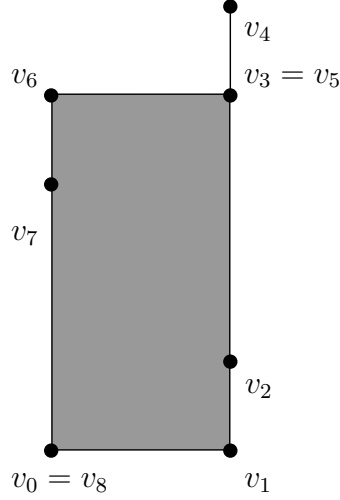


Figure 2.1: Polygon with antenna and adjacent parallel edges.

In all the following, $\mathcal{P} \subset \mathbb{R}^2$ is a finite n -sided polygon (possibly with antennas), where $n \geq 4$ is even, and lengths of edges come in pairs, i.e. the set of edges with any given length has even cardinality.

Definition 2.3. A surface is a two-dimensional topological manifold, that is, a topological manifold which is locally homeomorphic to $\mathbb{R}^2$.

Definition 2.4. A flat surface M is a surface obtained from pairwise isometric identification of the edges of some polygon $\mathcal{P}$.

Identifying edges also causes identification of vertices. A flat surface has a natural metric that makes it locally isometric to $\mathbb{R}^2$ everywhere except possibly at the points corresponding to those vertices of $\mathcal{P}$ that have a neighborhood isometric to a

neighborhood of the tip of a cone with angle other than 2π . Such points are called cone singularities of a flat surface.

Definition 2.5. *A flat surface whose edges are identified only through translation is called a translation surface.*

Identified edges of a translation surface are necessarily parallel. Consequently, all cone singularities of a translation surface have an angle which is an integer multiple of 2π . A familiar example of a translation surface is a torus, which arises from identifying opposite edges of any parallelogram; see Figure 2.2. However, in this case there are no cone singularities as the vertices all identify to a single point with angle 2π .

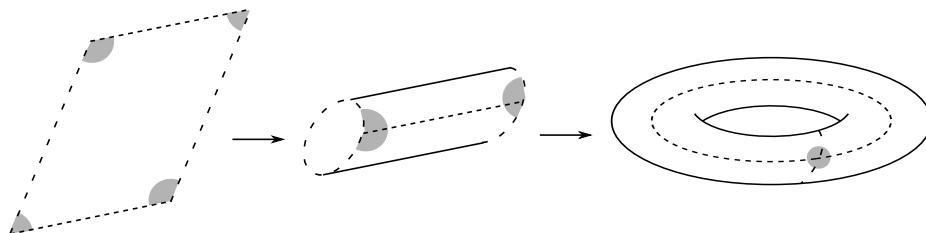


Figure 2.2: Identifying opposite edges of a parallelogram to get a torus; first the top and bottom edges, forming a cylinder, then the ends of the cylinder, forming a torus

In general, if M is a translation surface of genus $g > 1$, it will exhibit one or more cone singularities. This can be seen by considering the Euler characteristic of M . Without going into details, let us perform a quick calculation based on partitioning M into rectangles in such a way that any rectangle which touches the vertex of another rectangle also has that point as its own vertex. Then for each vertex v , let $i(v) \in \mathbb{N}$ be such that the angle at v is $2\pi \cdot i(v)$. Since each vertex is common to $4i(v)$ rectangles

(faces) and $4i(v)$ edges, the Euler characteristic of M is

$$\chi(M) = V - E + F = \sum_v 1 - \frac{1}{2} \sum_v 4i(v) + \frac{1}{4} \sum_v 4i(v) = \sum_v [1 - i(v)]. \quad (2.1)$$

Since also $\chi(M) = 2 - 2g$, for a surface of genus $g > 1$, $\chi(M) < 0$ and thus at least some v must have $i(v) > 1$, i.e., the angle is greater than 2π .

For genus two, there are therefore two possibilities: one point z_0 with angle $2\pi \cdot (2 + 1) = 6\pi$, or two points z_0 and z_1 , each with angle $2\pi \cdot (1 + 1) = 4\pi$. Indeed, all translation surfaces of genus g , $g \geq 2$, can be stratified according to the configuration of their cone singularities. The subcollection of surfaces having n cone singularities, each with angle $2\pi \cdot i(v_k)$, $k = 1, 2, \dots, n$, is denoted

$$\mathcal{H}(i(v_1) - 1, \dots, i(v_n) - 1), \quad (2.2)$$

where by convention singularities are organized such that $i(v_k) \geq i(v_\ell)$ for $k \leq \ell$. The two strata for genus two are therefore $\mathcal{H}(2)$ and $\mathcal{H}(1, 1)$. Examples of polygons corresponding to surfaces in $\mathcal{H}(2)$ are seen in Figure 2.3 and examples of polygons corresponding to surfaces in $\mathcal{H}(1, 1)$ are seen in Figure 2.4.

Definition 2.6. A saddle connection on a translation surface M is a curve whose endpoints are singularities and all other points are regular (non-singular).

Note that for $M \in \mathcal{H}(2)$, all saddle connections are necessarily loops (as there is only one singular point in this case). For $M \in \mathcal{H}(1, 1)$ a saddle connection can either be a loop or a path which connects z_0 and z_1 .

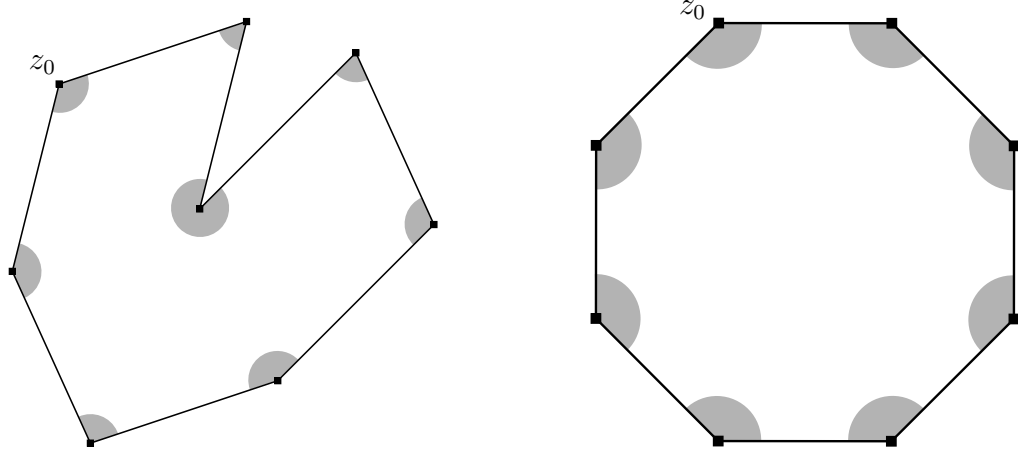


Figure 2.3: Elements of $\mathcal{H}(2)$. Note in each case the marked cone singularity labeled z_0 , with an angle of 6π .

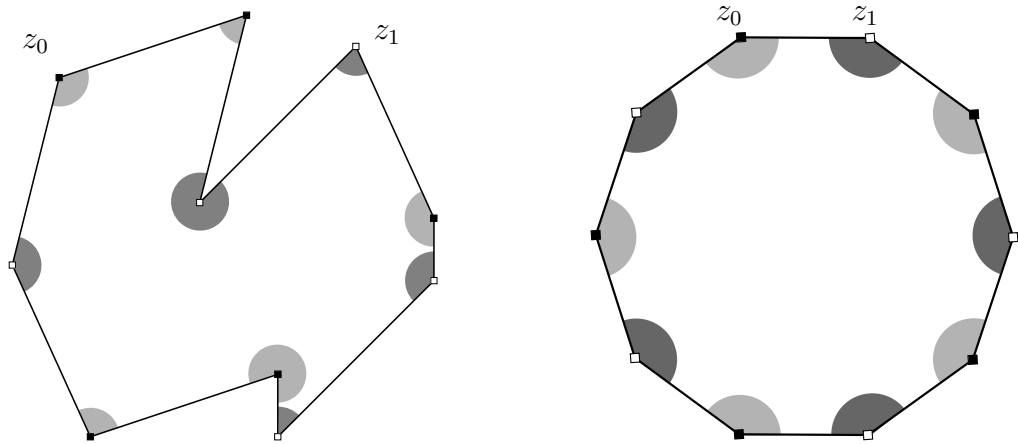


Figure 2.4: Elements of $\mathcal{H}(1,1)$. Note in each case the marked cone singularities labeled z_0 and z_1 , each with an angle of 4π .

Holomorphic One-Forms

A translation surface M can also be thought of as a Riemann surface with a (non-zero) holomorphic one-form ω . [17] includes this as an equivalent definition of translation surface. Indeed, if M is obtained from $\mathcal{P}$, then, identifying $\mathbb{R}^2$ with $\mathbb{C}$ in the usual way, the one-form $dz = dx + idy$ on $\mathcal{P}$ determines a holomorphic one-form ω on M such that ω equals dz on the part of M identified with the interior of $\mathcal{P}$. We shall write ω^s and ω^u for the harmonic forms consisting of the real and imaginary parts, respectively, of ω , so that $\omega = \omega^s + i\omega^u$.

Locally, near $p \in M$, $\omega = f(z)dz$ for some holomorphic function $f : U \rightarrow \mathbb{C}$, where $U \subset \mathbb{C}$ is some open, simply connected neighborhood of 0 and z is the local coordinate with $z = 0$ corresponding to p . Let $F(z)$ be an antiderivative of $f(z)$ on U satisfying $F(0) = 0$. From complex analysis, we know there is a biholomorphic function $H : V \rightarrow W$, where $V \subset U$ and W are open neighborhoods of 0, satisfying $H(0) = 0$ and $F(z) = \frac{H(z)^{k+1}}{k+1}$ for some $k \in \{0, 1, 2, \dots\}$. Thus, locally, $\omega = f(z)dz = F'(z)dz = dF(z) = d\left(\frac{H(z)^{k+1}}{k+1}\right) = H(z)^k dH(z)$, so taking $w = H(z)$ as a new local coordinate, we have $\omega = w^k dw$ near p .

Definition 2.7. *A point p is called a zero of ω if $k > 0$, and the number k is called the degree of the zero at p .*

The horizontal and vertical foliations associated to ω are locally given by the equations $\operatorname{Re} f(z) = 0$ and $\operatorname{Im} f(z) = 0$, respectively. By inspecting the foliations of

$w^k dw$, we see from Figure 2.5 that the angle at p is $2\pi \cdot (k + 1)$. Thus, in particular, ω is zero exactly at the cone singularities of M . We denote by $Z(\omega)$ this set of singularities. In the case of M of genus 2 (where our focus lies), $Z(\omega)$ consists of one or two points (either $\{z_0\}$ or $\{z_0, z_1\}$).

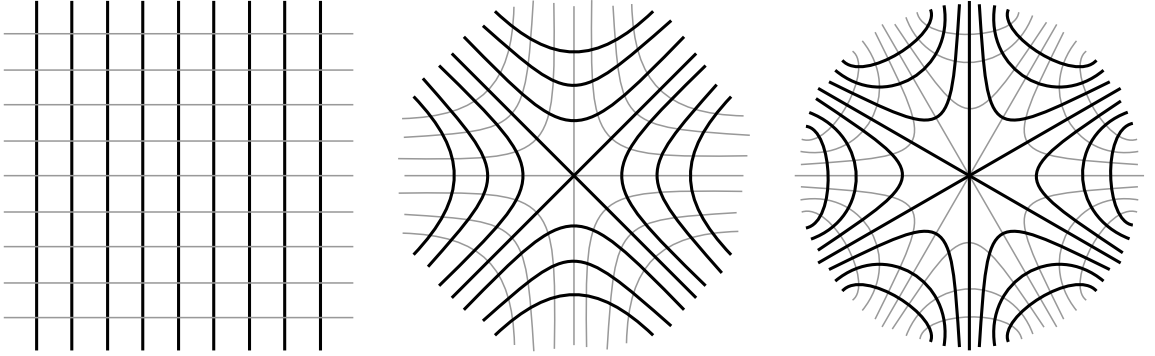


Figure 2.5: Foliations of $w^k dw$ for $k = 0, 1, 2$.

Without going into details (which can be found in [17]), we mention that a holomorphic one-form ω determines a Riemannian metric $|\omega|$ on $M \setminus Z(\omega)$, locally given by $|\omega| = |f(z)||dz| = |f(z)|\sqrt{dx^2 + dy^2}$. This turns $M \setminus Z(\omega)$ into a metric space for which M is a completion, and one can show that M is a translation surface. See also [18] or the surveys [19, 20] for more on the relationship between various perspectives on translation surfaces.

Hyperelliptic Involution

Definition 2.8. A hyperelliptic involution η on a translation surface M of genus $g \geq 2$ is a conformal involution (i.e., a holomorphic map satisfying $\eta^2 = \text{Id}$) that

fixes exactly $2g + 2$ points.

When such an involution exists, it is unique, so henceforth we shall refer to it as *the hyperelliptic involution on M .*

Definition 2.9. *A Riemann surface M of genus $g \geq 2$ is called hyperelliptic if it admits a hyperelliptic involution.*

It is a well-known fact in the theory of Riemann surfaces (see [21], for example), which will be shown by elementary means in Chapter 3, that all surfaces of genus two are hyperelliptic. This fact will prove to be invaluable to our analysis, and one of the primary reasons why consideration of genus two is much simpler than consideration of higher genera. (Another reason is that the combinatorial and geometric complexity increases very rapidly with genus.)

Definition 2.10. *The fixed points of the hyperelliptic involution are called Weierstrass points.*

Surfaces of genus two have $2(2) + 2 = 6$ Weierstrass points. An example of a genus two surface embedded¹ in $\mathbb{R}^3$ and its hyperelliptic involution and Weierstrass points are seen in Figure 2.6. A more detailed analysis of the behavior of the hyperelliptic involution on genus two surfaces, as well as the possible configurations of Weierstrass points, begins on page 71.

¹An embedded surface inherits a Riemannian metric from $\mathbb{R}^3$ and thus acquires a Riemann surface structure (which can be viewed as a class of conformally equivalent Riemannian metrics). It was shown in [22] that all closed Riemann surfaces can be conformally embedded in $\mathbb{R}^3$ and the same for arbitrary Riemann surfaces in [23].

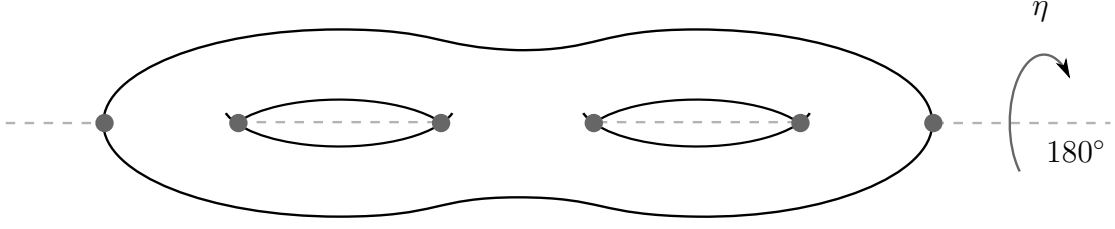


Figure 2.6: Hyperelliptic involution on an example of a genus two surface. The Weierstrass points here are the six points where the axis of rotation by 180° meets the surface.

Connected Sums

We give an informal introduction to the concept of connected sum². Let M_1 and M_2 be two Riemann surfaces. On each $M_i, i = 1, 2$, take a smooth embedding $\gamma_i : [a, b] \rightarrow M_i$ and “slice” each M_i along γ_i , resulting in a slit surface $\tilde{M}_i$. γ_i induces $\gamma_i^+ : [a, b] \rightarrow \tilde{M}_i$ and $\gamma_i^- : [a, b] \rightarrow \tilde{M}_i$ that parametrize the two sides of the slit in $\tilde{M}_i$.

Definition 2.11. The connected sum $M = M_1 \# M_2$ is formed from the union $\tilde{M}_1 \cup \tilde{M}_2$ by identifying $\gamma_1^+(t)$ with $\gamma_2^-(t)$ and $\gamma_1^-(t)$ with $\gamma_2^+(t)$ for all $t \in [a, b]$.

See Figure 2.7. M is taken with the quotient topology. The slits correspond in M to two curves given by $\gamma^+(t) := [\gamma_1^+(t)] = [\gamma_2^-(t)]$ and $\gamma^-(t) := [\gamma_1^-(t)] = [\gamma_2^+(t)]$.

One can show the following:

Proposition 2.12. If M_1 and M_2 have genus g_1 and g_2 , respectively, then the connected sum $M = M_1 \# M_2$ is a Riemann surface of genus $g = g_1 + g_2$,

Definition 2.13. A Riemann surface M with a smooth curve γ in M is said to

²A rigorous exposition in a more general setting required for our purposes appears on page 46.

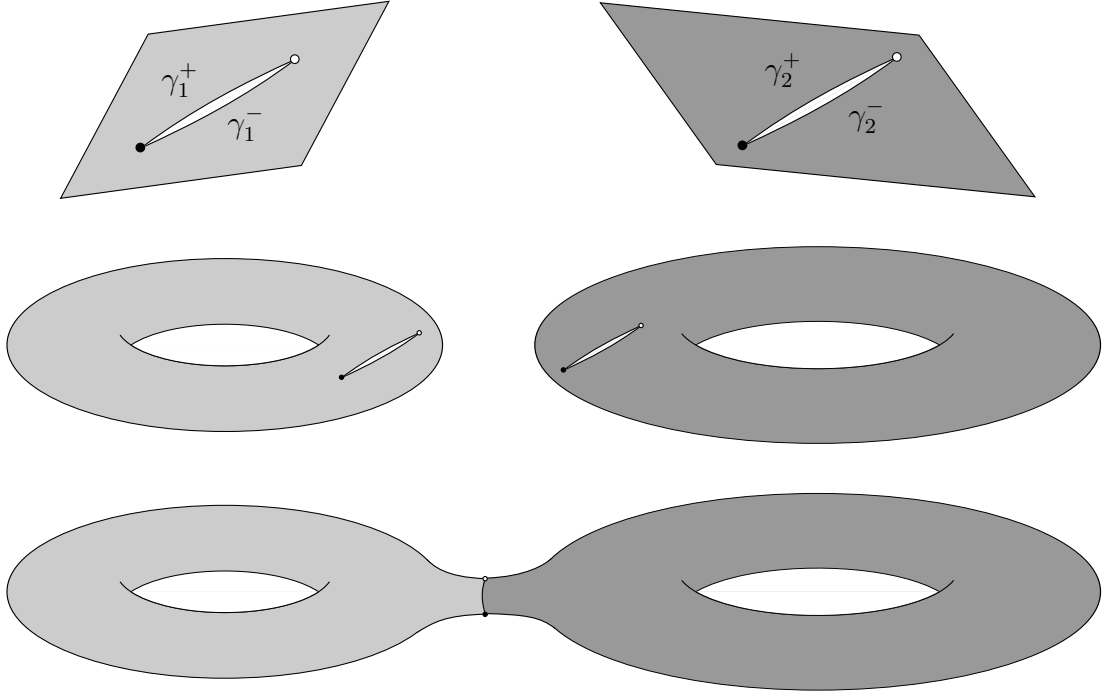


Figure 2.7: Connected sum of tori; first slit each torus, then identify γ_1^+ with γ_2^- and γ_1^- with γ_2^+ .

split along γ as a connected sum if there exist surfaces M_1, M_2 such that M is homeomorphic to the connected sum $M_1 \# M_2$, and γ corresponds via the homeomorphism to the concatenation of γ^+ with the reverse of γ^- .

In this classical definition of a connected sum, the gluing curve γ is an embedding (except if $\gamma(a) = \gamma(b)$). Later we will have to consider situations where a Riemann surface M splits along a curve γ as a (generalized) connected sum, where even $\gamma|_{(a,b)}$ is not simple, as a parametrization of γ passes over the same segment twice. However, with some care taken to define the identifications involved, this still can result in a successfully defined generalized connected sum. This will be explored further in Chapter 4.

CHAPTER 3

POLYBAND CONSTRUCTION AND MCMULLEN'S THEOREM

Polyband Construction

Consider the strip $\mathcal{S} := \{(x, y) : 0 \leq x \leq 1\}$ and the infinite cylinder $\mathcal{S}/\sim$ in $\mathbb{R}^2$ obtained by identifying the boundary lines of $\mathcal{S}$ by the translation $(x, y) \mapsto (x+1, y)$. Let L^+ be a polygonal chain in $\mathcal{S}/\sim$ that is a graph of a continuous function over the *equatorial circle* $E := \{(x, 0) : 0 \leq x \leq 1\}/\sim$ and consists of m ($m \geq 2$) segments (see Figure 3.1, right). Suppose that L^- in $\mathcal{S}/\sim$ is another polygonal chain lying below E and obtained from L^+ by rearrangement of its segments by translations. The topological annulus $\mathcal{A}$ in $\mathcal{S}/\sim$ bounded by L^+ and L^- forms a translation surface upon identifying the corresponding segments of L^+ and L^- .

Definition 3.1. *Call the annulus $\mathcal{A}$ constructed above a polyband and the whole process a polyband construction.*

We assume that L^+ and L^- are given the orientation induced from the standard orientation on E by projecting. Label the edges of L^+ as $L_1^+, \dots, L_m^+$ by going once around L^+ in the positive direction starting from some edge. Label the edges of L^- as $L_1^-, \dots, L_m^-$ where L_i^- is the segment corresponding to L_i^+ under the rearrangement that formed L^- from L^+ . Going around L^- in the positive direction starting from some edge, the encountered edges are $L_{\pi(1)}^-, \dots, L_{\pi(m)}^-$ where π is a permutation of

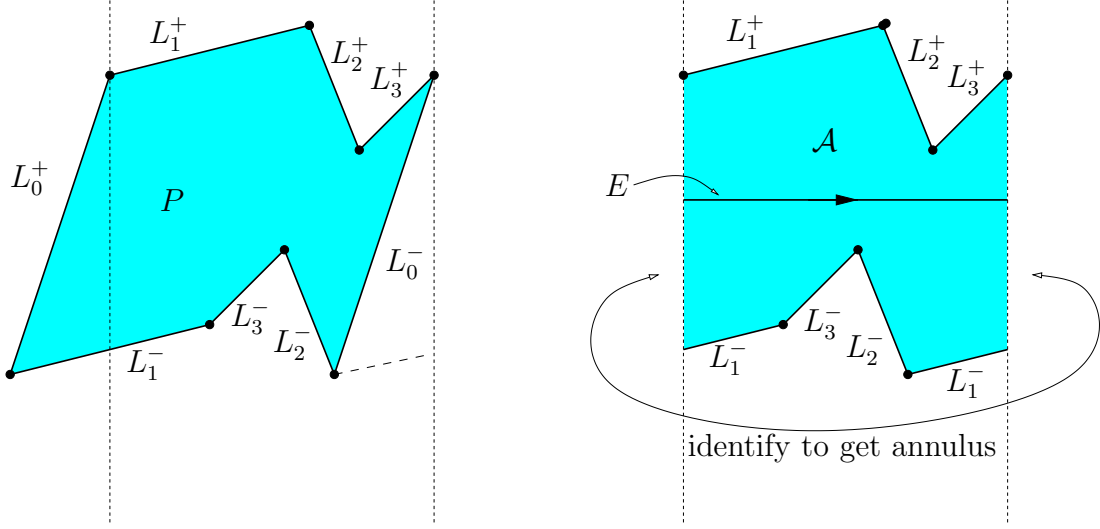


Figure 3.1: $M \in \mathcal{H}(2)$ arises from gluing sides of a polygon P or a polyband $\mathcal{A}$.

$\{1, \dots, m\}$. In Figure 3.1, starting from L_3^- , π is $(3, 2, 1)$, i.e., $\pi(1) = 3$, $\pi(2) = 2$, and $\pi(3) = 1$. This permutation depends on the choice of the starting edges on L^+ and L^- , so to remove this ambiguity, we consider two permutations π and π' as equivalent iff $\pi = c_1 \circ \pi' \circ c_2$ where c_i are some powers of the cyclic permutation $(2, \dots, m, 1)$.

Definition 3.2. The discrete datum of the polyband construction is the equivalence class of permutations, also called a reduced permutation.

We will denote discrete datum by using square brackets; in Figure 3.1 the discrete datum is $[3, 2, 1]$. One can check that for $m = 3$ there are only two distinct reduced permutations: $[1, 2, 3]$ and $[3, 2, 1]$, while for $m = 4$ there are three: $[1, 2, 3, 4]$, $[4, 3, 1, 2]$, and $[4, 3, 2, 1]$. The following result is our way of expressing the combinatorial simplicity of genus two translation surfaces.

Theorem 3.3 (Discrete Datum Theorem). 1. Any $M \in \mathcal{H}(2)$ is isometric to a

translation surface obtained from a polyband construction with discrete datum

$[3, 2, 1]$.

2. *Any $M \in \mathcal{H}(1, 1)$ is isometric to a translation surface obtained from a polyband construction with discrete datum $[4, 3, 2, 1]$.*

To prove the theorem, we use that any translation surface has a regular closed geodesic, i.e. a closed geodesic disjoint from $Z(\omega)$. This innocent fact is not trivial and has been originally shown by Masur with help from Teichmüller theory [24]. A nice elementary proof has been found by Smillie [25] (see also [26]). The crux is in deforming P by an affine transformation so that M has area $A = 1$ and the diameter of M is so large that some point $p \in M$ is further than $1/\sqrt{\pi}$ from the singularities (see Figure 3.2)¹. (Affine transformations map closed geodesics to closed geodesics.) Since $A = 1$, upon increasing $r > 0$ from zero, the r -neighborhood $B_r(p)$ in M centered at p must cease to be an embedded Euclidean disk for some $r_0 \in (0, 1/\sqrt{\pi}]$. For r that is a little larger than r_0 , $B_r(p)$ is still free of singularities but “laps over itself” and thus contains a flat cylinder made of a multitude of parallel regular closed geodesics.

As another standard preliminary, let us fix a regular closed geodesic E and some direction transversal to E . Consider the subset M' of the points p' of M that can be reached from a point of E by traveling along geodesic segments of that fixed direction.

Here we do not insist that a single segment is used for any given point p' and allow

¹The delicate point is that the diameter of M is generally smaller than the diameter of the deformed P .

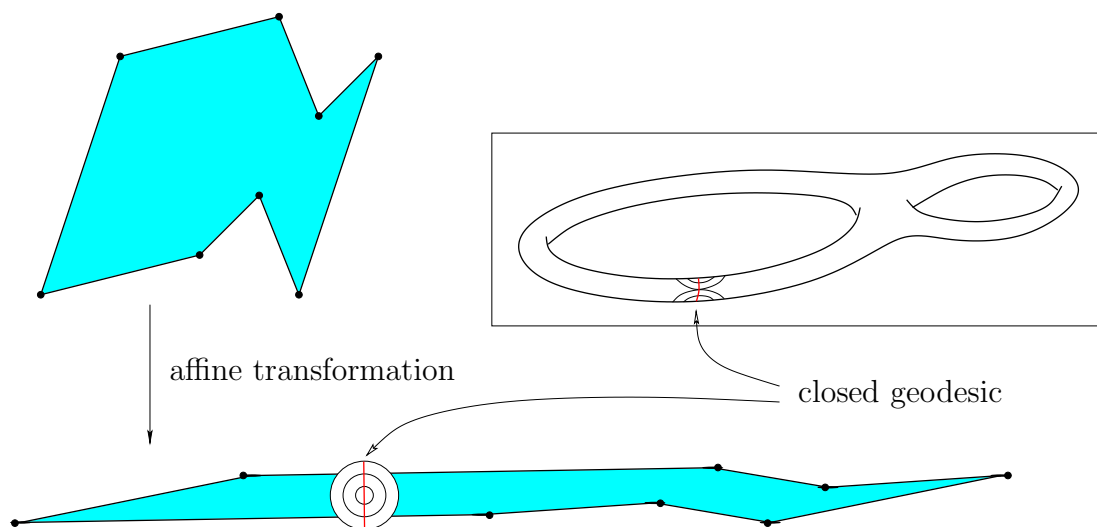


Figure 3.2: Smillie's construction of a (regular) closed geodesic.

unions of segments joined at singularities of M . It is not hard to see that, unless M' coincides with all of M , its boundary must be a union of saddle connections. In particular, by choosing a generic direction, that is, not one of the countably many directions of all possible saddle connections, one can ensure that $M' = M$. We are now ready to prove Theorem 3.3.

Proof of Theorem 3.3. Consider a translation surface M of genus two. Let E be a regular closed geodesic. Pick a generic direction transversal to E as in our preliminary discussion so that $M' = M$ and every geodesic segment in the chosen direction contains at most one singularity. It is convenient to apply an affine transformation to make the direction vertical and E horizontal.

Consider the first return map $T : E \rightarrow E$ under the vertical flow, i.e., the movement of points of M with unit speed in the vertical direction, which is unambiguously

defined except at the singularities, which have several outgoing verticals. T is an arc exchange: it is a well defined local isometry apart from the finitely many points p whose outgoing vertical geodesic hits a singularity s before returning to E . T acts by cutting E at all such “cut points” p and rearranging the resulting arcs in E by translations.

To be precise, for small $\varepsilon > 0$, the horizontal arc $(p - \varepsilon, p + \varepsilon)$ centered at a cut point p flows vertically intact and sweeps a rectangle in M until it hits a singularity s where the rectangle is slit along two verticals outgoing from s and forming angle 2π in M . (By our choice of direction, the slitted rectangle has its two parts returning to E without further encounters with singularities, granted $\varepsilon > 0$ is small enough.) In particular, a small portion of the slitted rectangle forms a small radius 2π -sector with a tip at s (darkly shaded in Figure 3.3). Such sectors for different p are disjoint and (their closures) form a neighborhood of the set of singularities. Therefore, if $M \in \mathcal{H}(2)$ then there are $6\pi/2\pi = 3$ cut points p in E and if $M \in \mathcal{H}(1, 1)$ then there are $(4\pi + 4\pi)/2\pi = 4$ cut points.

Let us focus on the case $M \in \mathcal{H}(2)$ when E is cut into three (open) arcs E_1, E_2, E_3 (Figure 3.3). Each arc E_i sweeps an open rectangle R_i before returning to E . The vertical portions of the boundaries of any two adjacent rectangles $\overline{R}_i$ and $\overline{R}_{i+1}$ that are below the singularity impact are identified in M .

Definition 3.4. *The disjoint union $\mathcal{R}$ of $\overline{R}_1, \overline{R}_2$, and $\overline{R}_3$ with those identifications is called a tower of rectangles.*

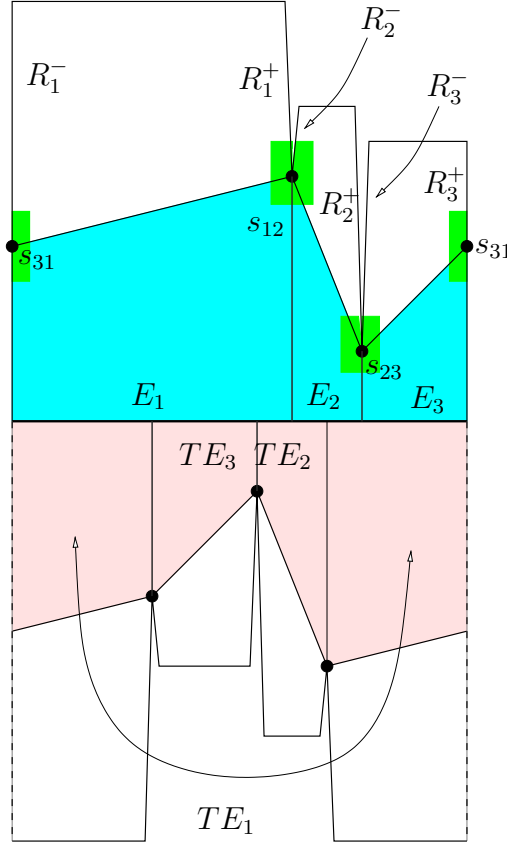


Figure 3.3: For $M \in \mathcal{H}(2)$, there are three cut points splitting E into E_1, E_2, E_3 . The tower of three rectangles above E rearranges into a polyband. The $[3, 2, 1]$ datum is forced because the three slitted rectangles (one presented in two pieces) have to glue so that a neighborhood of the sole singularity with angle 6π is homeomorphic to a disk.

As discussed, the vertical sides of the tower (above the singularities) are identified so that M contains a neighborhood of s with the angle at s equal to 6π . In Figure 3.3, the neighborhood is glued together from the three small darkly shaded slitted rectangles labeled by the three tower vertices s_{12}, s_{23}, s_{31} they abut.

In circumnavigating the 6π -singularity $s \in M$ clockwise, one must visit the slitted rectangles in one of the two (cyclic) orders s_{12}, s_{23}, s_{31} or s_{12}, s_{31}, s_{23} . The second order is simply the reversal of the first so we deal only with the first. (Both lead to the

same discrete datum $[3, 2, 1]$.) Looking at Figure 3.3, we see that the first order forces the identifications of the vertical sides of $\mathcal{R}$ to be $R_1^+ \leftrightarrow R_3^-$, $R_2^+ \leftrightarrow R_1^-$, $R_3^+ \leftrightarrow R_2^-$. That means that upon returning to E , the E_i appear in the cyclic order E_3, E_2, E_1 along E . (Indeed, E_3 is followed by E_2 to its right and E_2 is followed by E_1 .)

Finally, to uncover the polyband (still looking at Figure 3.3), we let L^+ be the polygonal chain obtained by joining s_{12} to s_{23} by a segment in R_2 and s_{23} to s_{31} by a segment in R_3 and s_{31} to s_{12} by a segment in R_1 . Cutting $\mathcal{R}$ along L^+ and moving the top pieces below E yields a polyband representing M . It is bounded by L^+ and its rearrangement L^- with discrete datum $[3, 2, 1]$.

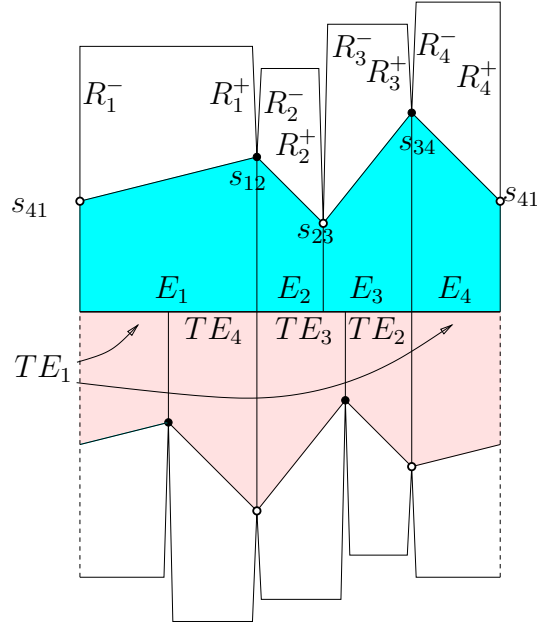


Figure 3.4: For $M \in \mathcal{H}(1, 1)$, the tower of four rectangles above E rearranges into a $[4, 3, 2, 1]$ polyband.

This finishes the proof for $M \in \mathcal{H}(2)$ and we move to the case when $M \in \mathcal{H}(1, 1)$,

which proceeds along the same lines except that now E is divided into four subarcs E_1, E_2, E_3, E_4 . Each of the two singularities corresponds to a pair of the four vertices of $\mathcal{R}$ with angle 2π (named $s_{12}, s_{23}, s_{34}, s_{41}$ according to the rectangles they belong to, see Figure 3.4).

First observe that such a pair cannot be formed by two vertices that are adjacent (i.e. belong to the same rectangle R_i). Indeed, if that were the case and, say s_{12} and s_{23} were identified in M , then the vertical sides of R_2 above the singularities would be identified as well and so $T(E_2)$ would have to be an arc whose endpoints coincide in M , contradicting $T(E_2)$ being a proper subarc of a simple closed curve E . Therefore, it must be that s_{12} and s_{34} identify to one singularity and s_{23} and s_{41} to the other, as depicted in Figure 3.4. The identifications of the vertical sides of the tower are then given by $R_1^+ \leftrightarrow R_4^-, R_3^+ \leftrightarrow R_2^-, R_2^+ \leftrightarrow R_1^-, R_4^+ \leftrightarrow R_3^-$. This means that, upon returning to E , the arcs E_1, E_2, E_3, E_4 appear in the order $T(E_4), T(E_3), T(E_2), T(E_1)$. As before, cutting the tower along the polygonal chain L^+ joining $s_{12}, s_{23}, s_{34}, s_{41}$ leads to a rearrangement into a polyband representing M and with discrete datum $[4, 3, 2, 1]$. $\square$

Hyperellipticity

As the first application of Theorem 3.3, we give a geometric reason for the hyperellipticity of all genus two Riemann surfaces — a property usually derived via the Riemann-Roch Theorem, [21]. Intuitively, this follows from the central symmetry

in the polybands in Figures 3.3 and 3.4. Let us explain and provide some complex analytic context for this fact.

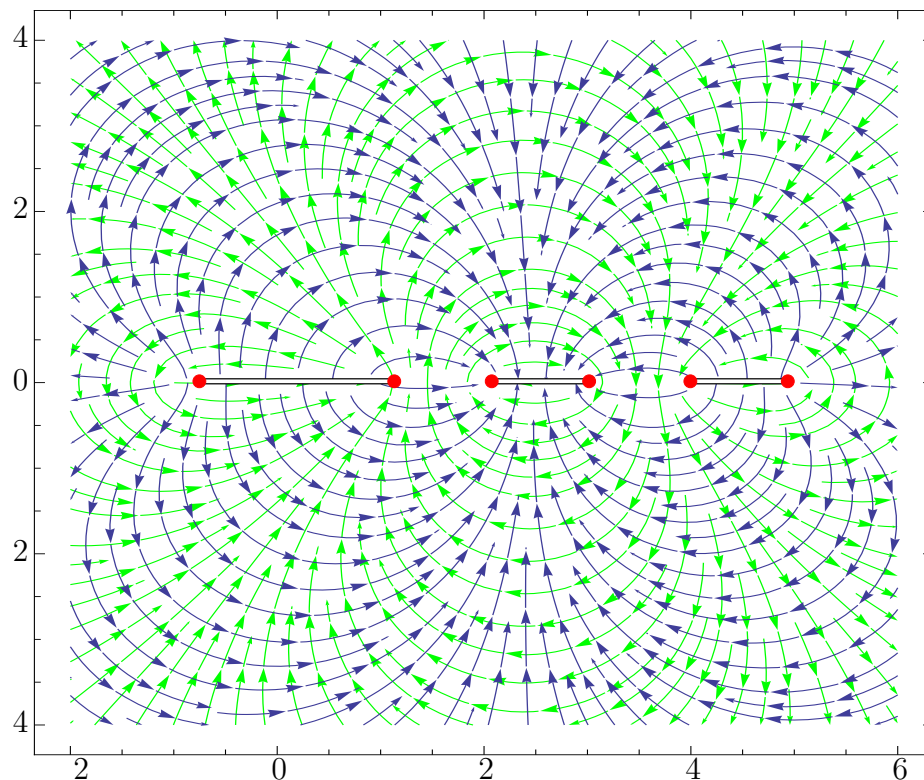


Figure 3.5: The horizontal and vertical lines of the 1-form $\omega = w^{-1}dz$ on one of the two copies of the triply slitted plane making up the Riemann surface of $w = \sqrt{(z+1)(z-1)(z-2)(z-3)(z-4)(z-5)}$. The real part of ω is $\text{Re}(w^{-1})dx - \text{Im}(w^{-1})dy$ and the darker flowlines are those of the corresponding vectorfield $(\text{Re}(w^{-1}), -\text{Im}(w^{-1}))$. The lighter flowlines similarly correspond to the imaginary part of ω . (Here x and y are the standard coordinates in the depicted z -plane, not the polyband $\mathcal{A}$.)

Recall that the Riemann surface M of the (two-valued) analytic function

$$w = \sqrt{(z - a_1)(z - a_2) \cdots (z - a_n)}$$

(where the a_i are distinct) is conformally equivalent to the Riemann sphere when $n = 1, 2$ and to the complex torus (a quotient of $\mathbb{C}$ by a lattice) if $n = 3, 4$. In the latter

case, M is called an elliptic curve. The hyperelliptic curves are the Riemann surfaces obtained by taking $n > 4$. They are of the smallest genus $g = 2$ for $n = 5, 6$. An abstract Riemann surface M is hyperelliptic if and only if it is conformally equivalent to a hyperelliptic curve, and this is characterized by existence of the hyperelliptic involution described on page 16. Indeed, if such involution η exists, then the quotient M/η is equivalent to the Riemann sphere², and the inverse of the natural factor map $\pi : M \rightarrow M/\eta$ is essentially the square root function above, with the action of η corresponding to the choice of sign.

To connect with our development, recall (from the discussion on page 15) that an abstract Riemann surface M can be turned into a translation surface by fixing a holomorphic one-form³ ω on M and letting its real and imaginary parts play the roles of the coordinate forms dx and dy (so ω corresponds to $dz = dx + idy$ in the $M = P/\sim$ or $M = \mathcal{A}/\sim$ presentations). Against this backdrop, note that, due to $1, 2, 3, \dots$ and $\dots, 3, 2, 1$ being “flips” of each other, any polyband $\mathcal{A}$ with datum $[3, 2, 1]$ or $[4, 3, 2, 1]$ is left invariant under an isometry of the infinite cylinder $\mathcal{S}/\sim$ that rotates it by 180° about a suitable point, i.e., a mapping given by $(x, y) \mapsto (-x + a, -y + b)$ where x is mod 1 and the parameters a, b are picked so that each segment of L^+ goes to the corresponding segment of L^- . (For the polybands constructed in the proof of Theorem 3.3 and depicted in Figures 3.3 and 3.4, $b = 0$ and a is easy to guess.) We

²as can be seen by using Riemann-Hurwitz Formula

³Such forms correspond to incompressible irrotational flows on M . One can be usually constructed by hand in any given example (say dz/w in ours) and its a priori existence is *clear* to a physicist. Mathematicians construct them from suitable harmonic functions obtained either as limits of subharmonic functions or as minimizers of a Dirichlet integral, see [21].

refer to this isometry as the central symmetry of $\mathcal{A}$ — although, it has two fixed points: $(a/2, b/2)$ and $(a/2 + 1/2, b/2)$.

The central symmetry respects the boundary identifications in $\mathcal{A}$ that glue it into the surface M and thus induces an isometry $\eta : M \rightarrow M$ that has exactly 6 fixed points: the two in the interior of the polyband $\mathcal{A}$, one at the center of each side, and the singularity if $M \in \mathcal{H}(2)$. For $M \in \mathcal{H}(1, 1)$, the two singularities are interchanged. Viewed on the original Riemann surface, η is conformal — obviously so away from the singularities but also there (if only because isolated singularities of conformal maps are removable). We established the following classical result:

Corollary 3.5 (Hyperellipticity). *Any Riemann surface M of genus $g = 2$ is hyperelliptic: it admits a conformal involution $\eta : M \rightarrow M$ with 6 fixed points.*

Before moving on, let us offer the following visually appealing presentation of any $M \in \mathcal{H}(2)$ illustrated by Figure 3.6.

Theorem 3.6 (Stamped Torus). *Any $M \in \mathcal{H}(2)$ is isometric to the translation surface obtained by identifying the opposite sides of a parallelogram (perhaps degenerate) in a flat torus (with the interior of the parallelogram removed).*

Proof. By Theorem 3.3, we may suppose that M is obtained from a polyband construction. First assume that L^+ is not straight (i.e., not all of its segments have the same direction). For each vertex of L^+ , consider the (possibly degenerate) triangle

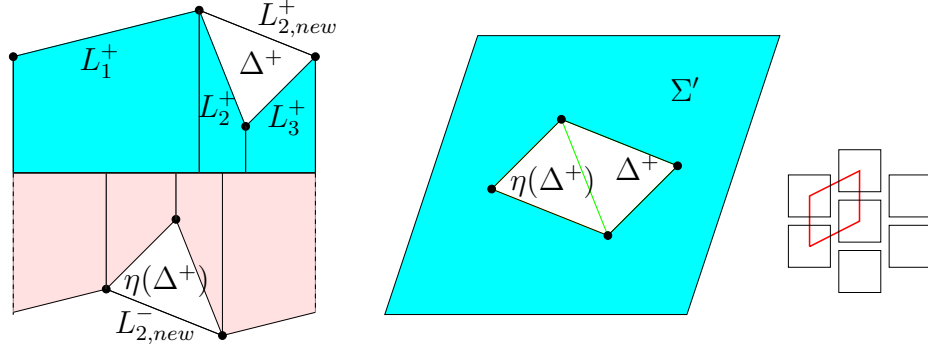


Figure 3.6: Any $M \in \mathcal{H}(2)$ (left) arises from a flat torus (center) with a parallelogram removed (“stamped out”). Generally (right), the removed parallelogram need not fit into any fundamental parallelogram.

spanned by this vertex and the adjacent vertices. One of the triangles has to be non-degenerate and have interior above the polyband — as otherwise L^+ would be convex and thus straight. We may as well suppose that the two sides of this triangle Δ^+ are L_2^+ and L_3^+ (as in Figure 3.6). By adjoining Δ^+ and its symmetric counterpart $\eta(\Delta^+)$ to the polyband we get a new polyband with the discrete datum $[1, 2]$, which yields a flat torus. The desired parallelogram is obtained by identifying Δ^+ and $\eta(\Delta^+)$ along their edges that are not in $L^\pm$. In the case when L^+ is straight (and thus horizontal), a degenerate parallelogram $L_2^+ \cup L_3^+ = L_3^- \cup L_2^-$ can be used. $\square$

The situation depicted on the right of Figure 3.6 serves as a warning that some $M \in \mathcal{H}(2)$ cannot be presented as a “big” parallelogram with a “small” parallelogram stamped out (as it happened in the center of Figure 3.6).

McMullen’s Theorem

We turn to the main objective of this chapter, which is the following restatement

of the key combinatorial result in [1] (Theorem 7.4 lumped together with a much easier Theorem 7.3). It implies that any genus two translation surface can be constructed from two slitted parallelograms in a way suggested by Figure 2.7. Observe that the depicted surface is in $\mathcal{H}(1,1)$. To get a surface in $\mathcal{H}(2)$ one of the slits has to start and end at the same point of the torus.

Theorem 3.7 (McMullen). *Any translation surface M of genus $g = 2$ contains a saddle connection J such that $\eta(J) \neq J$ (where η is the hyperelliptic involution). The surface M splits along $J \cup \eta(J)$ into a connected sum of two tori. Moreover, given a (maximal) open flat cylinder C , such J can be found so that neither J nor $\eta(J)$ cross ∂C (although they may be contained in ∂C or have endpoints in ∂C).*

Proof. Given a genus $g = 2$ surface M and a cylinder C , we can instantiate the proof of Theorem 3.3 with E being the equator of C to present M via the polyband construction so that C is the maximal horizontal cylinder in the polyband $\mathcal{A}$ — the shaded region in Figures 3.8 and 3.9.

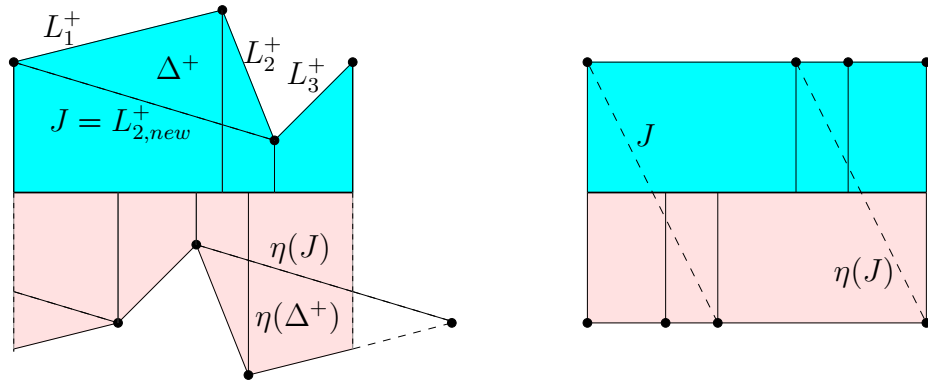


Figure 3.7: The construction of J for $M \in \mathcal{H}(2)$; the degenerate case on the right.

We start with the easier case when $M \in \mathcal{H}(2)$. As in the previous argument, we first assume that L^+ is not straight and consider the triangles spanned by each vertex of L^+ together with the two adjacent vertices. Again, one such triangle, call it Δ^+ , has to be non-degenerate and contained in $\mathcal{A}$ (see Figure 3.7). The desired J is the lower side of Δ^+ (the side not in L^+). Indeed, cut away from $\mathcal{A}$ along $J \cup \eta(J)$, Δ^+ and $\eta(\Delta^+)$ naturally identify to form a torus and $\mathcal{A} \setminus (\Delta^+ \cup \eta(\Delta^+))$ is a new polyband with the discrete datum $[2, 1]$, yielding a torus as well. For the case when L^+ is straight, see Figure 3.7.

It remains to deal with the case when $M \in \mathcal{H}(1, 1)$. The situation is a bit more complicated due to L^+ containing four segments. L^+ is a graph of a function over E and we may well assume that s_{41} is the minimum, the lowest point of L^+ . The segment L_1^+ must then go *up*, i.e., have a non-negative slope. Subsequent segments L_2^+, L_3^+ can go *up* or *down* and L_4^+ has to go *down*, i.e., have a non-positive slope. Below we consider the four possibilities for the slopes of $L_1^+, L_2^+, L_3^+, L_4^+$: *up, up, up, down*; *up, down, down, down*; *up, up, down, down*; and *up, down, up, down*. (These are not exactly mutually exclusive because some slopes may be zero.) We first restrict attention to the “non-degenerate” situation when the minimum on L^+ is attained at the sole location s_{41} .

up, up, up, down: Let $J = J_{234}$ connect s_{12} to the “rightmost” point of L^+ , $s_{41, \text{right}}$ in Figure 3.8. Observe that the segments J, L_2^+, L_3^+, L_4^+ bound a quadrilateral Q^+ . Q^+ and $\eta(Q^+)$ glue (along J and $\eta(J)$) to form a hexagon, which identifies to

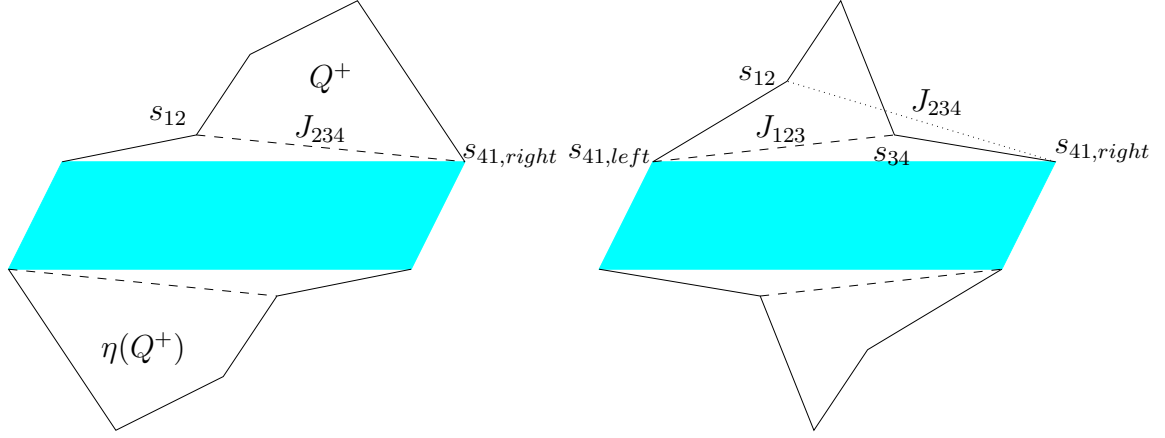


Figure 3.8: *up, up, up, down* and *up, up, down, down* (non-degenerate cases).

a torus. What remains of the polyband $\mathcal{A}$ has discrete datum $[2, 1]$ so also forms a torus.

up, down, down, down: This case is (vertical) axis-symmetric to the previous one.

up, up, down, down: Consider J_{234} as in the first case and its symmetric counterpart J_{123} connecting the leftmost point $s_{41,left}$ of L_4^+ to s_{34} , see Figure 3.8. $J = J_{234}$ is as desired unless it is not entirely contained in the polyband $\mathcal{A}$, which is when s_{34} is below J_{234} . Likewise, $J = J_{123}$ works unless s_{12} is below J_{123} . It is impossible for s_{12} and s_{23} to be both “below.”

up, down, up, down: Now s_{23} is a local minimum on L^+ (but it is not in ∂C by our temporary non-degeneracy assumption). It is really easy to find a suitable J that crosses ∂C ; see Figure 3.9. For J that does not cross ∂C we have to work a bit harder.

Let K^+ be the triangle in $\mathcal{A}$ with L_3^+ and L_4^+ for its two sides. Upon identifying a pair of parallel sides in $K^+ \cup \eta(K^+)$ we get a cylinder $K \subset M$. Consider the oriented

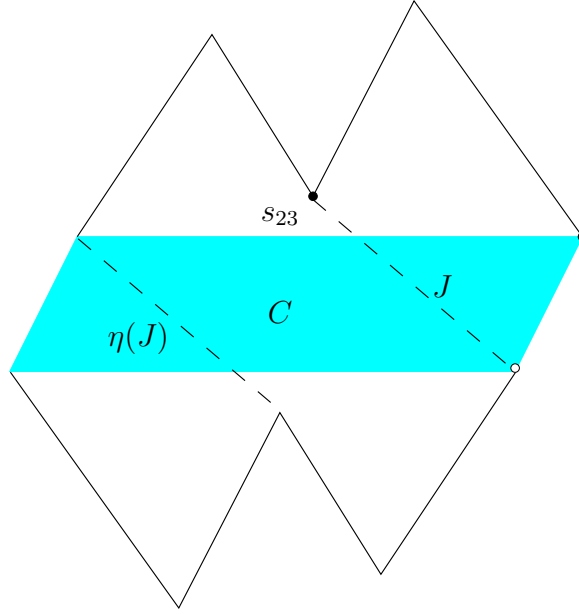


Figure 3.9: *up, down, up, down* (non-degenerate case): an “easy” J could be found if it were allowed to cross ∂C .

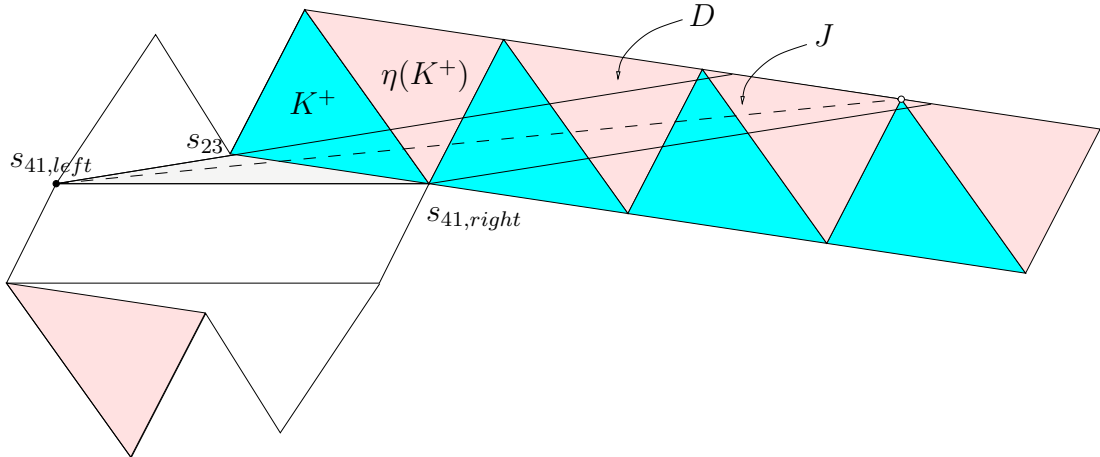


Figure 3.10: *up, down, up, down* (non-degenerate case): K^+ and $\eta(K^+)$ form a cylinder, which is shown “unfolded” to unveil the trapezoid inside which J is found.

half-line D originating from $s_{41, \text{left}}$ and passing through s_{23} . The portion of D in the cylinder K cuts K into a parallelogram. The union of this parallelogram and the triangle with vertices $s_{41, \text{left}}, s_{23}, s_{41, \text{right}}$ forms a trapezoid (see Figure 3.10). Note that the side opposite to the $s_{41, \text{left}}, s_{41, \text{right}}$ -side in the trapezoid must contain a point that represents a singularity.

Let J be the segment in the trapezoid connecting that point to $s_{41, \text{left}}$. Note that $\eta(J) \neq J$, if only because, presented⁴ in $\mathcal{A}$, $\eta(J)$ ends at $\eta(s_{41, \text{right}})$ which is not in J . From $\eta(J) \neq J$, it already follows that M cut along $\gamma := \eta(J) \cup J$ disassociates into two slitted tori (cf. Theorem 7.3 in [1]). Indeed, since η acts on the first homology $H_1(M, \mathbb{Z})$ by $-\text{Id}$ and $\eta(\gamma) = \gamma$, γ is a null-homologous loop and thus cuts M into two subsurfaces of lower genus. For concreteness, Figure 3.11 identifies those two tori explicitly.

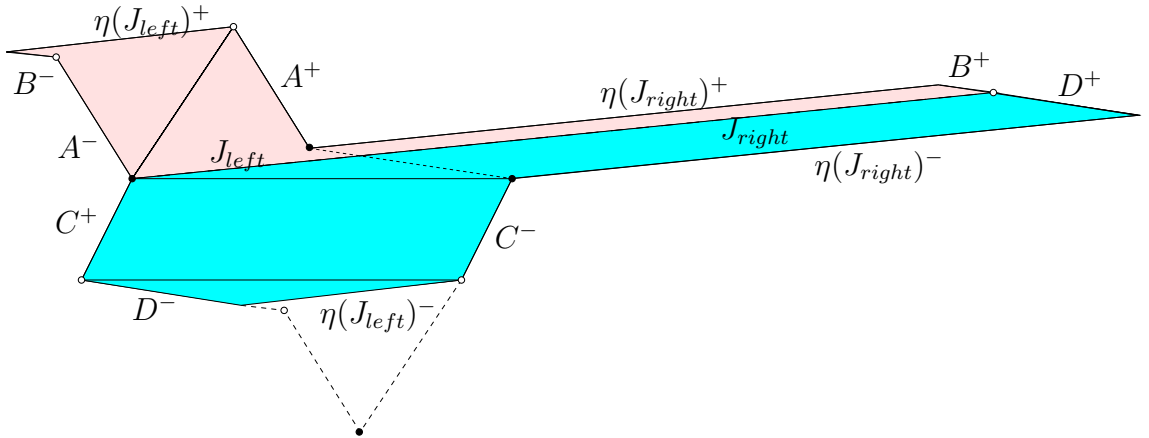


Figure 3.11: J and $\eta(J)$ cut M into two tori (presented by the two shaded hexagons).

The proof is complete save for the degenerate cases when s_{41} is not the only vertex

⁴We abuse the notation η : in the context of $\mathcal{A}$, η is the *obvious central symmetry*.

of L^+ in ∂C . These are all easy and taken care of in Figure 3.12.

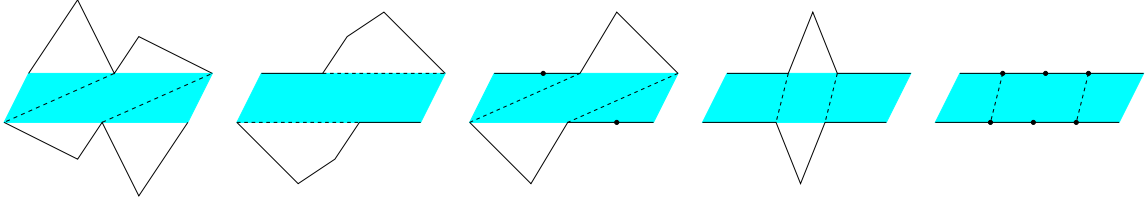


Figure 3.12: Degenerate cases with 2, 3, and all 4 vertices of L^+ in ∂C . (The second and third pictures are essentially the same.)

□

CHAPTER 4

CONNECTED SUMS USING L-CUTS

Definitions

In the following paragraphs, we formally define L-cuts, which enter the principal result, Theorem 4.7. In brief, an L-cut is a generalization of a straight saddle connection: it consists of one vertical segment followed by one horizontal segment (or vice versa). Recall from Chapter 2 (page 15) that $\omega = \omega^s + i\omega^u$ is a holomorphic one-form on M . Foliations of ω^s are horizontal and of ω^u are vertical. We begin by defining what it means for a segment to be “vertical” or “horizontal,” in the sense that it follows level lines of ω^s and ω^u , respectively¹.

Definition 4.1. A parametrized vertical segment on M is a continuous $\gamma : [a, b] \rightarrow M$ such that, for finitely many points $a = c_0 < c_1 < c_2 < \dots < c_n = b$, $\gamma_i := \gamma|_{(c_i, c_{i+1})}$ is one-to-one for $i \in \{0, \dots, n-1\}$ and $\int_{\gamma_i|_{(t, t')}} \omega^s = 0$ for all $t, t' \in (c_i, c_{i+1})$.

Note that $\int_{\gamma_i|_{(t, t')}} \omega^s$ stands for $\int_t^{t'} \omega^s(\gamma_i(\tau))[\dot{\gamma}_i(\tau)]d\tau$, where $\omega^s(a)[v]$ is the one-form ω^s evaluated at a and applied to the vector v . By requiring that $\omega^s(= \operatorname{Re} \omega)$ is zero on this segment, we see that the direction of the segment has no horizontal component (i.e., it is vertical).

¹In charts, ω^s corresponds to dx , so its level lines will be vertical, and ω^u corresponds to dy , so its level lines will be horizontal.

Definition 4.2. A vertical segment on M is an equivalence class of parametrized vertical segments where $\gamma_1 \sim \gamma_2$ if and only if there is some orientation-preserving homeomorphism $h : [a, b] \rightarrow [\tilde{a}, \tilde{b}]$ with $\gamma_2 = \gamma_1 \circ h$.

A parametrized horizontal segment and horizontal segment are defined analogously by replacing ω^s with ω^u .

An L-cut is now merely a concatenation of a vertical and a horizontal segment forming a path connecting the cone singularities of M .

Definition 4.3. A parametrized L-cut on M is given by a continuous $\gamma : [a, b] \rightarrow M$ such that $\gamma(a) \in Z(\omega), \gamma(b) \in Z(\omega)$, and for some $c \in [a, b]$, $\gamma|_{[a, c]}$ is a parametrized vertical (or horizontal) segment and $\gamma|_{[c, b]}$ is a parametrized horizontal (respectively, vertical) segment. We say the parametrized L-cut is degenerate if either $c = a$ or $c = b$, so that the parametrized L-cut is really a parametrized horizontal or vertical segment.

For our purposes, it will be useful to only consider L-cuts whose endpoints are different cone singularities of M , if more than one such singularity exists on M . Thus, for $M \in \mathcal{H}(1, 1)$, add a condition that $\gamma(a) \neq \gamma(b)$.

Definition 4.4. An L-cut is an equivalence class of parametrized L-cuts where $\gamma_1 \sim \gamma_2$ if and only if there is some orientation-preserving homeomorphism $h : [a, b] \rightarrow [\tilde{a}, \tilde{b}]$ with $\gamma_2 = \gamma_1 \circ h$.

Examples of L-cuts are seen in Figure 4.1. Note that the top surface is in $\mathcal{H}(2)$,

thus the L-cut is a loop. As above, we say the L-cut is *degenerate* if one of its segments is degenerated to a point, so the L-cut is really a horizontal or vertical saddle connection.

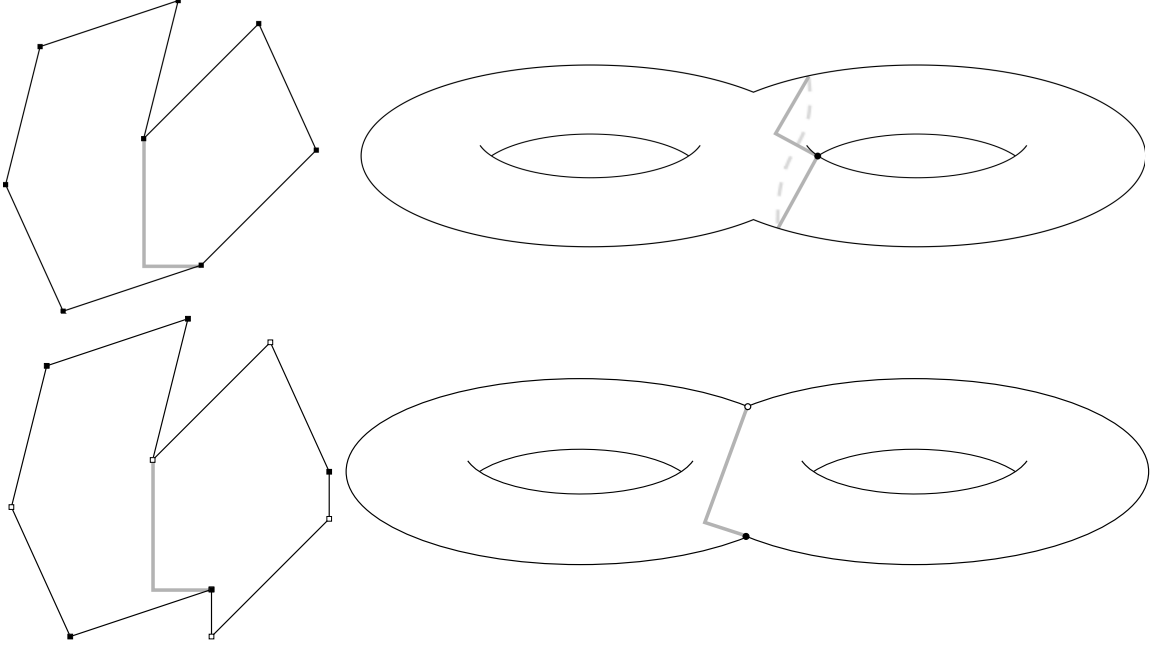


Figure 4.1: L-cuts on $M \in \mathcal{H}(2)$ (top) and $M \in \mathcal{H}(1,1)$ (bottom). Note that for $M \in \mathcal{H}(2)$, the L-cut necessarily forms a loop, while for $M \in \mathcal{H}(1,1)$ this is specifically disallowed.

In this dissertation, for most $M \in \mathcal{H}(1,1)$, it is possible to find γ which is one-to-one on $[a, b]$. For most $M \in \mathcal{H}(2)$, γ can be found which is one-to-one on (a, b) (while $\gamma(a) = \gamma(b) = z_0$, the sole singular point of M). However, for some M we will have to use L-cuts which pass through a singularity at one or more points other than the endpoints; these are worthy of special attention now to clarify how they look in the surface M .

Definition 4.5. An L-cut is called singular if $\gamma(t) \in Z(\omega)$ for some $t \in (a, b)$.

See Figure 4.2 for an illustration of a singular L-cut. Note that existence of a singular L-cut on M implies that there is a horizontal or vertical saddle connection on M . We also observe that singular L-cuts appear only for a non-generic set of translation surfaces within any stratum $\mathcal{H}(\dots)$.

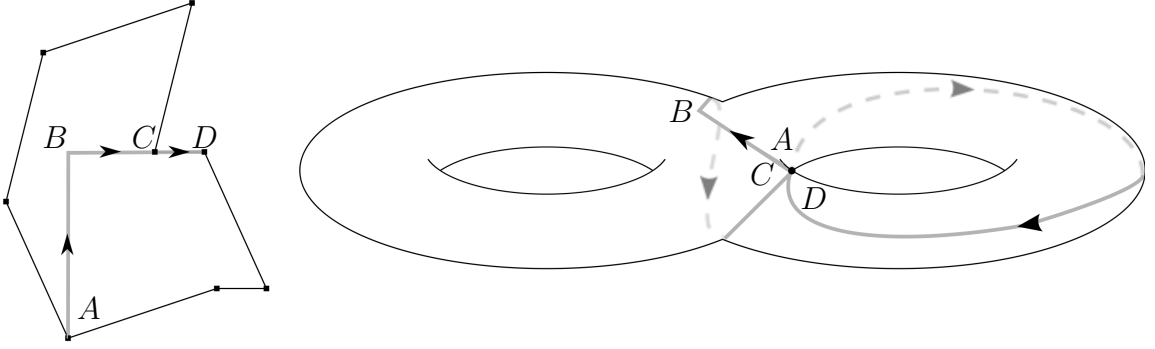


Figure 4.2: A singular L-cut; the path starts at A , turns at B , continues past C , and ends at D . Note that A, C , and D are all the same point (the cone singularity).

Our primary use of L-cuts will be to use a pair of them to represent a Riemann surface M as a connected sum. This pair will satisfy the following definition:

Definition 4.6. The pair of L-cuts (K, K') given by parametrizations $\gamma : [a, b] \rightarrow M$ and $\gamma' : [a, b] \rightarrow M$ is called parallel (equivalently, K is parallel to K') if and only if $\gamma(a) = \gamma'(a)$, $\gamma(b) = \gamma'(b)$, and the loop $K'K^{-1}$ is null-homologous.

Note that $K'K^{-1}$ being null-homologous means that $\int_{K'K^{-1}} \omega = 0$, so $\int_{K'} \omega = \int_K \omega$. In particular, upon choosing suitable parametrizations γ and γ' of K, K' , we

have $\int_{\gamma|_{[a,t]}} \omega = \int_{\gamma'|_{[a,t]}} \omega$ for all $t \in [a, b]$.

Stated more plainly, this means that two parallel L-cuts K and K' start at the same point (although they follow different verticals/horizontals emanating from this singular point), end at the same point, and have the same “shape” and segment lengths. Indeed, enforcing equality of integrals of $\omega = \omega^s + i\omega^u$ along K and K' ensures that both the vertical segment and horizontal segment comprising K are of identical length to those comprising K' . For a generic M , two L-cuts of the same shape and segment lengths are parallel. A pair of parallel L-cuts is seen in Figure 4.3, and Figure 4.4 shows an example of a (non-generic) surface having three L-cuts of the same shape and segment lengths, only two of which are parallel.

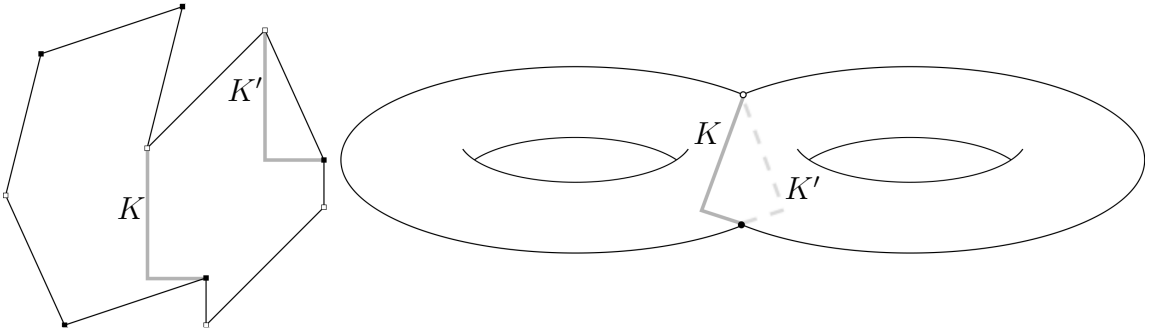


Figure 4.3: A parallel pair of non-singular L-cuts. In the picture of M on the right, K' appears behind M .

Extending McMullen’s Theorem

In McMullen’s theorem, the saddle connections have directions which cannot be prescribed, and so are generally neither horizontal nor vertical. The following theorem

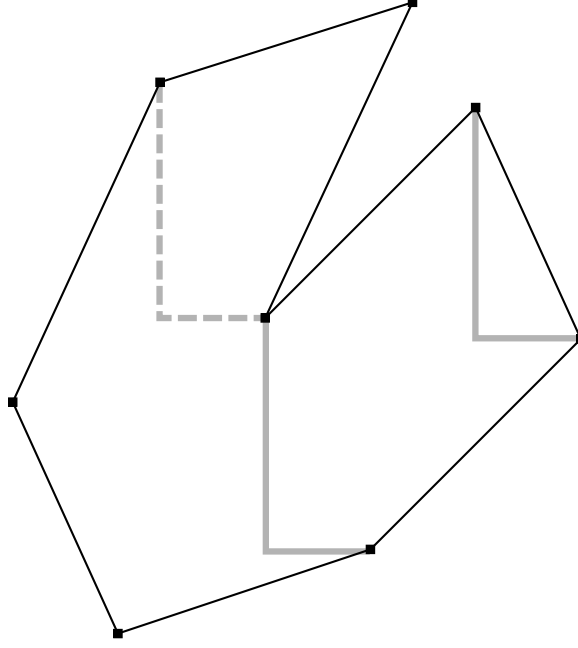


Figure 4.4: An example showing the possibility of a third L-cut (here dashed) with the same shape and segment lengths as the other two, but which does not satisfy the null-homologous condition when paired with either of them. Note that the pair which is null-homologous would survive a slight deformation of the polygon (while maintaining all identifications), whereas the third L-cut is an artifact which disappears under perturbation.

is the main result of this chapter.

Theorem 4.7 (Existence of L-cuts). *Suppose M is a genus two translation surface with vertical flow which is minimal. Then there exists a parallel pair of L-cuts (K, K') in M . Moreover, if M has no vertical and no horizontal saddle connections then the L-cuts K and K' are non-singular.*

The strategy to prove Theorem 4.7 will be to find a rectangle R on M , with horizontal and vertical sides and singular points lying at two opposite vertices, such that $R \neq \eta(R)$. Then, we take for K two adjacent sides of R that jointly connect

the singularities. An L-cut K' parallel to K is then obtained from two sides of $\eta(R)$. Note that $\eta(K')$ joins with K to form the boundary of R , and $\eta(K)$ joins with K' to form the boundary of $\eta(R)$ (see Figure 4.5).

In practice, a candidate rectangle can be easily found in the plane containing a polygon representation $\mathcal{P}$ of M . The crux is in ensuring that the interior of the rectangle is free of singularities, and so defines a rectangle in M . The proof will proceed by considering several candidate rectangles, and identifying in every case one which is indeed a rectangle in M . Additionally, note that if the boundary of R contains one or more singular points (apart from the two which appear at the endpoints of K), then one or both of K and K' will be singular. When M has no vertical and no horizontal saddle connections, non-singular L-cuts can be found. When M has vertical or horizontal saddle connections, it is possible for L-cuts will be singular, and it will be seen that this is commonly the case.

Splitting along L-Cuts

As a consequence of Theorem 3.7, we see that a parallel pair of L-cuts on M can be used as McMullen used a saddle connection J and its image $\eta(J)$ under hyperelliptic involution to write M as a connected sum of tori. Formally stated:

Theorem 4.8 (L-cut Splitting). *Given a genus 2 translation surface M with a parallel pair of L-cuts (K, K') as in Theorem 4.7, M splits along $K \cup K'$ as a connected sum of two flat tori.*

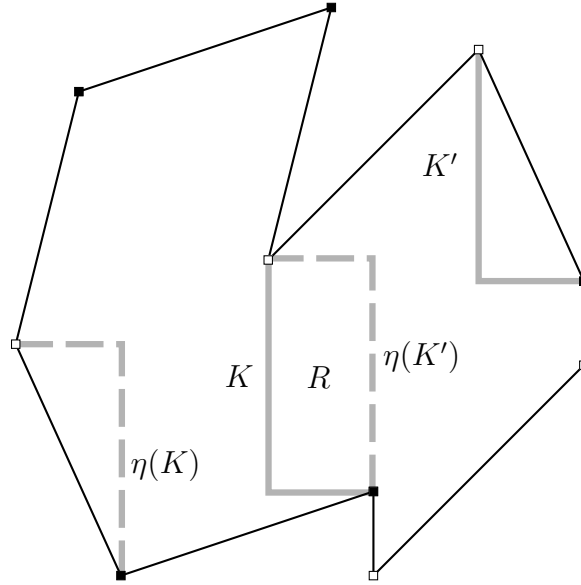


Figure 4.5: A parallel pair of non-singular L-cuts K and K' , with K and $\eta(K')$ bounding a rectangle R in M .

In this section, we describe the precise means of splitting M along $K \cup K'$ as asserted by Theorem 4.8; this is illustrated in Figure 4.6. For a polygon $\mathcal{P}$ representing M and parallel L-cuts K and K' on M , let $\gamma, \gamma' : [a, b] \rightarrow \mathcal{P}$ be parametrizations of K and K' . Then $\mathcal{P} \setminus (K \cup K')$ is a disjoint union of open (possibly disconnected) sets $\mathcal{A}, \mathcal{B} \subset \mathcal{P}$, with $\mathcal{P}_{\mathcal{A}} := \mathcal{A} \cup K \cup K'$ and $\mathcal{P}_{\mathcal{B}} := \mathcal{B} \cup K \cup K'$, such that γ and γ' induce maps $\gamma_{\mathcal{A}}, \gamma'_{\mathcal{A}} : [a, b] \rightarrow \mathcal{P}_{\mathcal{A}}$ and $\gamma_{\mathcal{B}}, \gamma'_{\mathcal{B}} : [a, b] \rightarrow \mathcal{P}_{\mathcal{B}}$. For $*$ representing $\mathcal{A}$ and $\mathcal{B}$, we require that $\mathcal{P}_*$, upon performing boundary identifications induced by the identifications of $\mathcal{P}$ restricted to $\mathcal{P}_*$, and by identifying $\gamma_*(t)$ with $\gamma'_*(t)$ for all $t \in [a, b]$, becomes an orientable translation surface M_* of genus one, i.e., M_* is homeomorphic to $\mathbb{T}^2$.

In the above, the word “induced” refers to the process of transitive closure of the

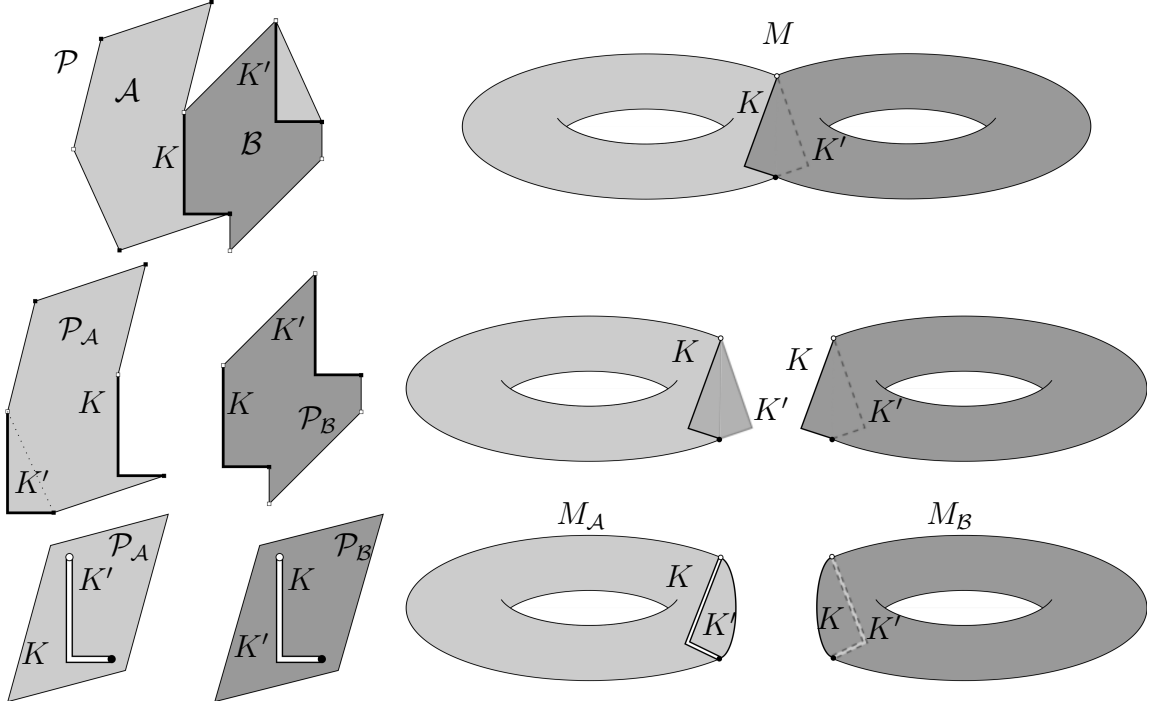


Figure 4.6: A split via a parallel pair of (non-singular) L-cuts. As in the case of a split via straight saddle connections, $\mathcal{P}_A$ and $\mathcal{P}_B$ can each be seen as a parallelograms with a slit.

relation used to identify the boundary points of $\mathcal{P}_A$ and $\mathcal{P}_B$. This is due to the fact that if K or K' is singular, one (or possibly both) of the subpolygons $\mathcal{P}_A$ and $\mathcal{P}_B$ may feature an antenna (as described on page 10). This antenna will carry important identification information for reconstructing $\mathcal{P}$ from $\mathcal{P}_A$ and $\mathcal{P}_B$.

Let $\mathcal{P}_*$ represent such a subpolygon with antenna and M_* be its associated genus-one surface (so $*$ stands for $\mathcal{A}$ or $\mathcal{B}$). Some of the identification done within this $\mathcal{P}_*$ to construct M_* will identify more than two points. Specifically, to get M_* from $\mathcal{P}_*$ it is not sufficient to identify only the pairs of points $(\gamma_*(t), \gamma'_*(t))$, but one has to take the quotient of $\mathcal{P}_*$ by the transitive closure of the relation consisting of such pairs.

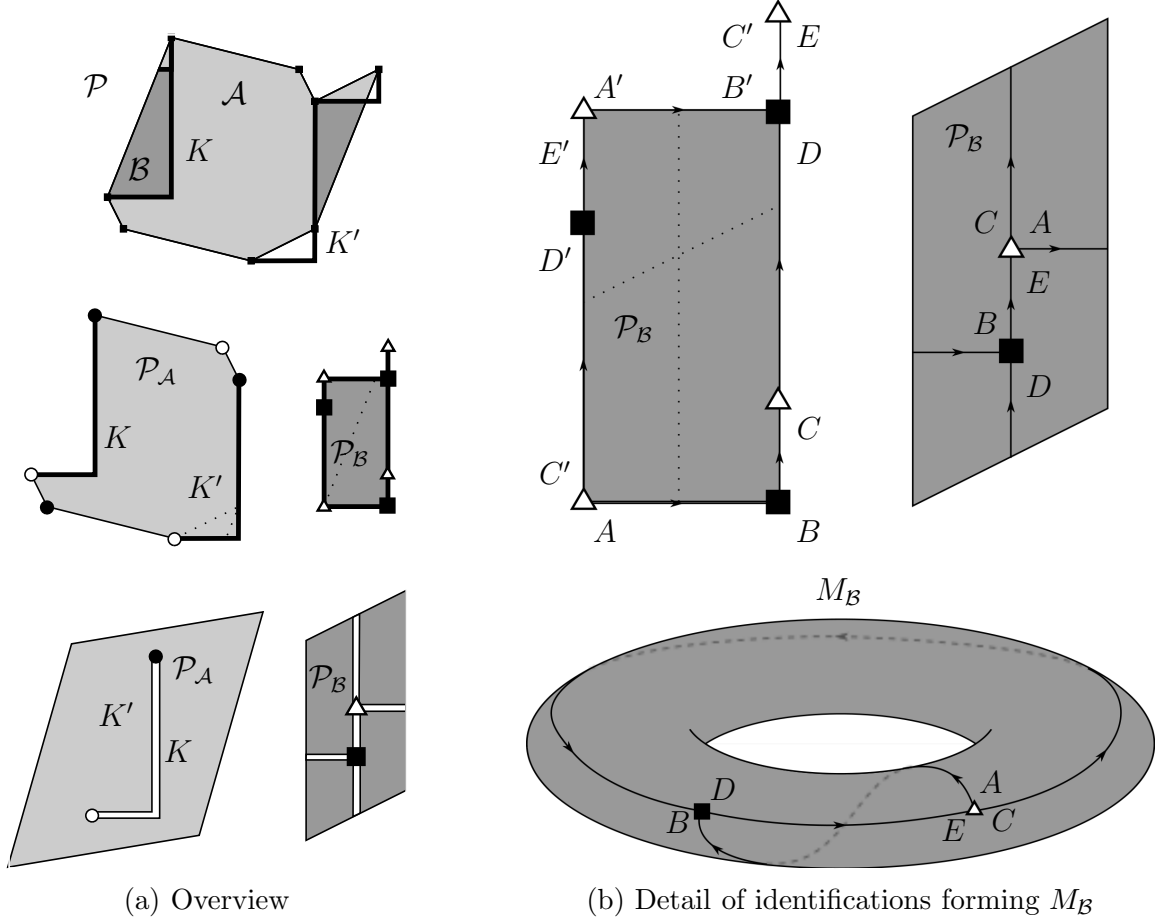


Figure 4.7: Split via a parallel pair of singular L-cuts; note that the antenna of $\mathcal{P}_B$ represents a portion of the L-cuts which is passed twice by the parametrization (between points B and C and again between D and E).

An example of this, for $M \in \mathcal{H}(2)$, is seen in Figure 4.7. In this case, the formation of M_A is straightforward, but the identifications of K and K' needed to form M_B are not immediately clear, because the L-cuts overlap in $\mathcal{P}_B$. Let A be the point at the beginning of K , symbolized in Figure 4.7 by a triangle. This same point also appears at C' on K' and thus also E on K (which is K 's ending point, so K is really a loop in M_B). If K and K' are to be identified, this means this point also appears at C

and at A' (which is also E'). The turning point of K , labeled B and symbolized by a square, gets identified with the point B' on K' , which is also D on K and therefore D' on K' . Note that the antenna of $\mathcal{P}_B$ causes the segment of K between B and C to be identified with the segment of K' between D' and E' , so this segment is traversed twice by γ and γ' ; this is, specifically, where the transitive closure is necessary to produce M_B .

Theorem 4.8 is demonstrated by explicitly exhibiting the splitting for each (K, K') constructed in the course of the proof of Theorem 4.7 which occupies most of the rest of this chapter, and requires some preliminary groundwork.

Preliminaries for the Proof of Theorem 4.7

At this point it is convenient to introduce the following notation:

- Let $\mathcal{H}^\circ(2)$ be the set of those $M \in \mathcal{H}(2)$ which have no vertical saddle connections.
- Let $\mathcal{H}^*(2)$ be the set of those $M \in \mathcal{H}(2)$ which have at least one vertical saddle connection. Further partition $\mathcal{H}^*(2)$ into $\mathcal{H}_{\text{VM}}^*(2)$ and $\mathcal{H}_{\text{NVM}}^*(2)$, which consist of surfaces on which the vertical flow is respectively minimal and not minimal.

Note that vertical flow is necessarily minimal for all $M \in \mathcal{H}^\circ(2)$. Similarly, define $\mathcal{H}^\circ(1, 1)$, $\mathcal{H}^*(1, 1)$, $\mathcal{H}_{\text{VM}}^*(1, 1)$, and $\mathcal{H}_{\text{NVM}}^*(1, 1)$.

As minimality of vertical flow is a condition of Theorem 4.7, this section will focus on only those surfaces belonging to $\mathcal{H}^\circ(2) \cup \mathcal{H}_{\text{VM}}^*(2) = \mathcal{H}(2) \setminus \mathcal{H}_{\text{NVM}}^*(2)$ or

$\mathcal{H}^\circ(1, 1) \cup \mathcal{H}_{\text{VM}}^*(1, 1) = \mathcal{H}(1, 1) \setminus \mathcal{H}_{\text{NVM}}^*(1, 1)$. Surfaces belonging to one of $\mathcal{H}_{\text{NVM}}^*(2)$ and $\mathcal{H}_{\text{NVM}}^*(1, 1)$ will be discussed in the next section.

The proof of Theorem 4.7 presented here will rest on the description of *zippered rectangles*, due to Veech; see [3]. We only briefly survey the tools and concepts we shall use; a more detailed exposition of this material can be found in [27] and [28]. Note, however, that Veech only handles generic surfaces, i.e., those without vertical saddle connections, so we will have to adapt his construction to serve these cases.

Interval Exchange Map

Suppose M is a translation surface of genus two with a one form ω on M prescribing vertical and horizontal directions. Let $I \subset M$ be a horizontal segment whose left endpoint is a singular point of M , and consider the vertical flow² on M , i.e. the flow with unit speed in a fixed direction perpendicular to I .

We would like to use I to define an interval exchange transformation $T : I \rightarrow I$. To ensure T has certain desirable properties, we will lay out three conditions labeled $(\star)$, $(\star\star)$, and $(\star\star\star)$ which I shall be chosen to satisfy.

First, in order to use T to study M , it should be the case that all vertical orbits in M hit I . Precisely, we require:

Any point in M not in the outgoing vertical of a singularity hits I under vertical flow in some negative time. ($\star$)

Clearly any I satisfies $(\star)$ in cases where the vertical flow on M is minimal. When

²Of course, this flow is not well defined (even for short times) at singularities, where there is more than one choice of outgoing vertical, and is only well defined for a finite time for points on the incoming verticals of singularities. On all other points the flow is defined for all time.

it is not, there may or may not be a suitable I . For the rest of this section, we assume M has vertical flow which is minimal. M with non-minimal vertical flow will be handled separately (Theorem 4.17 on page 76).

The general theory presented here, as seen in [27] and [28], applies to $M \in \mathcal{H}^\circ(2)$ and $M \in \mathcal{H}^\circ(1,1)$. Along the way, we will adapt it to include $M \in \mathcal{H}_{\text{VM}}^*(2)$ and $M \in \mathcal{H}_{\text{VM}}^*(1,1)$.

Consider $M \in \mathcal{H}^\circ(2)$ first. Each point on I will either flow into a singularity or return to a unique point on I . As the sole singular point z_0 has a neighborhood with angle 6π (and there are no vertical saddle connections), the image of I under the vertical flow will hit z_0 at exactly three different points p_1, p_2, p_3 . The low on either side of each of these points is separated, resulting in the flow from I being cut into four “ribbons.” Since the vertical flow is minimal on $M \in \mathcal{H}^\circ(2)$, each of these ribbons eventually returns to I . See Figure 4.8.

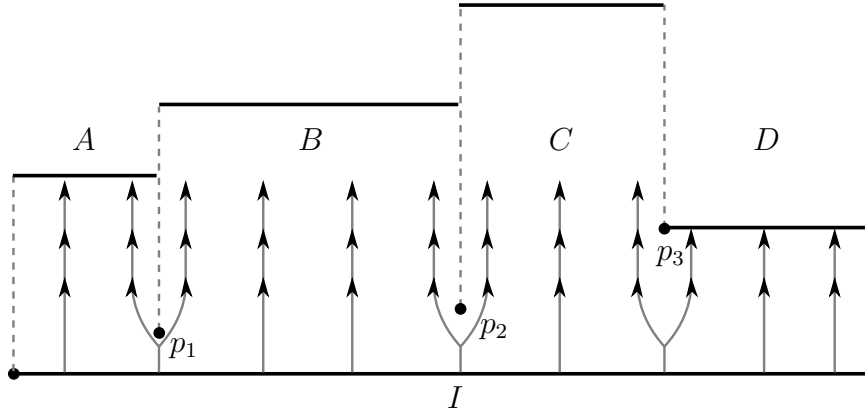


Figure 4.8: A patch of $M \in \mathcal{H}^\circ(2)$, with the vertical flow separating I into 4 segments. The angle about the singularity is 6π .

To ensure the four ribbons return to I without being cut again, we impose the following condition on I :

The right endpoint of I lies on an outgoing or incoming vertical of a singularity. Further, the segment of this vertical between the endpoint of I and the singularity does not intersect I . (★★)

Without this, there is one more cut and there are five ribbons. For I satisfying (★★), the first return map $T : I \rightarrow I$ is an interval exchange map on four segments. Label these four segments A, B, C , and D , and their respective (positive) lengths $\lambda_A, \lambda_B, \lambda_C$, and λ_D .

Before proceeding, we note that the situation here is very similar to what we encountered in the proof of Theorem 3.3 in Chapter 3, but there we used as a cross section a closed geodesic E in the place of a segment I . As a result we will arrive with a representation of M via polygons, not polybands. In fact, this is the standard way to represent translation surfaces in the literature.

For $M \in \mathcal{H}_{\text{VM}}^*(2)$, the existence of vertical saddle connections means some singular points may not be hit by the vertical flow from I at all, which happens when a saddle connection lies entirely above I (meaning I does not cross the saddle connection). Indeed, exactly one ribbon is lost for each vertical saddle connection which is not crossed by I . (This will be the object of the *Keane condition*, Definition 4.11.) See Figure 4.9.

So, there may be fewer than four ribbons (though it is still possible to have exactly four). To mark the existence and position of each vertical saddle connection that does

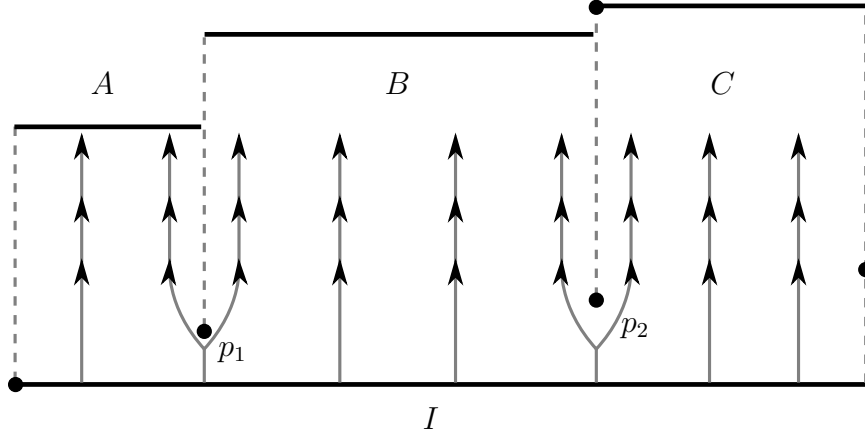


Figure 4.9: A patch of $M \in \mathcal{H}_{\text{VM}}^*(2)$, with the vertical flow separating I into fewer than 4 segments. Here, there is one vertical saddle connection so there are 3 segments.

not cross I , we still assign to it one of the names A, B, C, D , with its corresponding subsegment of I having a length of zero. Using this convention, we will still be able to define T as being an interval exchange on four segments.

For $M \in \mathcal{H}^\circ(1, 1)$, the vertical flow from a horizontal segment I satisfying $(\star\star)$ encounters the two 4π singularities z_0 and z_1 at two points each, splitting I into five ribbons. In this case, the induced map $T : I \rightarrow I$ is an interval exchange on five segments, labeled A, B, C, D , and E , with respective (positive) lengths $\lambda_A, \lambda_B, \lambda_C, \lambda_D$, and λ_E . For $M \in \mathcal{H}_{\text{VM}}^*(1, 1)$, we again allow these segments to have length zero, and it can be arranged that T is as an exchange on five segments.

For M in $\mathcal{H}_{\text{VM}}^*(2)$ or $\mathcal{H}_{\text{VM}}^*(1, 1)$, T can be defined in the following manner. Consider the flow which is in a direction perturbed from vertical by a small angle θ , for which there are no saddle connections. (Almost every θ satisfies this.) The outgoing/incoming vertical segment which is the object of $(\star\star)$ will perturb slightly and we

adjust I to I_θ accordingly to keep $(\star\star)$ satisfied (for the perturbed vertical flow). Note that when the right endpoint of I is sitting on a vertical saddle connection, there will be a choice of which of the two endpoints of this vertical saddle connection to track with I_θ .

Let $T_\theta : I_\theta \rightarrow I_\theta$ be the first return map to I_θ under the perturbed vertical flow. Now, since the perturbed vertical flow has no saddle connections (and thus the corresponding surface is in $\mathcal{H}^\circ(2) \cup \mathcal{H}^\circ(1, 1)$), Each T_θ is a proper interval exchange on four or five segments of nonzero length. It is not hard to see that, with I_θ perturbed as described above, T_θ varies continuously³ in θ for each of the cases $\theta > 0$ and $\theta < 0$ (though not as θ changes sign).

Thus, we can take two limits $T_+ := \lim_{\theta \rightarrow 0^+} T_\theta$ and $T_- := \lim_{\theta \rightarrow 0^-} T_\theta$. $T_{+/-}$ are also interval exchanges, but with some segments of length zero. So, for $M \in \mathcal{H}_{\text{VM}}^*(2) \cup \mathcal{H}_{\text{VM}}^*(1, 1)$, given I satisfying $(\star)$, $(\star\star)$, T can be defined one of two ways if I does not end on a vertical saddle connection, and one of four ways if it does (doubled due to the two choices of which point I_θ should track for each of T_+ and T_-).

Definition 4.9. *The combinatorial datum of the interval exchange map $T : I \rightarrow I$ for $M \in \mathcal{H}^\circ(2) \cup \mathcal{H}_{\text{VM}}^*(2)$ consists of two arrangements of the set of four labels $\{A, B, C, D\}$, representing the order in which the segments appear in I before and after applying T . For $M \in \mathcal{H}^\circ(1, 1) \cup \mathcal{H}_{\text{VM}}^*(1, 1)$ they are arrangements of the five labels $\{A, B, C, D, E\}$.*

³say, in the sense of Hausdorff metric between the closures of the graphs of T_θ

We will record the combinatorial datum of T by listing the order in which the segments appear before applying T above the order in which they appear after applying T , e.g., $\begin{bmatrix} ABCD \\ DCBA \end{bmatrix}$ or $\begin{bmatrix} ABCDE \\ EDCBA \end{bmatrix}$. Note that for $M \in \mathcal{H}_{\text{VM}}^*(2) \cup \mathcal{H}_{\text{VM}}^*(1, 1)$, the two (or four) possible maps T may have different combinatorial data.

Again, the nature of the combinatorial datum we use here is different from the discrete datum of a polyband presented in Chapter 3, it being an interval exchange on a segment I rather than on a closed geodesic E . In particular, recall that for a polyband, the permutation representing the combinatorial type depended on the choice of starting edges, so the discrete datum was an equivalence class of permutations called a reduced permutation.

Definition 4.10. *The combinatorial datum of T is called admissible if the associated permutation of segments is irreducible, equivalently if there is no segment $I' \subsetneq I$ of positive length which contains the left endpoint of I , where $I \setminus I'$ also has positive length, such that $T(I') = I'$.*

Admissibility disallows combinatorial data such as $\begin{bmatrix} ABCD \\ BCAD \end{bmatrix}$ and $\begin{bmatrix} ABCDE \\ CBAED \end{bmatrix}$; in such cases, the vertical flow on M is necessarily not minimal, being split into acting on distinct invariant subsurfaces. Thus, any T representing $M \in \mathcal{H}^\circ(2) \cup \mathcal{H}_{\text{VM}}^*(2) \cup \mathcal{H}^\circ(1, 1) \cup \mathcal{H}_{\text{VM}}^*(1, 1)$ will necessarily have admissible combinatorial datum.

Note, however, that for $M \in \mathcal{H}_{\text{VM}}^*(2) \cup \mathcal{H}_{\text{VM}}^*(1, 1)$, for certain configurations of the vertical saddle connections, some admissible combinatorial data cannot occur. For example, suppose $M \in \mathcal{H}_{\text{VM}}^*(2)$ has a single vertical saddle connection which has been

designated C . Then M does not admit an I such that the induced interval exchange map $T : I \rightarrow I$ has the admissible combinatorial datum $\begin{bmatrix} ADBC \\ DCAB \end{bmatrix}$, since this would result in points of I corresponding to the segment B returning to themselves under T , a violation of minimality.

For M with vertical saddle connections, it will be desirable to identify the way in which those vertical saddle connections interact with T . The following condition will prove important for studying $M \in \mathcal{H}_{\text{VM}}^*(2) \cup \mathcal{H}_{\text{VM}}^*(1, 1)$:

Definition 4.11. *$T : I \rightarrow I$ is said to satisfy the Keane condition if no (positive or negative) power of T maps an endpoint of one of the subsegments of the interval exchange to an endpoint of another subsegment (apart from those which necessarily map to the endpoints of I).*

Stated another way, T satisfying the Keane condition is such that M has no vertical saddle connection which crosses the interior of I (though M may have other vertical saddle connections). Thus our third condition on I is:

I does not cross a vertical saddle connection (though its right endpoint may be on one). (***)

Note that any I for $M \in \mathcal{H}^\circ(2) \cup \mathcal{H}^\circ(1, 1)$ satisfies (***) by default and thus the map T associated to I automatically satisfies the Keane condition. However, T for $M \in \mathcal{H}_{\text{VM}}^*(2) \cup \mathcal{H}_{\text{VM}}^*(1, 1)$ satisfies the Keane condition if and only if its domain I satisfies (***) .

Observe that any T satisfying the Keane condition necessarily has admissible

combinatorial datum, though the converse is not true. (Indeed, Keane shows in [29] that when M has a T satisfying the Keane condition, vertical flow on M must be minimal. [27] and [28] also have proofs of this fact.)

Zippered Rectangles and Polygons

Suppose $M \in \mathcal{H}^\circ(2)$. The four segments A, B, C , and D , which together make up I , can be seen as the bases of four rectangles $R_1, \dots, R_4$. As $T : I \rightarrow I$ is the first return map to I , the top of each rectangle is also a segment of I . Thus, the same four rectangles can be translated (according to the combinatorial datum of T) so that their tops combine to form I . The same situation arises for $M \in \mathcal{H}^\circ(1, 1)$, except that now there are five rectangles. For $M \in \mathcal{H}_{\text{VM}}^*(2)$, there may be fewer than four rectangles, and for $M \in \mathcal{H}_{\text{VM}}^*(1, 1)$, there may be fewer than five rectangles. (Indeed, this will be the case precisely when I satisfies $(\star \star \star)$.) Again, these cases are best understood by taking limits of the $\mathcal{H}^\circ(\dots)$ cases by perturbing the vertical by a small angle θ .

Pairs of adjacent rectangles will have a portion of their common side identified, while a slit forms at each appearance of a singular point, separating part of this common side into distinct segments which are not identified with each other (though those segments each still get identified with some other segment). The union of these rectangles⁴, with the proper identifications (see [3] for the formal definition), forms a

⁴At first glance, it may seem that this is a tower, defined in Chapter 3 on page 25. However, once again the difference is that the bases of these rectangles form the segment I , rather than the closed geodesic E . In particular, the left edge of the leftmost rectangle and right edge of the rightmost rectangle is not automatically identified, as they were in a tower.

surface which is isomorphic (i.e., isometric via an isometry preserving horizontal and vertical) to M . We note that $(\star\star\star)$ implies that each vertical saddle connection (if any) will appear at the side of a rectangle.

Definition 4.12. A zippered rectangle associated to the pair (M, T) consists of two such unions of rectangles: one whose base forms I , and one whose top forms I .

A zippered rectangle for $M \in \mathcal{H}^\circ(2)$ with T having combinatorial datum $\begin{bmatrix} ABCD \\ DCBA \end{bmatrix}$ is seen in Figure 4.10. A zippered rectangle for $M \in \mathcal{H}_{\text{VM}}^*(2)$ is seen in Figure 4.11; recall that T may exhibit up to four different combinatorial data. Here there are 3 possibilities: $\begin{bmatrix} ABCD \\ DCBA \end{bmatrix}$, $\begin{bmatrix} ADBC \\ DCBA \end{bmatrix}$, and $\begin{bmatrix} ABCD \\ DBAC \end{bmatrix}$.

Conversely, for $M \in \mathcal{H}^\circ(2)$, the combinatorial datum of an interval exchange map $T : I \rightarrow I$ can be recovered by examining a given zippered rectangle. For convenience, we will speak of the zippered rectangle as having the combinatorial datum of its associated interval exchange map T . Again, for $M \in \mathcal{H}_{\text{VM}}^*(2)$, there can be up to four choices of combinatorial datum depending on which T is used.

Now, for $M \in \mathcal{H}^\circ(2)$, two (non-closed) polygonal chains $\mathcal{C}^+$ and $\mathcal{C}^-$ can be defined on a zippered rectangle representation of M , both starting at the appearance of z_0 at the left endpoint of I , with $\mathcal{C}^+$ connecting adjacent appearances of z_0 above I , and $\mathcal{C}^-$ connecting adjacent appearances of z_0 below I . Including one more segment joining the endpoints of $\mathcal{C}^+$ and $\mathcal{C}^-$ (the rightmost appearances of z_0 above and below I , respectively) results in an octagon $\mathcal{O}$, seen in Figure 4.12. $\mathcal{O}$ has edges which come in four parallel pairs; identifying corresponding edges yields M . Give each edge of $\mathcal{O}$

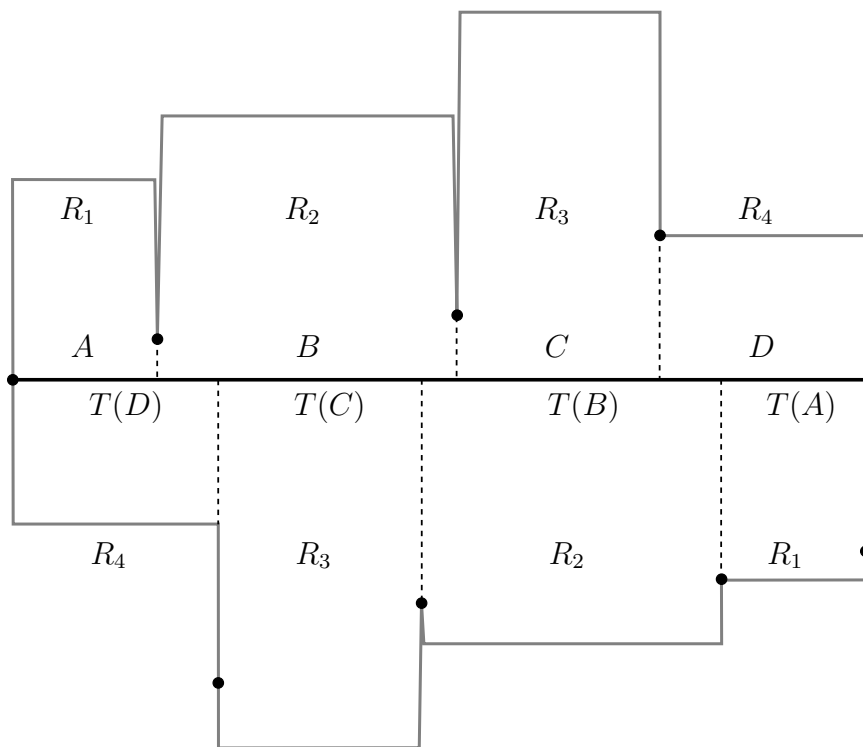


Figure 4.10: A zippered rectangle for $M \in \mathcal{H}^\circ(2)$ with T having combinatorial datum $\begin{bmatrix} ABCD \\ DCBA \end{bmatrix}$.

the label $A^{+/-}$, $B^{+/-}$, $C^{+/-}$, or $D^{+/-}$, according to the rectangle it came from, with $+$ indicating edges of $\mathcal{C}^+$ and $-$ indicating edges of $\mathcal{C}^-$.

For $M \in \mathcal{H}^\circ(1,1)$, as there are now five rectangles, the polygon resulting from this process (now with $\mathcal{C}^+$ and $\mathcal{C}^-$ having both z_0 and z_1 as vertices) is a decagon $\mathcal{D}$, whose sides come in five parallel pairs. As before, label each edge of $\mathcal{D}$ as $A^{+/-}$, $B^{+/-}$, $C^{+/-}$, $D^{+/-}$, or $E^{+/-}$. We will declare the combinatorial datum of an octagon or decagon representation of M to be that of its underlying zippered rectangle (and therefore also its associated interval exchange map T).

When M is in $\mathcal{H}_{\text{VM}}^*(2)$ or $\mathcal{H}_{\text{VM}}^*(1,1)$, a similar polygon can be drawn. The primary

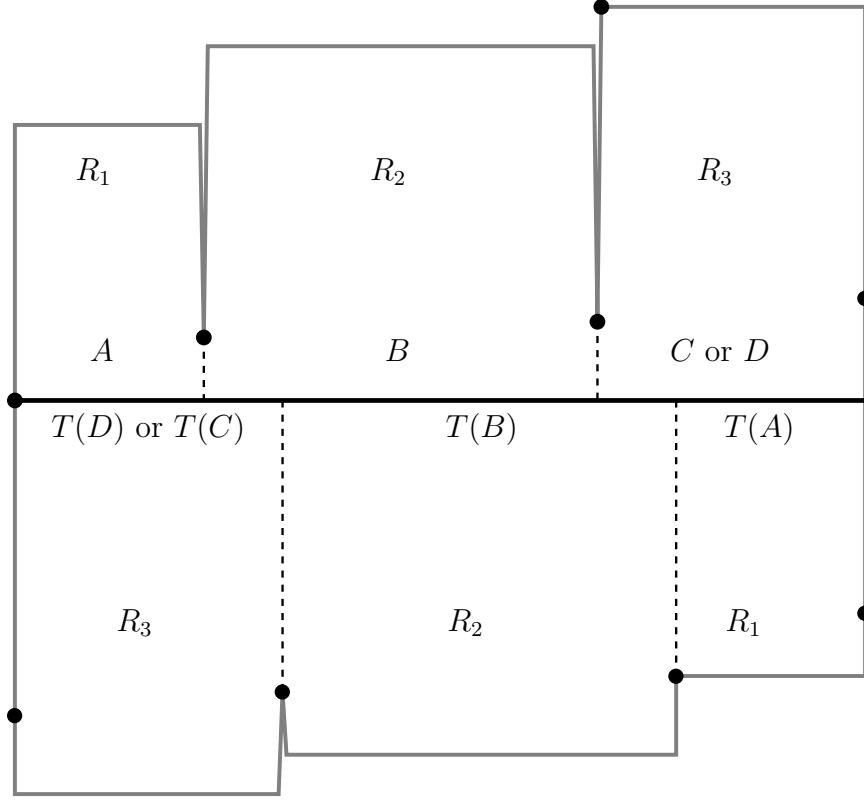


Figure 4.11: A zippered rectangle for $M \in \mathcal{H}_{\text{VM}}^*(2)$. Note that T has one of two combinatorial data $\begin{bmatrix} ABCD \\ DCBA \end{bmatrix}$ or $\begin{bmatrix} ABCD \\ DBAC \end{bmatrix}$, depending on how the vertical saddle connection is incorporated.

difference in these cases is that there is no longer a unique choice of the order in which $\mathcal{C}^+$ and $\mathcal{C}^-$ include the vertical saddle connection, so it is possible to get multiple distinct polygons from a given zippered rectangle. This choice precisely corresponds to the two (or four) choices of T described earlier, so there will be two different polygons in cases where I does not end on a vertical saddle connection and four otherwise. As previously mentioned, these polygons may (and likely will) have different combinatorial data. However, note that a given polygon has only one combinatorial datum. An example illustrating the four polygons generated from a particular zippered rectangle

is seen in Figure 4.13.

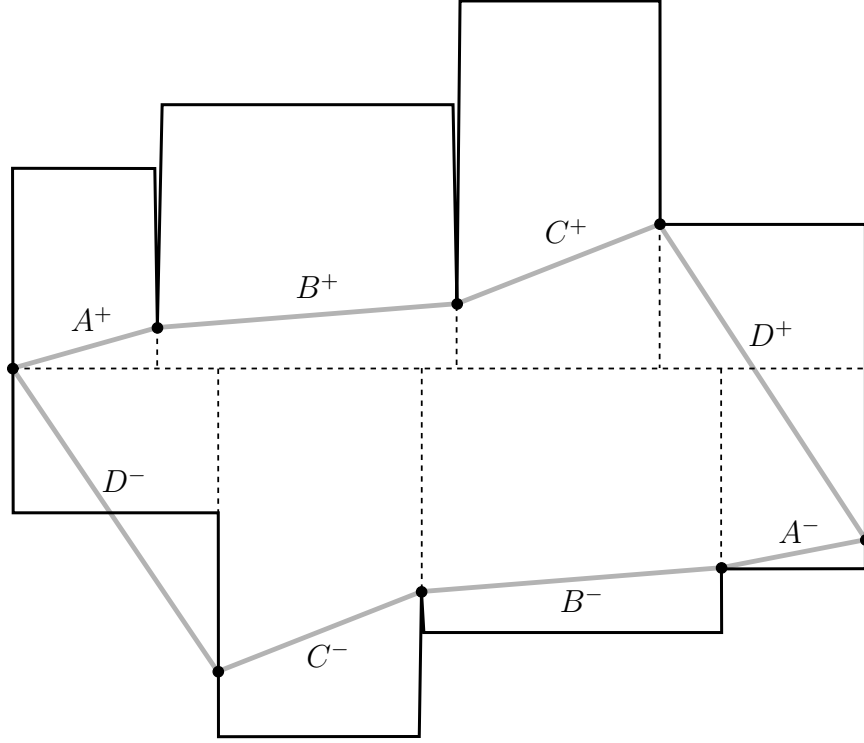
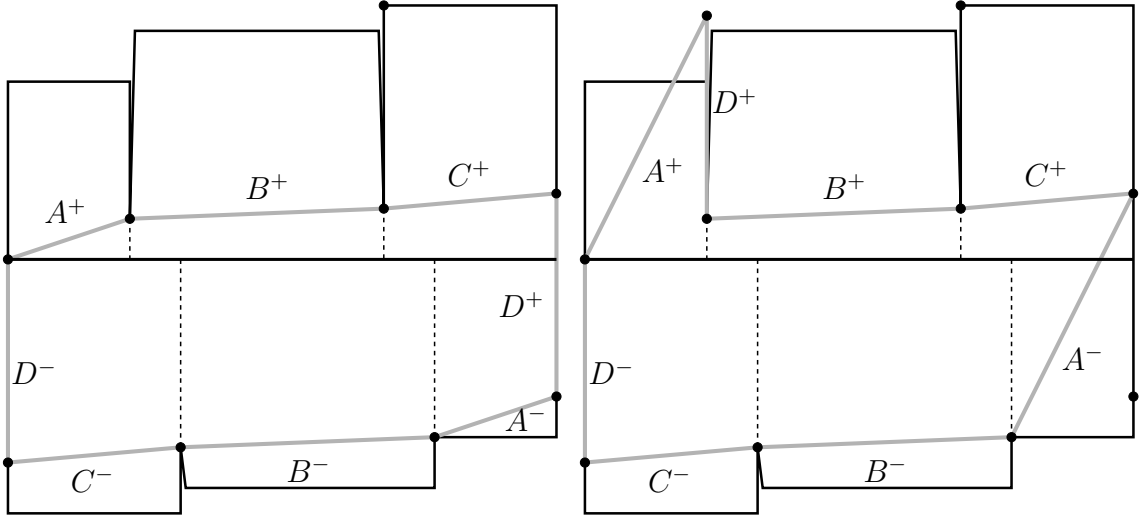
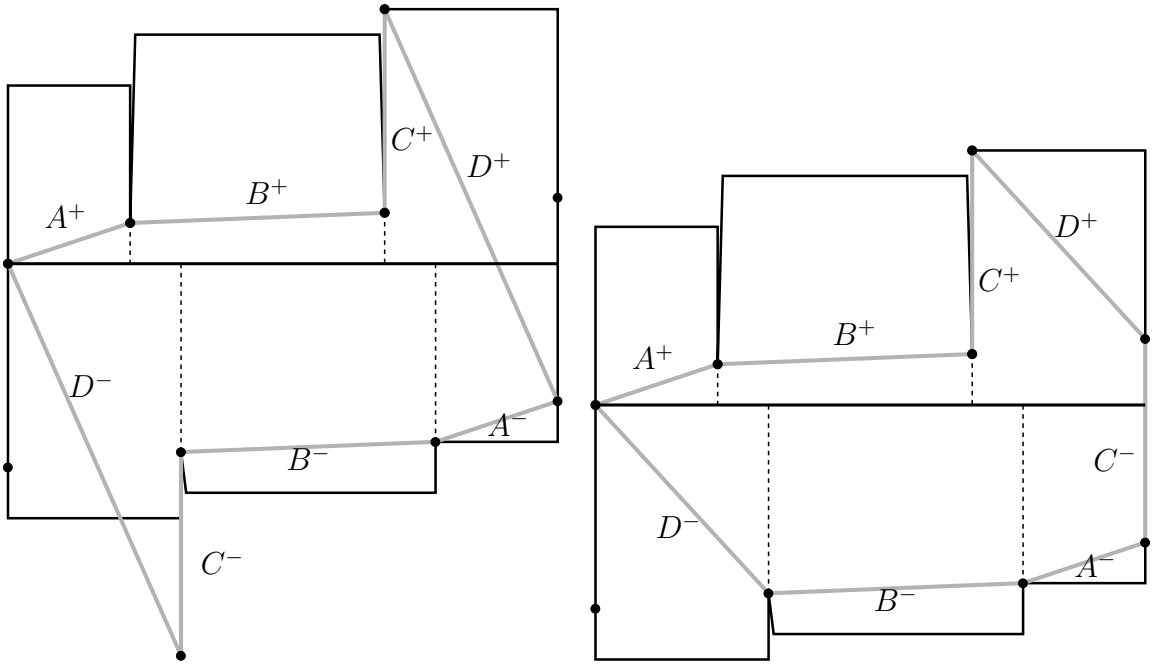


Figure 4.12: An octagon representing $M \in \mathcal{H}^o(2)$ formed from a zippered rectangle. Here the segment joining $\mathcal{C}^+$ and $\mathcal{C}^-$ is given the label D^+ .

One possible complication with attempting to construct a polygon from a zippered rectangle is that there may be a “fishtail” phenomenon, where joining the polygonal chains results in a non-simple polygon. This is illustrated in Figure 4.14. However, note in particular that this cannot happen when T has combinatorial datum $\begin{bmatrix} ABCD \\ DCBA \end{bmatrix}$ or $\begin{bmatrix} ABCDE \\ EDCBA \end{bmatrix}$ (and so T reverses the order of the segments of I), as in these cases the polygon is centrally symmetric. Because of this phenomenon, zippered rectangles are the preferred way of presenting M . However, polygons are more intuitive for our purposes, and we will take steps to avoid fishtails.



(a) $\theta < 0$, I_θ tracks the outgoing vertical. (b) $\theta < 0$, I_θ tracks the incoming vertical.
 Combinatorial datum is $\begin{bmatrix} ABCD \\ DCBA \end{bmatrix}$. Combinatorial datum is $\begin{bmatrix} ADBC \\ DCBA \end{bmatrix}$.



(c) $\theta > 0$, I_θ tracks the outgoing vertical. (d) $\theta > 0$, I_θ tracks the incoming vertical.
 Combinatorial datum is $\begin{bmatrix} ABCD \\ DCBA \end{bmatrix}$. Combinatorial datum is $\begin{bmatrix} ABCD \\ DBAC \end{bmatrix}$.

Figure 4.13: Four different octagons representing $M \in \mathcal{H}_{\text{VM}}^*(2)$ formed from a single zippered rectangle, not all with the same combinatorial datum. Also notice that the vertical saddle connection is labeled D in the top two pictures and C in the bottom two pictures.

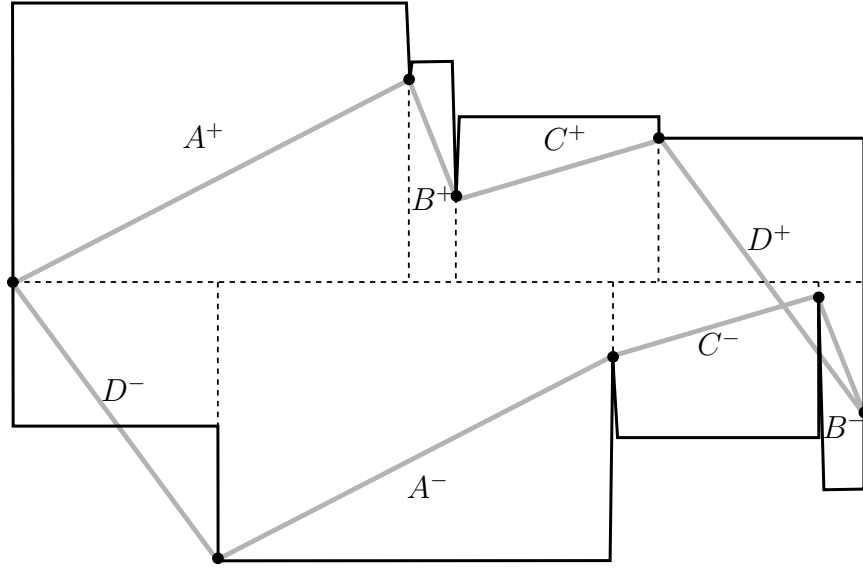


Figure 4.14: Here a zippered rectangle produces a non-simple polygon, having a “fishtail.” Here, the vertex between C^- and D^- , the last two segments of C^- , is so high that the edge D^+ formed to join C^+ and C^- must cross C^- , resulting in a non-simple octagon. As this can only happen on the right side (where the polygonal chains C^+ and C^- are joined), it is prevented in the centrally symmetric zippered rectangles of combinatorial datum $\begin{bmatrix} ABCD \\ DCBA \end{bmatrix}$ or $\begin{bmatrix} ABCDE \\ EDCBA \end{bmatrix}$.

Rauzy Operations and Rauzy Diagrams

For $M \in \mathcal{H}^\circ(2)$, if the rightmost rectangle on the bottom of a zippered rectangle is narrower than the rightmost rectangle on the top (i.e., its corresponding subsegment of I is shorter), it can be removed and reattached (with the same identification) to the other copy of the rectangle above it, with the newly orphaned portion of the top rectangle similarly relocated. This pair of translations results in a new zippered rectangle representing the same surface. (See Figure 4.15, top.) Conversely, if the rightmost top rectangle is narrower than the rightmost bottom rectangle, a similar pair of movements can be made. These reconfigurations are called *Rauzy operations*

of type 0 and type 1, respectively. (Again, their formal definition can be found in [27] and [28].)

Let X represent whichever of A , B , C , or D , (or E , for $M \in \mathcal{H}^\circ(1, 1)$) is the longer of the two rightmost segments, and Y represent the shorter of the two. Applying the appropriate Rauzy operation shortens λ_X by λ_Y , and the new zippered rectangle will typically (but not always) have a new combinatorial datum. (See Figures 4.17 and 4.18.) In the octagon or decagon representation of M , the Rauzy operations can be seen as the translation of a particular triangle. (See Figure 4.15, bottom.)

Note that if the rightmost rectangles on the top and bottom have the same length, neither Rauzy operation is defined. This cannot happen precisely when the associated interval exchange map T satisfies the Keane condition (which, again, is always true for $M \in \mathcal{H}^\circ(2)$ and $M \in \mathcal{H}^\circ(1, 1)$).

In most situations, it is a simple matter to perform these Rauzy operations on polygons representing $M \in \mathcal{H}_{\text{VM}}^*(2)$ or $M \in \mathcal{H}_{\text{VM}}^*(1, 1)$, provided I has been chosen so that T satisfies the Keane condition. Particularly, the operations behave exactly the same way when both of the rightmost segments in the combinatorial datum of the polygon have nonzero length. Additionally, when precisely one of these segments has nonzero length, the Rauzy operation still moves a triangle in a clear and unambiguous way. The zero-length segment is, of course, shorter than the one with nonzero length, so the triangle (with one vertical side) is moved as in the case where both segments have nonzero length. Note that in this situation $\lambda_Y = 0$ so λ_X does not actually get

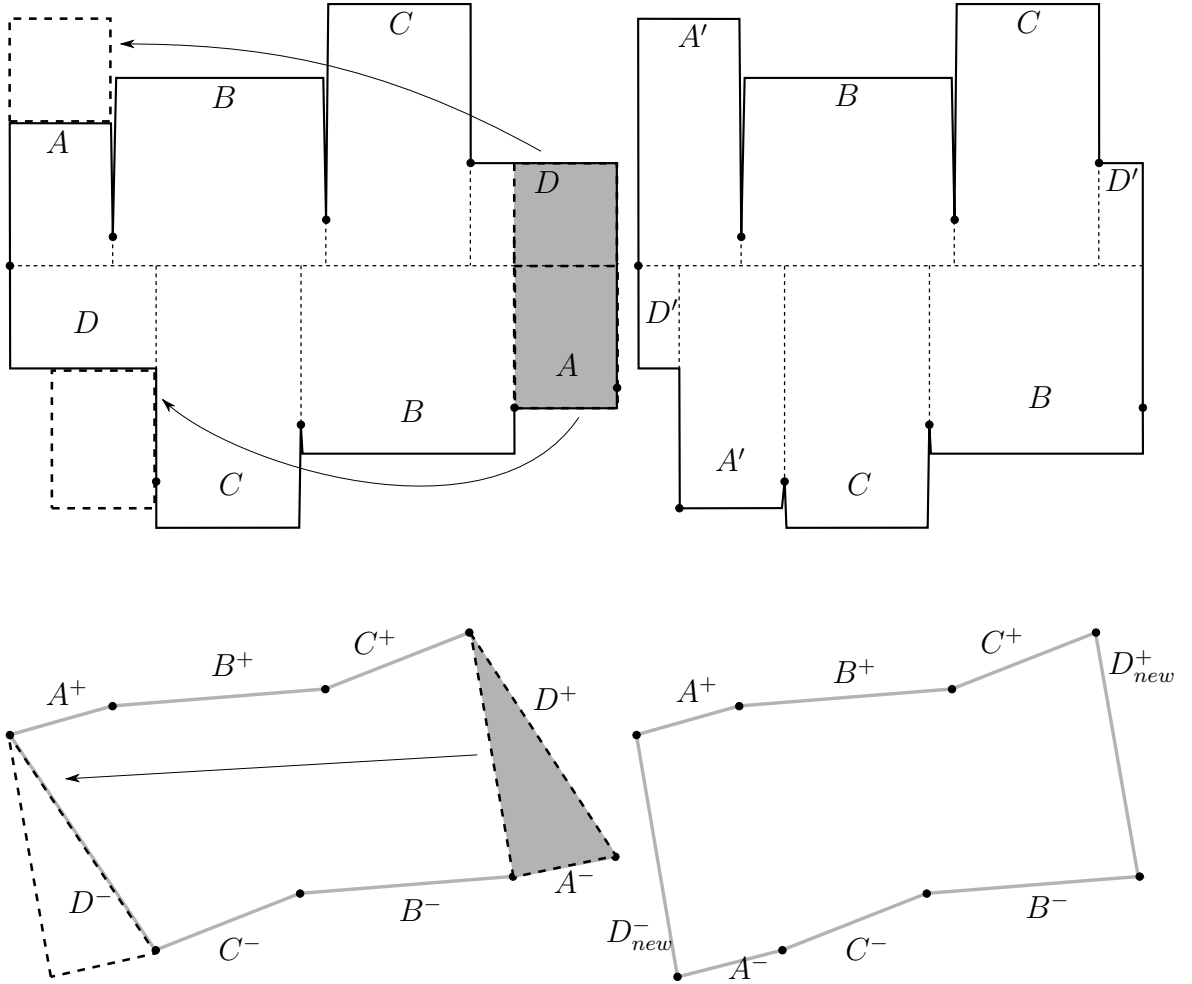


Figure 4.15: Performing a Rauzy operation “0” on a zippered rectangle, above, and on its associated octagon, below, both representing $M \in \mathcal{H}^\circ(2)$. Before the operation, the combinatorial datum is $\begin{bmatrix} ABCD \\ DCBA \end{bmatrix}$; after, it has become $\begin{bmatrix} ABCD \\ DACB \end{bmatrix}$.

shorter. See Figure 4.16 for an example.

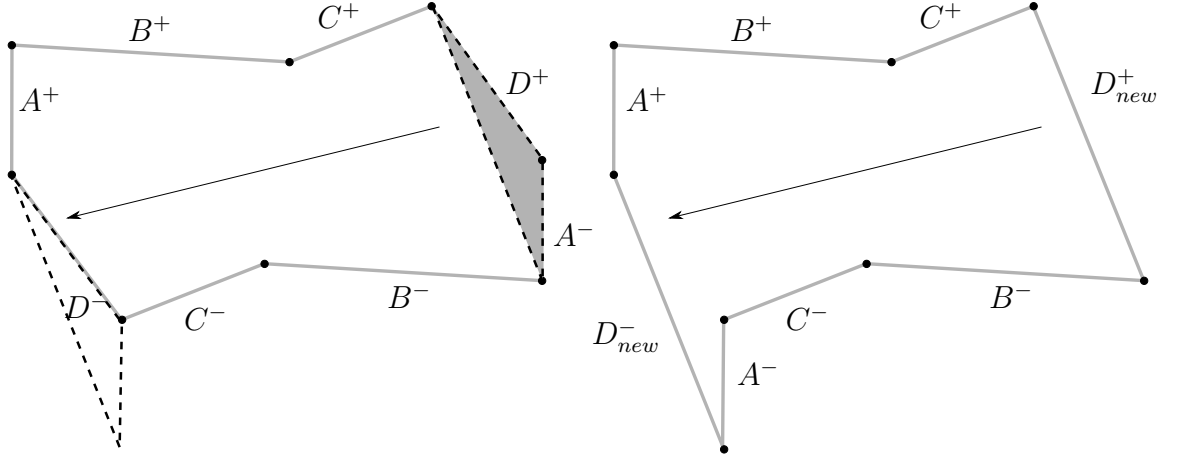


Figure 4.16: Performing a Rauzy operation “0” on an associated octagon representing $M \in \mathcal{H}_{\text{VM}}^*(2)$. Before the operation, the combinatorial datum is $\begin{bmatrix} ABCD \\ DCBA \end{bmatrix}$; after, it has become $\begin{bmatrix} ABCD \\ DACB \end{bmatrix}$.

However, when both of the rightmost segments in the combinatorial datum of the polygon have zero length, the triangle degenerates to a segment, and neither of the two segments is longer, so there is no deterministic choice of which of the two operations to perform (assuming it could even be determined how to do either of them). For future reference, designate such a polygon as *Rauzy-incompatible*. For our purposes, it will be acceptable to leave this problem unresolved, as it does not arise in our arguments.

There are seven admissible combinatorial data which can arise an interval exchange T representing vertically minimal $M \in \mathcal{H}(2)$; for vertically minimal $M \in \mathcal{H}(1,1)$ there are fifteen. Taking these as nodes of a graph, together with the Rauzy

operations acting as directed edges, we can construct the following:

Definition 4.13. A Rauzy diagram associated to a stratum of translation surfaces is a directed graph whose nodes are admissible combinatorial data for interval exchange maps T arising from translation surfaces in the stratum and whose edges have two types corresponding to the two Rauzy operations. A directed edge connecting one node to another indicates that the application of the associated Rauzy operation to a zippered rectangle (or polygon) having the combinatorial datum of the first node results in a zippered rectangle (or polygon) having the combinatorial datum of the second node.

Each node in a Rauzy diagram has exactly one of each of the following: an incoming 0 edge, outgoing 0 edge, incoming 1 edge, and outgoing 1 edge. Note that in some cases, a Rauzy operation doesn't result in a change of combinatorial datum, corresponding to looped edges in the Rauzy diagram.

One feature of the Rauzy diagram worth pointing out is the horizontal symmetry; replacing T by T^{-1} reverses the roles of the operations 0 and 1, as well as turns each combinatorial datum “upside down.” Thus, each edge of the Rauzy diagram has a corresponding edge on the other side of the diagram associated with the opposite operation, and each node of the Rauzy diagram (except the central one) has a corresponding node on the other side of the diagram with the same structure.

Figure 4.17 shows the Rauzy diagram for zippered rectangles (or polygons) representing vertically minimal surfaces in $\mathcal{H}(2)$ (combinatorial data involving four segments A, B, C , and D), and Figure 4.18 shows the Rauzy diagram for vertically minimal surfaces in $\mathcal{H}(1, 1)$ (combinatorial data involving five segments A, B, C, D , and E). In both of these cases, the Rauzy diagram has a single connected component, as the strata $\mathcal{H}(2)$ and $\mathcal{H}(1, 1)$ are connected. For higher genus, there is a Rauzy diagram for each connected component of the associated stratum. (See [3] and the discussions in [16] and [15].)

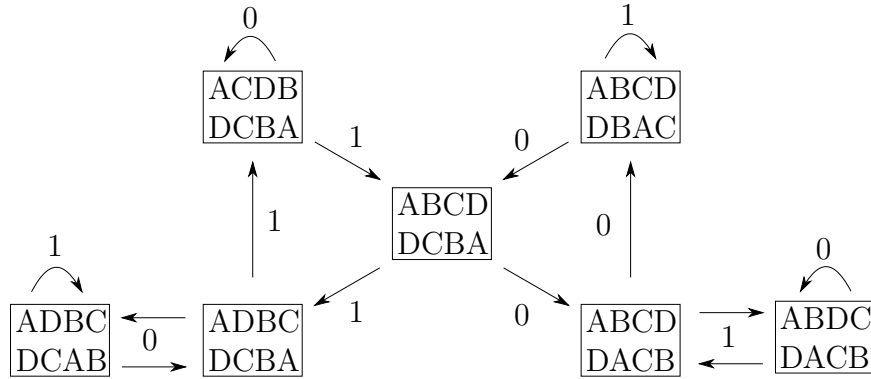


Figure 4.17: Rauzy diagram for the stratum $\mathcal{H}(2)$.

We can now take a moment explain why the existence of Rauzy-incompatible polygons, as described on page 66, where both of the rightmost edges of a polygon representing $M \in \mathcal{H}_{\text{VM}}^*(2) \cup \mathcal{H}_{\text{VM}}^*(1, 1)$ are vertical, does not affect our proof. To start, note that in each of the two Rauzy diagrams for genus two, for every node except the central one, one of the two rightmost segments cannot have length zero without violating minimality (as in the example described on page 55).

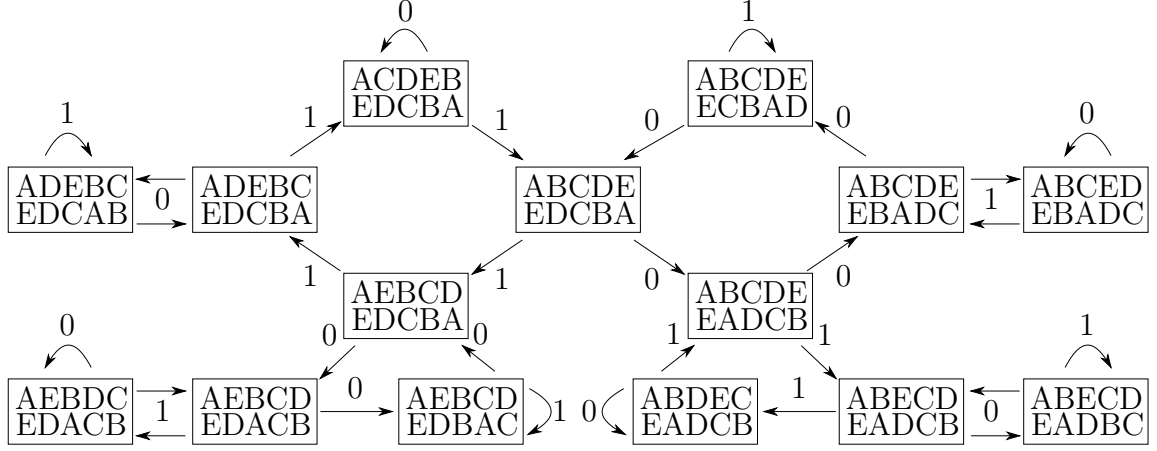


Figure 4.18: Rauzy diagram for the stratum $\mathcal{H}(1,1)$.

Thus, a Rauzy-incompatible polygon can only have the combinatorial datum of the central node. However, in the proof of Theorem 4.7, it will never be necessary to perform a Rauzy operation on such a polygon. Indeed, our primary use of Rauzy operations will be to perform a sequence of them, if necessary, to enable the supposition that the polygon we use to represent M has the combinatorial datum of the central node; this guarantees the polygon has no fishtail, as well as reduces the combinatorial complexity of the proof. In particular, there will be no call for performing a Rauzy operation on a Rauzy-incompatible polygon.

At most one of the two Rauzy operations can be performed on any given polygon. Thus, any initial polygon representation of M predetermines a sequence of Rauzy operations, which corresponds to a path in the Rauzy diagram. Provided the associated interval exchange map T satisfies the Keane condition (which, recall, always happens for $\mathcal{H}^\circ(2)$ and $\mathcal{H}^\circ(1,1)$, and can always be made to happen for $\mathcal{H}_{\text{VM}}^*(2)$ and

$\mathcal{H}_{VM}^*(1, 1)$), and the polygon is not Rauzy-incompatible, exactly one of the two operations can always be performed. Thus this path is infinite for $M \in \mathcal{H}^\circ(2) \cup \mathcal{H}^\circ(1, 1)$, and it can be arranged for $M \in \mathcal{H}_{VM}^*(2) \cup \mathcal{H}^\circ(1, 1)$ that the path is only terminated upon encountering a Rauzy-incompatible polygon.

Proposition 4.3 in [27] and Proposition 5.1 in [28] state that for $M \in \mathcal{H}^\circ(2) \cup \mathcal{H}^\circ(1, 1)$ the path in the Rauzy diagram is such that each letter appears last infinitely many times in both rows of the combinatorial datum. A direct consequence of this is that every node of the diagram is visited infinitely many times. We now adapt this argument to prove the following Lemmas 4.14 and 4.15. Note that the presence of zero-length segments dictates extra care.

Lemma 4.14. *For any initial polygon representation of $M \in \mathcal{H}^\circ(2) \cup \mathcal{H}_{VM}^*(2)$, where the associated interval exchange map T satisfies the Keane condition, the resulting path in the Rauzy diagram necessarily visits the central node $\begin{bmatrix} ABCD \\ DCBA \end{bmatrix}$.*

Proof. Suppose the path stays only on one side of the Rauzy diagram, for instance the left. Looking at the diagram (Figure 4.17), we see there are two possibilities: either the combinatorial datum is eventually always $\begin{bmatrix} ACDB \\ DCBA \end{bmatrix}$, or it never even reaches that node and cycles between $\begin{bmatrix} ADBC \\ DCAB \end{bmatrix}$ and $\begin{bmatrix} ADBC \\ DCBA \end{bmatrix}$.

The former case is only possible if $\lambda_A < \lambda_B$ for every iterate of the “0” operation which maps this node to itself. However, as this can only be the case if $\lambda_A = 0$, and vertical flow on this node is only minimal if $\lambda_A > 0$, we obtain a contradiction.

In the latter case, by assumption the “1” operation advancing $\begin{bmatrix} ADBC \\ DCBA \end{bmatrix}$ to

$\begin{bmatrix} ACDB \\ DCBA \end{bmatrix}$ is never performed, which requires that λ_A always be shorter than λ_C . Again, as λ_C is repeatedly shortened by λ_A and λ_B , this can only be the case if $\lambda_A = 0$. But, once again, minimality is violated unless $\lambda_A > 0$.

So, neither of these situations are possible, and we must eventually arrive at the central node. The argument for why the path cannot stay on the right half of the diagram is similar, with contradiction coming from the simultaneous requirement of $\lambda_D = 0$ and $\lambda_D > 0$. $\square$

Lemma 4.15. *For any initial polygon representation of $M \in \mathcal{H}^\circ(1, 1) \cup \mathcal{H}_{VM}^*(1, 1)$, where the associated interval exchange map T satisfies the Keane condition, the resulting path in the Rauzy diagram necessarily visits the central node $\begin{bmatrix} ABCDE \\ EDCBA \end{bmatrix}$.*

Proof. The argument is similar to that of Lemma 4.14. Consult Figure 4.18. First suppose the path stays on the left side. In order to avoid the central node, the path must eventually be trapped in one of three subgraphs (depending on the progress of the path through the large loop of “1” operations which leads to the central node). The path cannot stay in any one of these subgraphs unless $\lambda_A = 0$, but vertical flow is not minimal in a polygon with any of these combinatorial data unless $\lambda_A > 0$. Similarly, to stay on the right half of the diagram, λ_E must be zero, but this violates minimality. $\square$

Hyperelliptic Involution on Polygons

Although we shall not need it for the proof of Theorem 4.7, we take a moment to

discuss the nature of the hyperelliptic involution on polygon presentations of genus two M . A reader interested only in the proof may skip this section, with the understanding that η can be realized on a centrally symmetric polygon $\mathcal{P}$ which presents M by considering a simple 180° rotation of $\mathcal{P}$.

For other polygons, η on M can be realized on $\mathcal{P}$ by partitioning the polygon into centrally symmetric subpolygons $\mathcal{P}_i$ and subjecting each $\mathcal{P}_i$ to rotation by 180° . Note that the depictions here are for $M \in \mathcal{H}^\circ(2) \cup \mathcal{H}^\circ(1, 1)$, but the pictures are the same for $M \in \mathcal{H}_{\text{VM}}^*(2) \cup \mathcal{H}_{\text{VM}}^*(1, 1)$ except that some edges are vertical.

For $M \in \mathcal{H}(2)$ there are only 3 possible decompositions of the octagon $\mathcal{P}$ without fishtail into centrally symmetric subpolygons: (a) the whole octagon, (b) a hexagon and a parallelogram, or (c) three parallelograms. These are illustrated in Figure 4.19.

Recall that for M of genus two, the hyperelliptic involution fixes six points, called Weierstrass points. For $M \in \mathcal{H}(2)$, the cone singularity z_0 is one such Weierstrass point, appearing at all vertices of all $\mathcal{P}_i$ (and therefore all vertices of $\mathcal{P}$). There is also a Weierstrass point at the center of each $\mathcal{P}_i$ (accounting for up to three more Weierstrass points). The other two to four Weierstrass points will appear at the centers of identified pairs of edges of the $\mathcal{P}_i$ which are fixed under the 180° rotation; note that in Figure 4.19, only one of each pair is marked.

To see that the three stated possibilities are the only ones, we look again at the Rauzy diagram in Figure 4.17, which has seven total nodes. Each node (combinatorial datum), except for the central one, is paired up with the node which is symmetrically

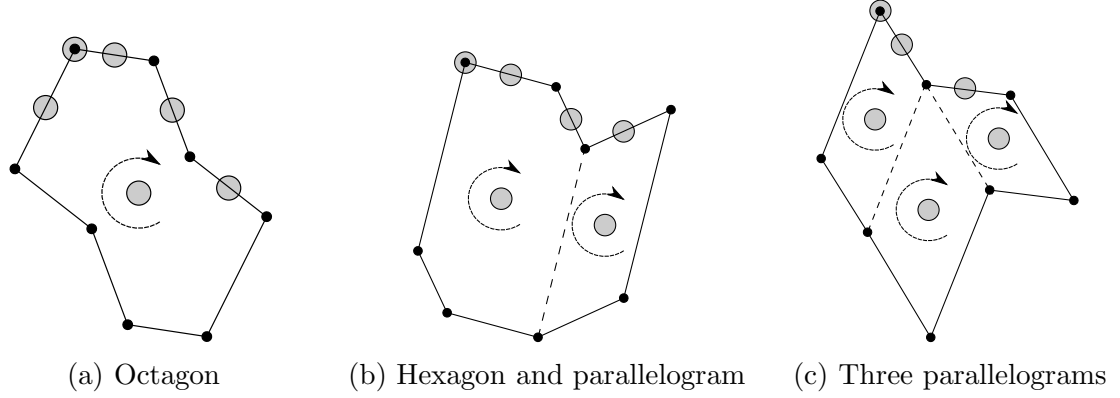


Figure 4.19: Behavior of the hyperelliptic involution in $\mathcal{H}(2)$; one appearance of each Weierstrass point is marked with a gray circle.

opposite in the diagram. Polygons associated to each pair of nodes (or the central node) all have a particular common pattern of edge identifications, determining how the 180° rotation behaves on these polygons. They break down into the 3 above categories as follows:

- Octagon: only the central node $\begin{bmatrix} ABCD \\ DCBA \end{bmatrix}$.
- Hexagon (left) and parallelogram (right): $\begin{bmatrix} ACDB \\ DCBA \end{bmatrix}$ and its symmetric opposite in the diagram $\begin{bmatrix} ABCD \\ DBAC \end{bmatrix}$.
- Hexagon (right) and parallelogram (left): $\begin{bmatrix} ADBC \\ DCBA \end{bmatrix}$ and its symmetric opposite in the diagram $\begin{bmatrix} ABCD \\ DACB \end{bmatrix}$.
- Three parallelograms: $\begin{bmatrix} ADBC \\ DCAB \end{bmatrix}$ and its symmetric opposite in the diagram $\begin{bmatrix} ABDC \\ DACB \end{bmatrix}$.

For $M \in \mathcal{H}(1,1)$ there are 5 possibilities for decomposing $\mathcal{P}$ (now a decagon) which has no fishtail into centrally symmetric subpolygons: (a) the whole decagon,

(b) an octagon and a parallelogram, (c) two hexagons, (d) a hexagon and two parallelograms, or (e) four parallelograms. These are illustrated in Figure 4.20.

The two singularities z_0 and z_1 (which together comprise all vertices of all $\mathcal{P}_i$) are exchanged by η , and so are not Weierstrass points. There are still Weierstrass points at the centers of the $\mathcal{P}_i$ (up to four of them), and the remaining two to five are contributed by the centers of identified pairs of edges of the $\mathcal{P}_i$ which are fixed under the 180° rotation. (Again, only one of each pair is marked in Figure 4.20.)

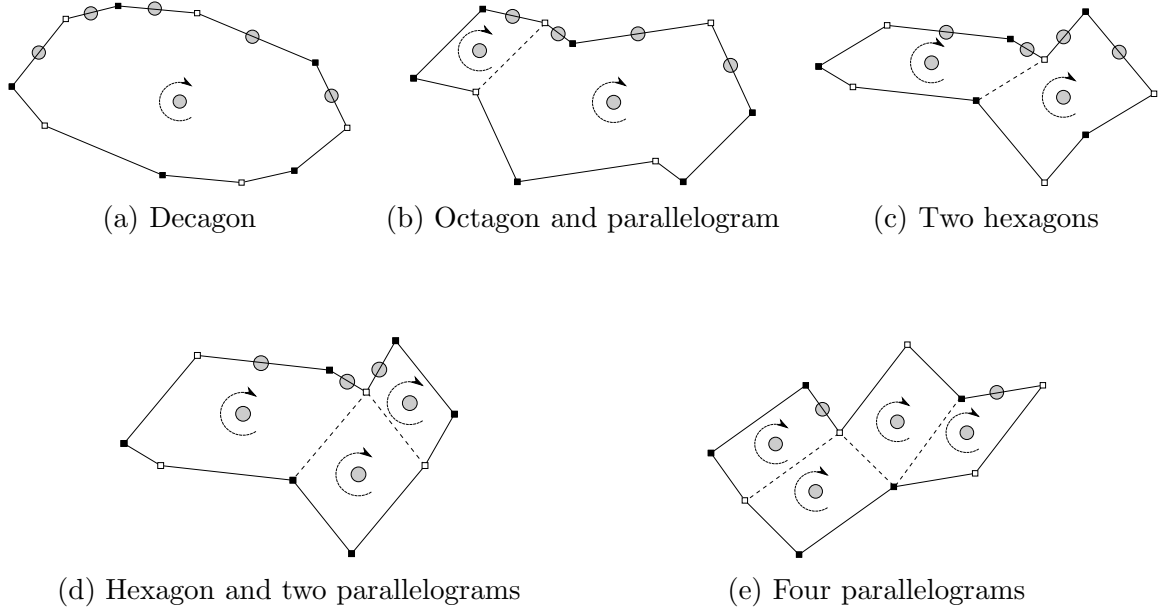


Figure 4.20: Behavior of the hyperelliptic involution in $\mathcal{H}(1,1)$; one appearance of each Weierstrass point is marked with a gray circle.

To see that there are only the five given possibilities, we look again at the Rauzy diagram in Figure 4.18. It has 15 total nodes, and again each node (except for the central one) pairs up with its symmetric opposite. As before, polygons associated

to each pair of nodes have a particular decomposition into subpolygons. They are grouped as follows:

- Decagon: only the central node $\begin{bmatrix} ABCDE \\ EDCBA \end{bmatrix}$.
- Octagon (left) and parallelogram (right): $\begin{bmatrix} ACDEB \\ EDCBA \end{bmatrix}$ and its symmetric opposite in the diagram $\begin{bmatrix} ABCDE \\ ECBAD \end{bmatrix}$.
- Octagon (right) and parallelogram (left): $\begin{bmatrix} AEBCD \\ EDCBA \end{bmatrix}$ and its symmetric opposite in the diagram $\begin{bmatrix} ABCDE \\ EADCB \end{bmatrix}$.
- Two hexagons: $\begin{bmatrix} ADEBC \\ EDCBA \end{bmatrix}$ and its symmetric opposite in the diagram $\begin{bmatrix} ABCDE \\ EBADC \end{bmatrix}$.
- Hexagon (left), and two parallelograms (center and right): $\begin{bmatrix} ADEBC \\ EDCAB \end{bmatrix}$ and its symmetric opposite in the diagram $\begin{bmatrix} ABCED \\ EBADC \end{bmatrix}$.
- Hexagon (center), and two parallelograms (left and right): $\begin{bmatrix} AEBCD \\ EDBAC \end{bmatrix}$ and its symmetric opposite in the diagram $\begin{bmatrix} ABDEC \\ EADCB \end{bmatrix}$.
- Hexagon (right), and two parallelograms (left and center): $\begin{bmatrix} AEBCD \\ EDACB \end{bmatrix}$ and its symmetric opposite in the diagram $\begin{bmatrix} ABECD \\ EADCB \end{bmatrix}$.
- Four parallelograms: $\begin{bmatrix} AEBDC \\ EDACB \end{bmatrix}$ and its symmetric opposite in the diagram $\begin{bmatrix} ABECD \\ EADBC \end{bmatrix}$.

Non-Minimal Vertical Flow

Before presenting the proof of Theorem 4.7, let us take a moment to consider M of genus two on which the vertical flow is not minimal. Recall that such M necessarily have vertical saddle connections, so $M \in \mathcal{H}_{\text{VM}}^*(2) \cup \mathcal{H}_{\text{NVM}}^*(2) \cup \mathcal{H}_{\text{VM}}^*(1, 1) \cup \mathcal{H}_{\text{NVM}}^*(1, 1)$.

Definition 4.16. The spine $\mathcal{S}$ of a surface M is the union of all vertical saddle connections and singular points of M .

For $M \in \mathcal{H}^\circ(2) \cup \mathcal{H}^\circ(1,1)$, the spine of M consists of a finite set of points: one for each singular point. Otherwise, cutting M along its spine $\mathcal{S}$ gives a finite collection $\{M_i\}_{i=1}^m$ of translation surfaces with boundary. The non-boundary points of M_i are naturally identified with points of M so that $\mathring{M}_i := M_i \setminus \partial M_i$ can be considered as an (open) subsurface of M . We have a natural map $\partial M_i \rightarrow \mathcal{S}$, but note that this map is not a bijection onto points $s \in \mathcal{S}$ such that M_i intersects more than one connected component of $B_\varepsilon(s) \setminus \mathcal{S}$ in M .

Theorem 4.17 (Decomposition of Surfaces with Non-Minimal Flow). *Suppose M is a genus two translation surface on which the vertical flow is not minimal. Then M can be written as $\bigcup_{i=1}^m \mathring{M}_i \cup \mathcal{S}$, with $m \leq 3$, where each $\mathring{M}_i$ is either a cylinder with periodic vertical flow or a genus one translation surface (a torus) with one or two slits, i.e., with one or two closed straight segments removed, on which vertical flow is minimal.*

The number and nature of the M_i and the way they identify to form M will be dictated by the so-called *ribbon graph* of M , a way to depict the structure of the spine $\mathcal{S}$ of M , which can also show how $\mathcal{S}$ lies in M . The proof of Theorem 4.17 will appear after a brief discussion of ribbon graphs.

First, we point out that Theorem 4.17 fits within the general framework of the following structure theorem for the vertical flow found in [30]:

Theorem 4.18 (Katok, Hasselblatt). *Let ϕ be a C^0 flow on a closed compact surface M defined by a uniquely integrable C^0 vector field X , and τ a transversal to X . Suppose that ϕ has a finite number of fixed points, which are orbit equivalent to (multiple) saddles or centers. Furthermore suppose that ϕ preserves a Borel probability measure that is positive on open sets. Then M splits in a ϕ -invariant way as $M = \bigcup_{i=1}^k P_i \cup \bigcup_{j=1}^l T_j \cup C$, where l is at most equal to the genus of M , P_i are open sets consisting of periodic orbits, each T_j is open, every semiorbit T_j that is not an incoming separatrix of a fixed point is dense in T_j , and C is a finite union of fixed points and saddle connections.*

Ribbon Graphs

We start with the standard definition of ribbon graph seen in [31] and will build it up to suit our needs. More about ribbon graphs can also be found in [1], [16], and [32].

Definition 4.19. A ribbon graph is a finite directed graph with an assignment of cyclic ordering on the set of undirected half-edges incident to each vertex.

Definition 4.20. For a translation surface M , the ribbon graph of M is a ribbon graph whose vertices are in bijection with the singular points of M and whose edges correspond to vertical saddle connections such that the cyclic order of half-edges incident to a vertex v is induced by the order of the saddle connections viewed in a chart near v .

The ribbon graph of M captures the essential features of the spine $\mathcal{S}$ of M . (Consequently, note that if M has no vertical saddle connections, its ribbon graph is simply a finite collection of points, one for each singular point of M .) Note also that the edges of a ribbon graph do not actually meet anywhere but at the vertices, although it may be necessary to render them with additional intersections in order to embed the graph into a plane.

A ribbon graph $\mathcal{G}$ is called so because one can construct from any such $\mathcal{G}$ an orientable surface $M_{\mathcal{G}}$ with boundary by thickening each edge. This process is described in more detail in [31], but we mention that $M_{\mathcal{G}}$ is a union of ribbons obtained by thickening the edges to topological rectangles and joining them while respecting the cyclic ordering of each vertex.

The graph $\mathcal{G}$ can be considered as a subset embedded in $M_{\mathcal{G}}$. Let a *half-ribbon* of $\mathcal{G}$ be a connected component of $M_{\mathcal{G}}$ with $\mathcal{G}$ removed. If M is a surface with $\mathcal{G}$ as its ribbon graph the half-ribbons of $\mathcal{G}$ are in bijection with open sets in M we shall call the *half-ribbons* of M . For $\varepsilon > 0$ small enough, the ε -neighborhood of the spine $\mathcal{S}$ in M is homeomorphic to $M_{\mathcal{G}}$. This homeomorphism sends $\mathcal{S}$ to $\mathcal{G}$ (treated as a subset of $M_{\mathcal{G}}$).

Definition 4.21. A half-ribbon of M is a connected component of the complement of $\mathcal{S}$ in an ε -neighborhood of $\mathcal{S}$ in M , for sufficiently small $\varepsilon > 0$.

Now, to better capture the way in which $\mathcal{S}$ lies in M , we introduce the following:

Definition 4.22. A measured ribbon graph is a ribbon graph with an assignment of

a number θ to every two half-edges e and e' incident to the same vertex that are adjacent in the cyclic order. θ is called the angle between e and e' . For a surface M , the measured ribbon graph of M is then the ribbon graph of M with the angle between any two edges given by the corresponding angle in M .

For the measured ribbon graph of a translation surface M , the sum of all angles between pairs of adjacent edges of a given vertex must be the angle at the corresponding singular point of M . Additionally, if two adjacent edges e and e' are both incoming or both outgoing (with respect to their common vertex), θ must be an even multiple of π , and if one of e and e' is incoming and the other is outgoing, then θ must be an odd multiple of π .

To each stratum $\mathcal{H}(\dots)$ of translation surfaces of a given genus we associate a family $\text{MRG}(\dots)$ of measured ribbon graphs. It consists of graphs with vertices in bijection with the singularities of the surfaces in this stratum and the configuration of half-edges at a vertex compatible with the configuration of verticals at that singularity. For instance, $\text{MRG}(2)$ consists of measured ribbon graphs with one vertex with total angle 6π and up to three edges, and $\text{MRG}(1, 1)$ consists of measured ribbon graphs with two vertices each with total angle 4π and up to four edges.

Thus the spines of the surfaces in $\mathcal{H}(\dots)$ yield graphs in $\text{MRG}(\dots)$, though some graphs in the family are not actually realizable by a surface. Theorem 4.24 identifies all the (isomorphism classes of) graphs in $\text{MRG}(2)$ and $\text{MRG}(1, 1)$ which are realizable by a surface in $\mathcal{H}(2)$ and $\mathcal{H}(1, 1)$, respectively.

Now, ribbon graphs of surfaces of genus two can be endowed with additional structure due to the presence of the hyperelliptic involution η on M . Hence, we define:

Definition 4.23. A hyperelliptic measured ribbon graph is a measured ribbon graph with an involution η of the underlying unoriented graph that reverses orientation of every edge and preserves the cyclic order of half-edges.

More precisely, for a particular hyperelliptic measured ribbon graph $\mathcal{G}$, η consists of a mapping η_0 of the vertices of $\mathcal{G}$ and a mapping η_1 of the set $\mathcal{E}$ of edges of $\mathcal{G}$ such that, for all $e \in \mathcal{E}$, where e_v and e_w are the vertices of e , $\eta_1(e)$ is an anti-edge of $\mathcal{G}$ with vertices $\eta_0(e_v)$ and $\eta_0(e_w)$. (At this point it would be helpful to recall that for $M \in \mathcal{H}(2)$, $\eta(z_0) = z_0$, and for $M \in \mathcal{H}(1, 1)$, we have $\eta(z_0) = z_1, \eta(z_1) = z_0$.)

The measured ribbon graph of a translation surface can be elevated to a hyperelliptic measured ribbon graph provided the surface is hyperelliptic. Thus, for surfaces M of genus two, the measured ribbon graph of M can always be considered as a hyperelliptic measured ribbon graph. So, $M \in \mathcal{H}^*(2) \cup \mathcal{H}^*(1, 1)$ can only have a measured ribbon graph on which can be defined an involution which reverses the orientation of each edge and preserves the cyclic order of half-edges incident to each vertex, as in definition 4.23.

In this section, such a hyperelliptic measured ribbon graph will be depicted with the action of η on half-edges represented by 180° rotation of the graph, with the edges following suit. Also, the angle between each pair of adjacent half-edges is indicated

by the depiction of a half-edge every π radians, even if this half-edge is not actually joined to another to complete it to a full edge, and so is not a proper part of the graph.

Additionally, although a measured ribbon graph is an oriented graph, each will be rendered here without orientation of edges indicated. For clarity, the orientation of half-edges (which alternates between outgoing or incoming, when including the “angle marker” half-edges mentioned above) at each vertex will be the same across all graphs from each family, consistent with that seen in Figure 4.21.



Figure 4.21: Each ribbon graph in $\text{MRG}(2)$ presented here will have its half-edges oriented as the figure on the left, and each ribbon graph in $\text{MRG}(1, 1)$ presented here will have its half-edges oriented as the figure on the right.

Towards the goal of classifying surfaces of genus two, we will first establish the following:

Theorem 4.24. *The hyperelliptic measured ribbon graph of any surface $M \in \mathcal{H}^*(2)$ coincides up to isomorphism with one of those seen in Figure 4.22. The hyperelliptic measured ribbon graph of any surface $M \in \mathcal{H}^*(1, 1)$ coincides up to isomorphism with one of those seen in Figure 4.23.*

The subsets of $\text{MRG}(2)$ and $\text{MRG}(1, 1)$ seen in Figures 4.22 and 4.23, respectively, can each be further grouped according to whether the vertical flow on an

associated surface is (a) minimal, (b) everywhere periodic, and (c) neither minimal nor everywhere periodic. This grouping will later be established in the course of proving Theorem 4.17, but is useful to point out now for clarity of notation. The letters M, P, and N are used to indicate these three cases, respectively, followed by the number of half-ribbons of a surface M which has the given hyperelliptic measured ribbon graph.

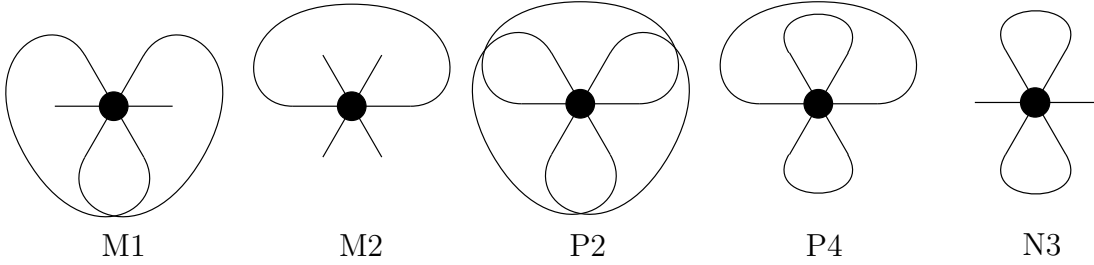


Figure 4.22: Hyperelliptic measured ribbon graphs for $M \in \mathcal{H}^*(2)$. Surfaces having one of the first two graphs have vertical flow which is minimal, surfaces having one of the next two have vertical flow which is everywhere periodic, and surfaces with having the last one have vertical flow which is neither everywhere periodic nor minimal.

Measured ribbon graphs from $\text{MRG}(2)$ and $\text{MRG}(1, 1)$ which are not realizable by a surface in $\mathcal{H}^*(2)$ or $\mathcal{H}^*(1, 1)$ are seen in Figures 4.24 and 4.25, respectively.

Proof of Theorem 4.24. Consider $\text{MRG}(2)$ first, seen in Figures 4.22 and 4.24. $\text{MRG}(2)$ contains 2 distinct graphs with one edge, 4 with two edges, and 3 with three edges. Of these 9 graphs, the 5 depicted in Figure 4.22 admit an involution η that reverses orientation of edges and preserves cyclic order of half-edges, as required by Definition 4.23. As mentioned earlier, this involution acts on the half-edges by rotation of the

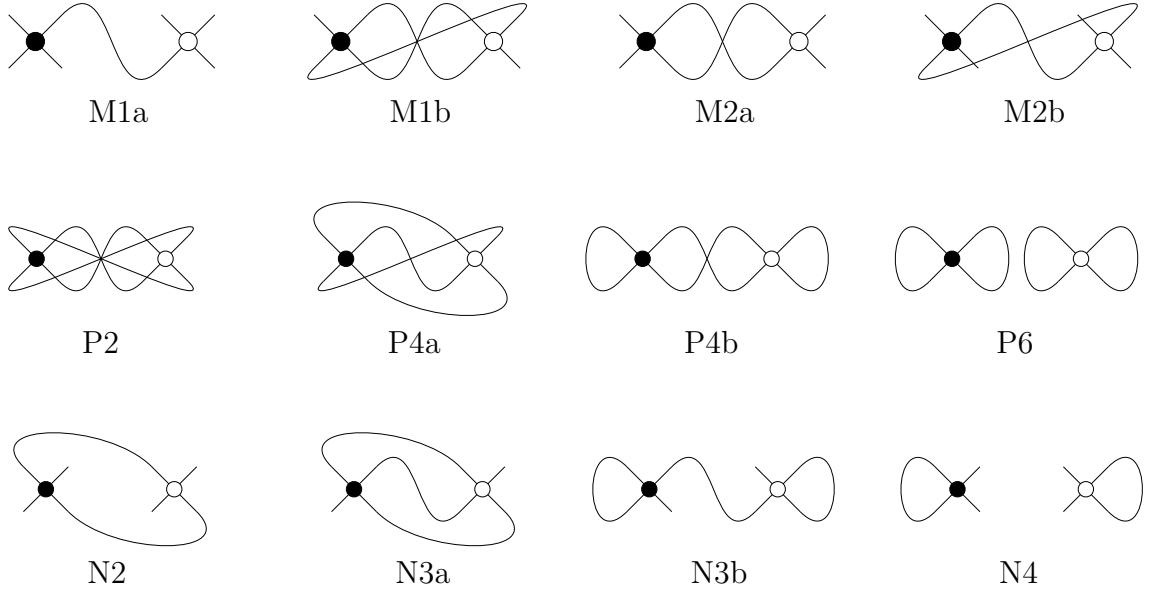


Figure 4.23: Hyperelliptic measured ribbon graphs for $M \in \mathcal{H}^*(1, 1)$. Surfaces having a graph in the first row have vertical flow which is minimal, surfaces having a graph in the second row have vertical flow which is everywhere periodic, and surfaces having a graph in the last row have vertical flow which is neither everywhere periodic nor minimal.

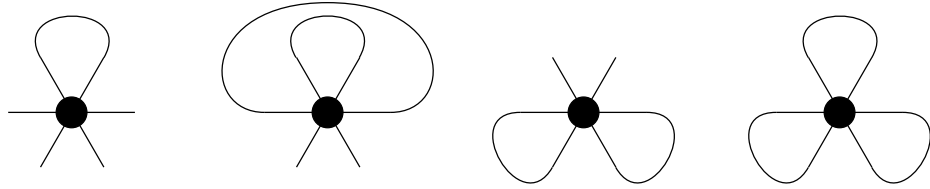


Figure 4.24: Measured ribbon graphs from $\text{MRG}(2)$ which are not realizable by a surface in $\mathcal{H}^*(2)$.

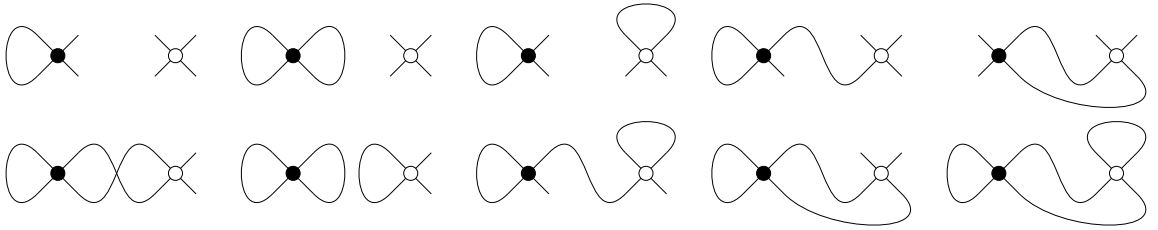


Figure 4.25: Measured ribbon graphs from $\text{MRG}(1, 1)$ which are not realizable by a surface in $\mathcal{H}^*(1, 1)$.

graph by 180° , and the edges map accordingly.

In contrast, every involution admitted by each of the graphs in Figure 4.24 which reverses orientation of edges does not preserve the cyclic order of half-edges, and every involution which preserves the cyclic order of half-edges does not reverse the orientation of edges. Thus each measured ribbon graph seen in Figure 4.24 cannot be elevated to a hyperelliptic measured ribbon graph, and so none of them are realizable by a surface in $\mathcal{H}^*(2)$.

For $\text{MRG}(1, 1)$, seen in Figures 4.23 and 4.25, there are 2 graphs with one edge, 8 with two edges, 7 with three edges, and 5 with four edges. (Note that some of these graphs are isomorphic when viewed only as ribbon graphs, without the measured structure.) As above, the 12 of these 22 graphs which are depicted in Figure 4.23 admit an involution η that reverses orientation of edges and preserves cyclic order of half-edges. For the other 10, seen in Figure 4.25, no such involution exists with both of these properties.

In the above argument we silently used that each measured ribbon graph in Figures 4.22 and 4.23 has a unique involution turning it into a hyperelliptic measured ribbon graph (namely, the one induced by the 180° rotation of its depiction). This is true with the exception of the graph labeled P2 in Figure 4.23, which has an additional involution that reverses the orientation of edges and preserves the cyclic order of half-edges incident to each vertex. This graph has been rendered in Figure 4.26a to realize this additional involution by 180° rotation. To complete the proof,

we show that the associated hyperelliptic measured ribbon graph cannot be realized.

Suppose a surface $M \in \mathcal{H}^*(1, 1)$ has this hyperelliptic measured ribbon graph, with the involution on the graph as indicated. Such a surface has four saddle connections $\alpha_1, \alpha_2, \beta_1, \beta_2$, so the spine $\mathcal{S}$ is given by the union of these saddle connections, together with the singular points z_0 and z_1 .

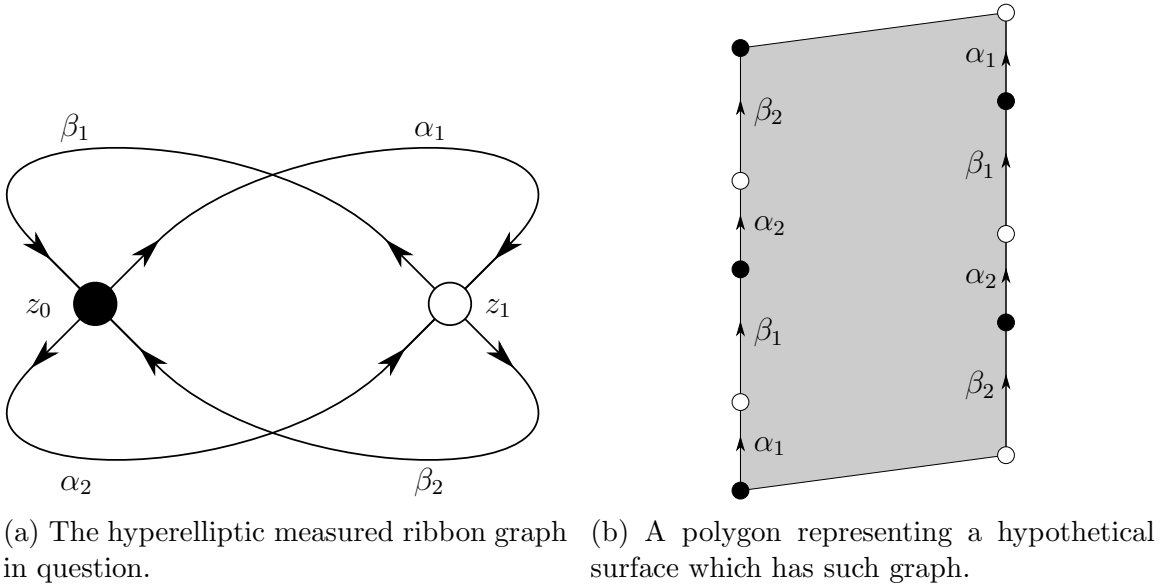


Figure 4.26: Detail of the last measured ribbon graph from Figure 4.25.

Both vertices encounter these edges in the (counterclockwise) cyclic order $\alpha_1, \beta_1, \alpha_2, \beta_2$.

Additionally, with the involution on the graph realized by 180° rotation, we have

$$\eta(\alpha_1) = -\alpha_2, \quad \eta(\beta_1) = -\beta_2, \quad \eta(\alpha_2) = -\alpha_1, \quad \eta(\beta_2) = -\beta_1. \quad (4.1)$$

So, none of these four edges are fixed by η and instead they come in two exchanged pairs. Note in particular that this means the lengths of α_1 and α_2 are equal, as are the lengths of β_1 and β_2 .

Define closed curves $\gamma_1 := \alpha_1\beta_1\alpha_2\beta_2$ and $\gamma_2 := \beta_2\alpha_2\beta_1\alpha_1$. We see that M has two half-ribbons: $\mathcal{R}_1$, which abuts the right side of γ_1 , and $\mathcal{R}_2$, which abuts the left side of γ_2 .

Vertical flow on $\mathcal{R}_1$ has closed trajectories parallel to γ_1 , and vertical flow on $\mathcal{R}_2$ has closed trajectories parallel to γ_2 , so flow on all of $M \setminus \mathcal{S}$ must be periodic. As neither γ_1 nor γ_2 is null-homologous, $\mathcal{R}_1$ and $\mathcal{R}_2$ must be part of a single periodic cylinder which comprises all of $M \setminus \mathcal{S}$. Indeed, M can be depicted as a polygon such as the one seen in Figure 4.26b. Here γ_1 forms the left side of the polygon and γ_2 forms the right.

Now, note that rotating this polygon 180° about its center point produces an involution which fixes 6 points (the center point of the polygon, the midpoint of the top/bottom edge, and the midpoints of each α_i and β_i). Thus, this rotation corresponds to the hyperelliptic involution. Since this rotation fixes each α_i and β_i , rather than exchanging them in pairs as required by (4.1), uniqueness of the hyperelliptic involution yields the conclusion that the given involution on the graph does not correspond to the hyperelliptic involution on the surface.

□

Thus the graphs seen in Figures 4.22 and 4.23 are the only ones which are realizable by a surface. In the next section, we will see examples of each of these, establishing that they are all indeed realized.

Classifying the M_i

Now, to address the claim in Theorem 4.17 about each M_i being a cylinder or slitted torus, we first recall the Gauss-Bonnet Theorem:

Theorem 4.25 (Gauss-Bonnet). *Let K be the Gaussian curvature of M_i and k be the geodesic curvature of ∂M_i . Then*

$$\int_{M_i} K + \int_{\partial M_i} k = 2\pi\chi(M_i), \quad (4.2)$$

where $\chi(M_i)$ is the Euler characteristic of M_i .

Let $X_i := Z(\omega) \cap \overset{\circ}{M}_i$ be the set of interior points of M_i corresponding to singular points of M and X'_i be the set of points of ∂M_i corresponding to singular points of M (noting that each singular point of M can contribute several points to X'_i).

As M is a translation surface, K is zero on the complement of X_i . Further, since we restrict to M of genus two, from the allowed ribbon graphs we see that all singular points of M are in the spine $\mathcal{S}$. Thus we have that X_i is empty and $K = 0$. In general, however,

$$K = \sum_{x \in X_i} 2\pi(1 - m_x)\delta_x, \quad (4.3)$$

where m_x is the degree of x in M_i and δ_x is the Dirac delta function concentrated at x .

Also, since ∂M_i is a finite union of geodesic segments, k is zero on the complement of X'_i . Indeed,

$$k = \sum_{x \in X'_i} -\theta_x \delta_x, \quad (4.4)$$

where θ_x is the angle by which ∂M_i exceeds π as it passes x , and again δ_x is the Dirac delta function concentrated at x .

Thus, for genus two surfaces, the Gauss-Bonnet Theorem states that

$$\chi(M_i) = -\frac{1}{2\pi} \sum_{x \in X'_i} \theta_x. \quad (4.5)$$

Now, recall that for a closed orientable surface M_i of genus g , $\chi(M_i) = 2 - 2g$. If M_i has b boundary components, each of these forms an additional edge in the formulation $\chi(M_i) = V - E + F$, so $\chi(M_i) = 2 - 2g - b$. In particular:

- If M_i is a cylinder, then $\chi(M_i) = 2 - 0 - 2 = 0$.
- If M_i is a torus with one slit (topologically, a torus with a disk removed), then $\chi(M_i) = 2 - 2 - 1 = -1$.
- If M_i is a torus with two slits, then $\chi(M_i) = 2 - 2 - 2 = -2$.

Proof of Theorem 4.17

Using Theorem 4.24 as a guide to what cases need to be considered, and the Gauss-Bonnet Theorem to classify the M_i , we are ready to prove Theorem 4.17.

Proof of Theorem 4.17. We start by considering each hyperelliptic measured ribbon graph from Figures 4.22 and 4.23. We will discuss the graphs labeled M first, followed by those of type P and then type N.

1. Graphs of type M. (See Figure 4.27.)

- $\mathcal{H}^*(2)$, M1: By tracing $\mathcal{G}$ we see that $M_{\mathcal{G}}$ consists of just one-half ribbon (which accounts for the full neighborhood of z_0). Thus there is exactly

one component M_1 , and vertical flow on M_1 is minimal. M_1 is obtained by cutting M open along $\mathcal{S}$, which is a loop in M (albeit not a simple one as it passes twice through z_0). ∂M_1 passes z_0 twice, with an angle of 2π (i.e., an excess of π) each time, so Gauss-Bonnet reads: $2 - 2g - b = -\frac{1}{2\pi}(\pi + \pi) = -1$. Hence $2g + b = 3$, and as ∂M_1 is a single loop, $b = 1$. Thus $g = 1$ and M_1 is a torus with a single slit.

- $\mathcal{H}^*(2)$, M2: Now there are two half-ribbons, so there may be two components. However, the lone saddle connection is not null-homologous in M (as can be seen by integrating ω^u along it), so it does not bound a proper subsurface. Thus again we only have one component M_1 , and vertical flow on it is minimal. Here ∂M_1 passes z_0 twice with an angle of 3π each time, so Gauss-Bonnet reads: $2 - 2g - b = -\frac{1}{2\pi}(2\pi + 2\pi) = -2$. Hence $b = 2$, $g = 1$, and M_1 is topologically a torus with two disks removed.
- $\mathcal{H}^*(1,1)$, M1a: There is a single half-ribbon, and thus a single minimal component M_1 . Gauss-Bonnet reads: $2 - 2g - b = -\frac{1}{2\pi}(3\pi + 3\pi) = -3$. Hence $2g + b = 5$, and as ∂M_1 is a single slit, $b = 1$ and $g = 2$. (So this case does not yield a cylinder or genus one surface. It will prove to be the only one.)
- $\mathcal{H}^*(1,1)$, M1b: Again there is a single ribbon and single minimal component M_1 . Gauss-Bonnet reads: $2 - 2g - b = -\frac{1}{2\pi}(\pi + \pi) = -1$ and M_1 is a slitted torus.

- $\mathcal{H}^*(1, 1)$, M2a: Now there are two half-ribbons, but the boundary between them is formed by a union of two saddle connections which is not homologous to zero. Thus again we have a single minimal component M_1 . Gauss-Bonnet reads: $2 - 2g - b = -\frac{1}{2\pi}(2\pi + 2\pi) = -2$ and M_1 is a doubly-slitted torus.
- $\mathcal{H}^*(1, 1)$, M2b: Again there are two-half-ribbons, bounded by a union of two saddle connections which is not homologous to zero, so there is a single minimal component M_1 . Gauss-Bonnet reads: $2 - 2g - b = -\frac{1}{2\pi}(\pi + \pi + \pi + \pi) = -2$ and again M_1 is a doubly-slitted torus.

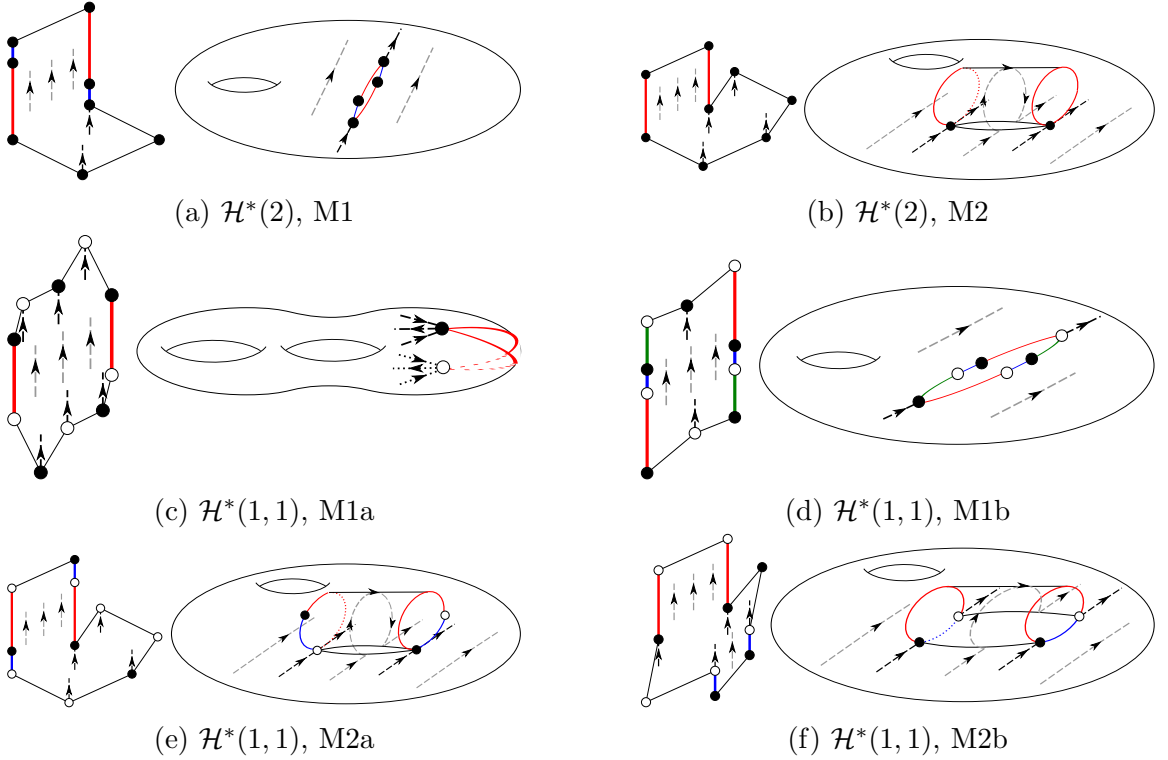


Figure 4.27: Geometric and conceptual depictions of the M cases.

So, on a surface with one of the 6 ribbon graphs of type M, vertical flow on $M = M_1$ is minimal and thus the surface belongs in the context of Theorem 4.7 rather than Theorem 4.17 (so indeed, the genus two surface arising in case M1a does not invalidate the claim in Theorem 4.17 regarding the nature of the M_i).

2. Graphs of type P. (See Figure 4.28.)

- As every half-edge of the graph is part of an edge, each half-ribbon in these cases contains closed trajectories. So, flow is periodic on all of M and all M_i are cylinders. Each half-ribbon represents half the boundary of a cylinder, so there are half as many cylinders as half-ribbons: one, two, or three of them.

3. Graphs of type N. (See Figure 4.29.)

- $\mathcal{H}^*(2)$, N3: Here there are three half-ribbons. Note that two of them have closed trajectories and thus pair up to form a cylinder M_1 . The third is part of some component M_2 on which vertical flow is minimal. We see that ∂M_2 passes z_0 twice, each time with an angle of 2π . Thus Gauss-Bonnet yields: $2 - 2g - b = -\frac{1}{2\pi}(\pi + \pi) = -1$, so M_2 is a slitted torus.
- $\mathcal{H}^*(1,1)$, N2: There are two half-ribbons. The union of the two saddle connections is homologous to zero, so there are two components M_1 and M_2 , and we see that on both of these components, vertical flow is minimal. In each of M_1 and M_2 , $\partial M_1 = \partial M_2$ passes z_0 and z_1 once each, with an

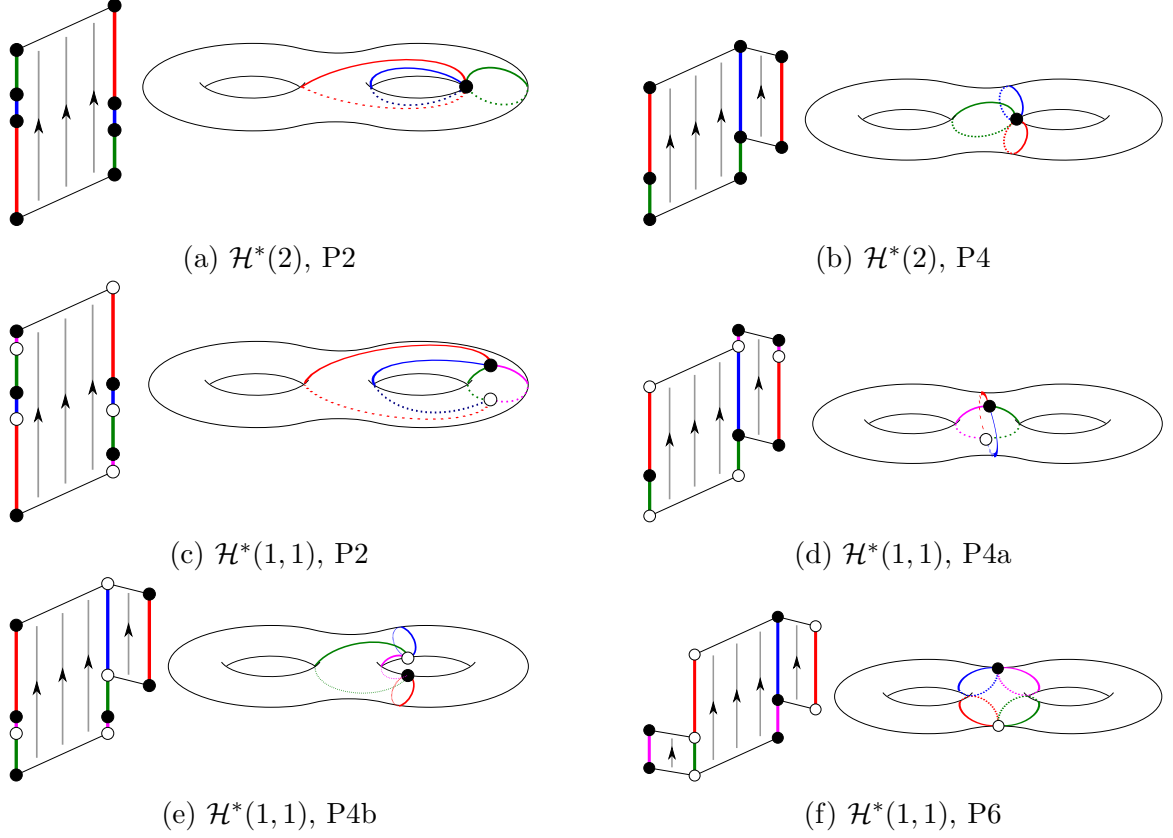


Figure 4.28: Geometric and conceptual depictions of the P cases.

angle of 2π each time. Gauss-Bonnet yields $\chi(M_1) = \chi(M_2) = -\frac{1}{2\pi}(\pi + \pi) = -1$, so M_1 and M_2 are both slitted tori.

- $\mathcal{H}^*(1,1)$, N3a: In this case, two of the three ribbons have closed trajectories and so must pair up to form a cylinder M_1 . The third is part of some component M_2 on which vertical flow is minimal. Gauss-Bonnet yields $\chi(M_2) = -\frac{1}{2\pi}(\pi + \pi) = -1$, so M_2 is a slitted torus.
- $\mathcal{H}^*(1,1)$, N3b: Again, two of the three ribbons have closed trajectories so let M_1 be the cylinder containing these two. The third is part of some

component M_2 on which vertical flow is minimal. Now Gauss-Bonnet yields

$$\chi(M_2) = -\frac{1}{2\pi}(2\pi + 2\pi) = -2. \text{ Since } \partial M_2 \text{ has two components, solving}$$

$2 - 2g - 2 = -2$ gives $g = 1$ and M_2 is a torus with two slits.

- $\mathcal{H}^*(1,1)$, N4: Here there are four ribbons. Two of them have closed trajectories, so they pair up to form a periodic cylinder M_1 . The other two must be part of the same component M_2 as there is nothing to separate them. Gauss-Bonnet yields $\chi(M_2) = -\frac{1}{2\pi}(2\pi + 2\pi) = -2$, so here again M_2 is a torus with two slits.

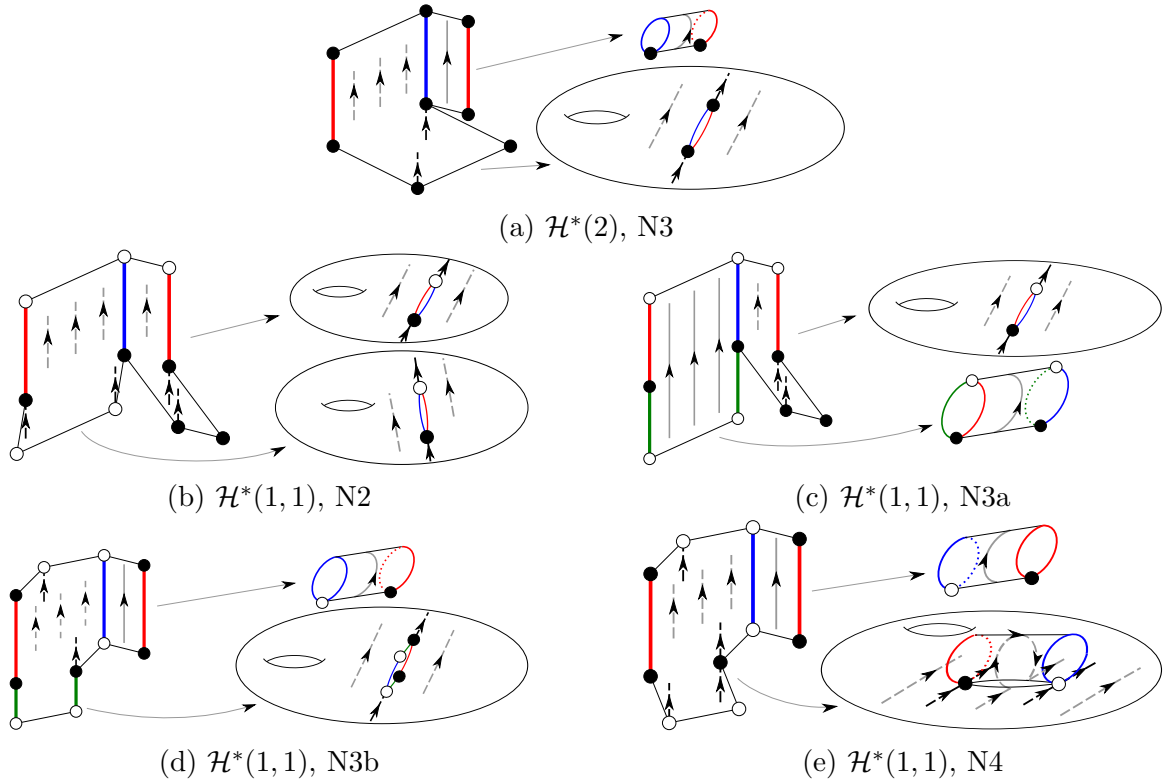


Figure 4.29: Geometric and conceptual depictions of the N cases.

□

Proof of Theorem 4.7

We are now ready to prove Theorem 4.7, reproduced here:

Theorem 4.7 *Suppose M is a genus two translation surface with vertical flow which is minimal. Then there exists a parallel pair of L -cuts (K, K') in M . Moreover, if M has no vertical and no horizontal saddle connections then the L -cuts K and K' are non-singular.*

Rectangles

The collections of surfaces $\mathcal{H}^\circ(2)$, $\mathcal{H}^\circ(1, 1)$, $\mathcal{H}_{\text{VM}}^*(2)$ and $\mathcal{H}_{\text{VM}}^*(1, 1)$ will be considered separately. By Lemmas 4.14 and 4.15, we can assume that M has a polygonal representation with combinatorial datum given by the $\begin{bmatrix} ABCD \\ DCBA \end{bmatrix}$ for $M \in \mathcal{H}^\circ(2) \cup \mathcal{H}_{\text{VM}}^*(2)$ or $\begin{bmatrix} ABCDE \\ EDCBA \end{bmatrix}$ for $M \in \mathcal{H}^\circ(1, 1) \cup \mathcal{H}_{\text{VM}}^*(1, 1)$.

As previewed on page 44, the argument will proceed by seeking a rectangle in M with certain desirable properties. Let us take a moment to formally define what makes a rectangle suitable for these purposes.

Definition 4.26. *A rectangle in M is a subset R of M that is a closure of the image of a homeomorphism from an open rectangle in $\mathbb{R}^2$ that is locally an isometry. We also require that the boundary of R is a union of segments of the horizontal and vertical foliations.*

We will find a rectangle in M by taking a polygon representation $\mathcal{P}$ of M and

considering *candidate rectangles*, proper rectangles (with vertical and horizontal sides) in the plane containing $\mathcal{P}$, with the tacit assumption that any protrusions (typically triangular) of these rectangles extending beyond $\mathcal{P}$ are to be translated to the interior of $\mathcal{P}$, as prescribed by edge identification of $\mathcal{P}$. Accordingly, it is possible for a candidate rectangle R_* to contain a vertex V of $\mathcal{P}$ (which corresponds to a singular point of M) in its interior. We will refer to such a situation as an *invasion* of R_* by V .

Since V has a neighborhood with angle 6π (if $M \in \mathcal{H}(2)$) or 4π (if $M \in \mathcal{H}(1, 1)$), a candidate rectangle R_* invaded by V cannot be a rectangle in M . On the other hand, if R_* has no invasion by a vertex of $\mathcal{P}$, then every point in the interior of R_* has an angle of 2π and the required local isometry exists. Thus, a candidate rectangle R_* is indeed a rectangle in M if and only if it is not invaded by a vertex of $\mathcal{P}$. See Figure 4.30.

$M \in \mathcal{H}^\circ(2)$

Lemma 4.27 (Rectangle in $M \in \mathcal{H}^\circ(2)$). *For $M \in \mathcal{H}^\circ(2)$, there exists a rectangle R in M , with z_0 appearing at two opposite corners of R , such that $R \neq \eta(R)$.*

Proof. R will emerge as one of two candidate rectangles: R_1 , with a diagonal homotopic to $D + A$, and R_2 , with a diagonal homotopic to $C + D + A$; both are depicted in Figure 4.31. Observe that neither of R_1 and R_2 (should it prove to be a rectangle in M) is fixed by η (which acts by 180° rotation of $\mathcal{O}$; recall Figure 4.19a on page

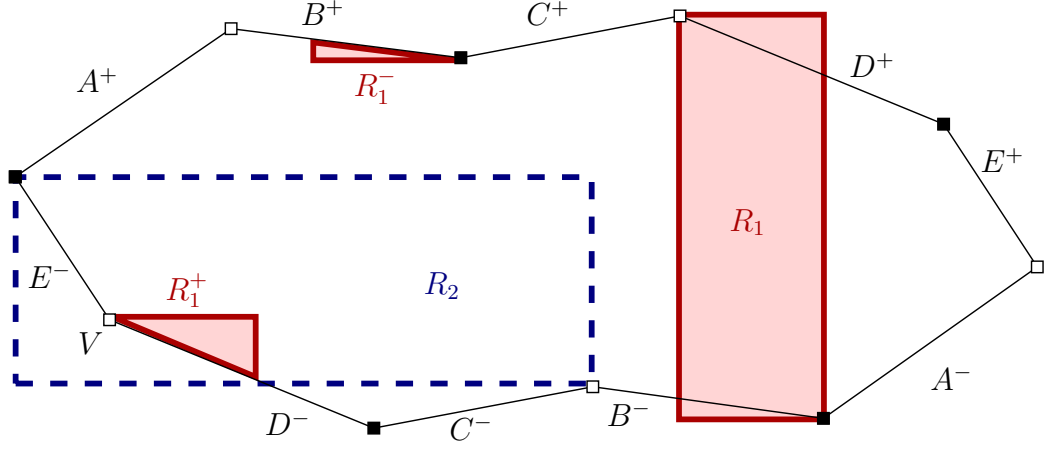


Figure 4.30: The candidate rectangle R_1 only encounters singularities at its corners; its interior is free of invasion, so it is indeed a rectangle in M . The triangular regions R_1^+ and R_1^- have been translated to the interior of the decagon $\mathcal{D}$ according to the identifications of $D^{+/-}$ and $B^{+/-}$, respectively. In contrast, the candidate rectangle R_2 is invaded by the vertex V , and so fails to be a rectangle in M . In particular, R_2 contains the point V which has a 4π angle, and so cannot be isometrically mapped to a rectangle in M .

73). The proof will proceed by establishing that for any $\mathcal{O}$ with combinatorial datum $\begin{bmatrix} ABCD \\ DCBA \end{bmatrix}$, at least one of the rectangles R_1 and R_2 is free of invasion by a vertex of $\mathcal{O}$.

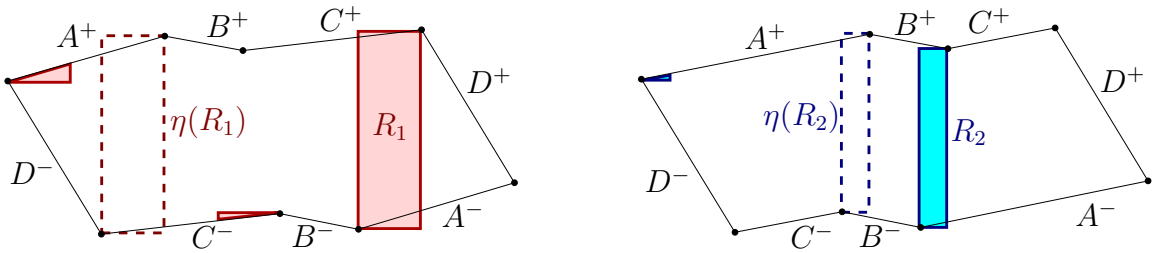


Figure 4.31: Candidate rectangles R_1 and R_2 for $\mathcal{H}^\circ(2)$.

Recall from page 59 that for $S \in \{A, B, C, D\}$, S^+ denotes the edge of $\mathcal{O}$ which

came from $\mathcal{C}^+$ (so $\mathcal{O}$ lies below S^+), and S^- denotes the edge of $\mathcal{O}$ which came from $\mathcal{C}^-$ (so $\mathcal{O}$ lies above S^-). Also recall that λ_S is the length of the segment of I corresponding to S . Note that since $M \in \mathcal{H}^\circ(2)$, $\lambda_S > 0$ for all S , and moreover, $\lambda_A \neq \lambda_D$, $\lambda_A \neq \lambda_C + \lambda_D$, and $\lambda_A \neq \lambda_B + \lambda_C + \lambda_D$ (among other such combinations of lengths which we won't explicitly mention). Add a temporary assumption that M has no horizontal saddle connections, which means no edge S of $\mathcal{O}$ can be horizontal.

It can be assumed that $\lambda_A > \lambda_D$, as if this is not the case, reversing the direction of the vertical flow and relabeling the segments in the opposite order would make it so. Also, if $\lambda_A > \lambda_B + \lambda_C + \lambda_D$, M 's path through the Rauzy diagram allows for a sequence of three successive “1” operations (consult Figure 4.17 on page 68), leading back to the central node and replacing $\mathcal{O}$ with its image under this sequence. Repeating this sequence as many times as necessary allows us to assume we are in the case where $\lambda_A < \lambda_B + \lambda_C + \lambda_D$. Thus, the only remaining variability in the relative lengths is whether λ_A is longer or shorter than $\lambda_C + \lambda_D$. Both cases are illustrated in Figure 4.32.

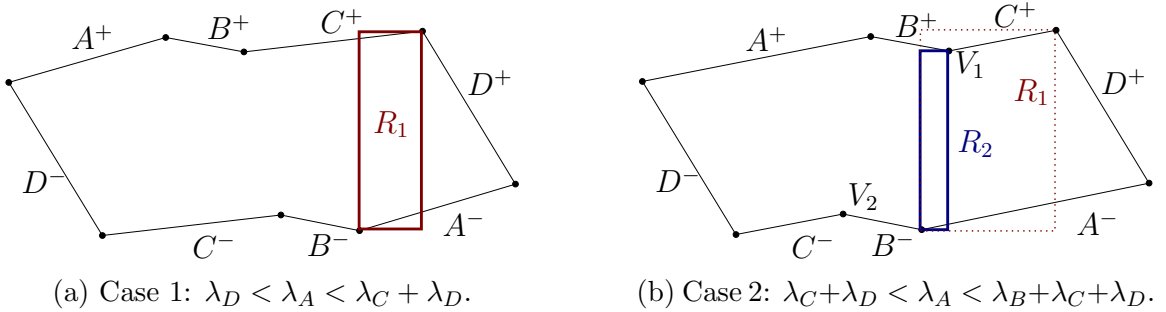


Figure 4.32: Representative diagrams for $M \in \mathcal{H}^\circ(2)$ (with $\lambda_D < \lambda_A < \lambda_B + \lambda_C + \lambda_D$).

Case 1: $\lambda_D < \lambda_A < \lambda_C + \lambda_D$.

- No vertex can invade R_1 .

Case 2: $\lambda_C + \lambda_D < \lambda_A < \lambda_B + \lambda_C + \lambda_D$.

- Now if C has positive slope the vertex V_1 at the left end of C^+ invades R_1 .

However, V_2 at the left end of B^- (which may have invaded R_2 in case 1) is now far enough to the left that R_2 has no invasion.

Now, if M does have horizontal saddle connections, the argument above can be repeated. However, some R may have horizontal edges which cross a vertex of $\mathcal{O}$, resulting in singular L-cuts. $\square$

$M \in \mathcal{H}^\circ(1, 1)$

Lemma 4.28 (Rectangle in $M \in \mathcal{H}^\circ(1, 1)$). *For $M \in \mathcal{H}^\circ(1, 1)$, there is a rectangle R in M , with z_0 and z_1 appearing at two opposite corners of R , such that $R \neq \eta(R)$.*

Proof. This time 5 candidate rectangles are considered for R : R_1 , with a diagonal homotopic to $D + E + A$, R_2 , with a diagonal homotopic to $E + A + B$, R_3 , with a diagonal homotopic to $A + B + C$, R_4 , with a diagonal homotopic to $C + D + E$, and R_5 , with a diagonal homotopic to $B + C + D$. These are depicted in Figure 4.33.

Once again, the proof will proceed by establishing that for any $\mathcal{D}$ with combinatorial datum $\begin{bmatrix} ABCDE \\ EDCBA \end{bmatrix}$, at least one of the rectangles R_1 through R_5 is free of invasion by a vertex of $\mathcal{D}$, and is thus a rectangle in M . As before, note that all five candidate rectangles satisfy $R_i \neq \eta(R_i)$ (for any of the R_i which is actually a

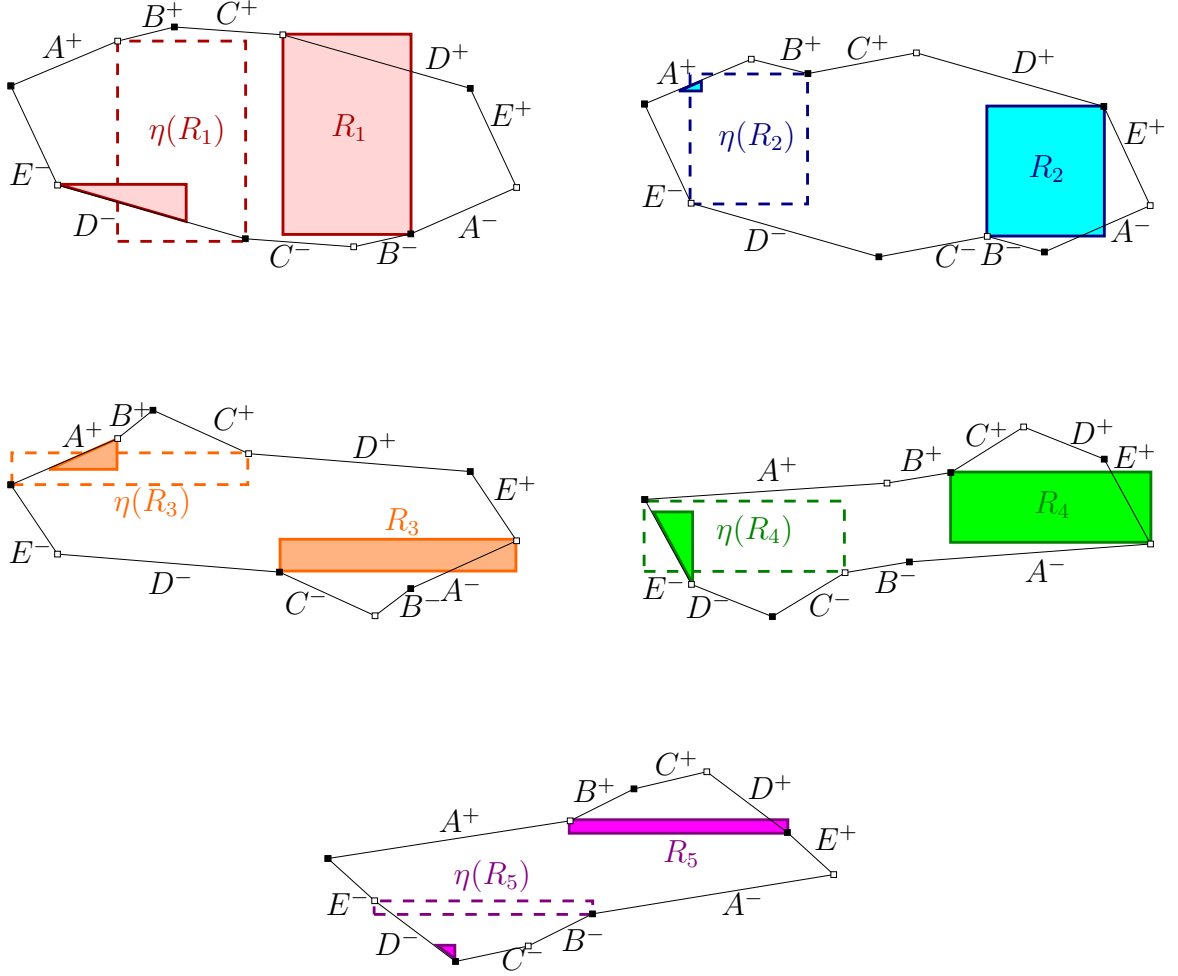


Figure 4.33: Candidate rectangles R_1 through R_5 for $\mathcal{H}^\circ(1,1)$.

rectangle in M).

Again, $M \in \mathcal{H}^\circ(1,1)$ means $\lambda_S > 0$ for $S \in \{A, B, C, D, E\}$ as well as $\lambda_A \neq \lambda_E$, $\lambda_A \neq \lambda_D + \lambda_E$, $\lambda_A \neq \lambda_C + \lambda_D + \lambda_E$, and $\lambda_A \neq \lambda_B + \lambda_C + \lambda_D + \lambda_E$. We temporarily assume M has no horizontal saddle connections, so no edge S of $\mathcal{D}$ can be horizontal.

As with the $\mathcal{H}^\circ(2)$ case, $\mathcal{D}$ can be taken to satisfy $\lambda_A > \lambda_E$, as otherwise we could reverse the direction of the vertical flow and relabel the edges in the opposite order.

Also, we can assume $\lambda_A < \lambda_B + \lambda_C + \lambda_D + \lambda_E$, as if this is not the case, M 's path through the Rauzy diagram allows for a sequence of 4 consecutive “1” operations (as seen in Figure 4.18 on page 69), leading back to the central node and replacing $\mathcal{D}$ by its image under this sequence.

This leaves 3 major cases for the lengths, again arranged by the relative length of λ_A compared to sums of other lengths. Illustrations for cases 1 and 2 are seen in Figure 4.34 and for case 3 are seen in Figure 4.35.

Case 1: $\lambda_E < \lambda_A < \lambda_D + \lambda_E$.

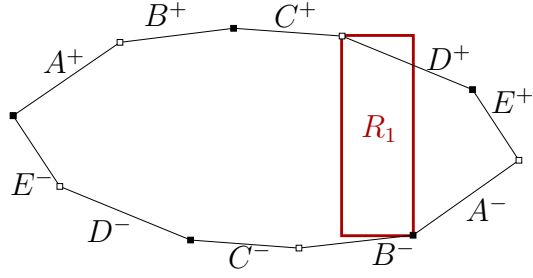
- Subcase (a), $\lambda_A + \lambda_B > \lambda_D + \lambda_E$: Here the rectangle R_1 is always free of invasion.
- Subcase (b), $\lambda_A + \lambda_B < \lambda_D + \lambda_E$: Now two vertices can invade R_1 . If B has negative slope, the vertex V_1 at the left end of B^- invades R_1 . However, in this case R_2 is free of invasion. Another possibility is that the vertex V_2 at the left end of C^- invades R_1 , but by doing so it must be higher than the vertex V_3 at the right end of B^- , making R_3 free of invasion. (Note that R_2 may still be suitable, but is invaded by V_3 if B has positive slope.)

Case 2: $\lambda_D + \lambda_E < \lambda_A < \lambda_C + \lambda_D + \lambda_E$.

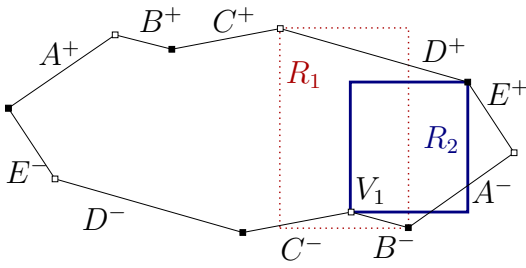
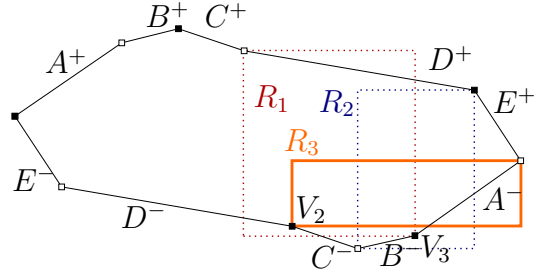
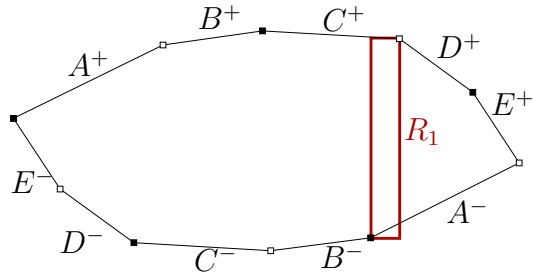
- Again R_1 can have no invasion.

Case 3: $\lambda_C + \lambda_D + \lambda_E < \lambda_A < \lambda_B + \lambda_C + \lambda_D + \lambda_E$.

- Subcase (a): If C has negative slope, the vertex V_4 at the left endpoint of C^+ will be above R_1 , so R_1 has no invasion.
- Subcase (b), supposing C has positive slope: R_4 is free of invasion if the vertex



(a) Case 1(a)

(b) Case 1(b), V_1 invasion of R_1 (c) Case 1(b), V_2 invasion of R_1 

(d) Case 2

Figure 4.34: Representative diagrams for $\mathcal{H}(1, 1)$, cases 1 ($\lambda_E < \lambda_A < \lambda_D + \lambda_E$) and 2 ($\lambda_D + \lambda_E < \lambda_A < \lambda_C + \lambda_D + \lambda_E$).

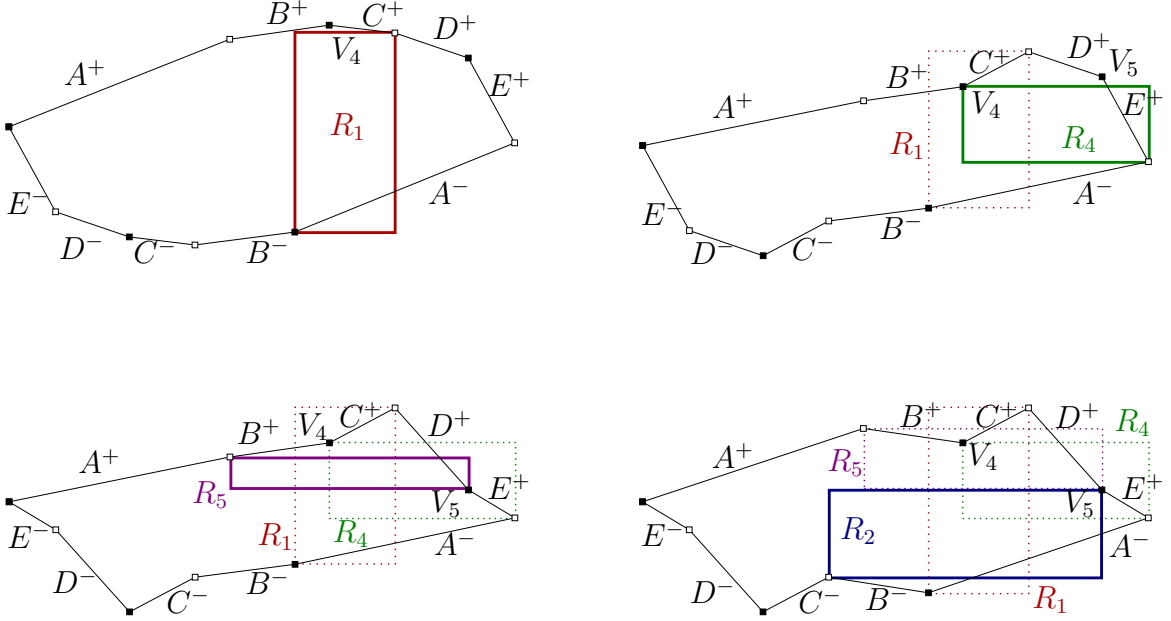


Figure 4.35: Representative diagrams for $\mathcal{H}^o(1,1)$, case 3 ($\lambda_C + \lambda_D + \lambda_E < \lambda_A < \lambda_B + \lambda_C + \lambda_D + \lambda_E$).

V_5 at the right end of D^+ stays above V_4 .

- Subcase (c): If V_5 drops lower than V_4 (thereby invading R_4), it does not invade R_5 . Further, if B has positive slope then V_4 does not invade.
- Subcase (d), supposing V_5 is lower than V_4 and B has negative slope: Here no vertex can invade R_2 .

Now, if $\mathcal{D}$ has horizontal saddle connections, the same 5 rectangles will still work.

However, as in the argument for $\mathcal{H}^o(2)$, one or both the L-cuts formed from the boundary of R and $\eta(R)$ may be singular in this case. $\square$

This ends the portion of the proof for M having no vertical or horizontal saddle connections. If M has horizontal saddle connections, the same rectangles can be found

exactly as above, noting that some resulting L-cuts may be singular. In particular, points where cases are distinguished by whether one vertex is higher than another may merge, giving a rectangle with a vertex at a midpoint of one or more edges.

What remains to be considered then are the cases with vertical saddle connections.

$$\underline{M \in \mathcal{H}_{\text{VM}}^*(2)}$$

If $M \in \mathcal{H}_{\text{VM}}^*(2)$, the vertical flow on M has one or two vertical saddle connections which are fixed by η . (As seen in the previous section, these correspond to the hyperelliptic measured ribbon graphs labeled M2 and M1, respectively.) As with the $\mathcal{H}^\circ(2)$ case, we will consider an octagon $\mathcal{O}$ representing M which has combinatorial datum $\begin{bmatrix} ABCD \\ DCBA \end{bmatrix}$. Since $\mathcal{O}$ has a Weierstrass point at the center of each of its edges, we can realize all $\mathcal{H}_{\text{VM}}^*(2)$ cases by considering octagons $\mathcal{O}$ with one or two vertical (pairs of) edges.

We will denote each case by indicating which edges of $\mathcal{O}$ are vertical, so the 10 cases are: $A, B, C, D, AB, AC, AD, BC, BD, CD$. We will say a vertical edge goes “up” or “down” according to how adjacent edges connect to it. For instance, if B is vertical and the vertex joining A and B is lower than the vertex joining B and C , we say B goes up.

We consider three candidate rectangles: R_1 with diagonal homotopic to $A + B$, R_2 with diagonal homotopic to $C + D$, and R_3 with diagonal homotopic to $B + C$.

- In cases A, AC , and AD , A is vertical and B is not, so R_1 is free of invasion.

See Figure 4.36a.

- In cases D and BD (and AD again), D is vertical and C is not, so R_2 is free of invasion. See Figure 4.36b.
- In cases B and BC , where B goes down (so C also goes down in case BC), R_1 is free of invasion. See Figure 4.36c.
- In cases C and BC , where C goes up (so B also goes up in case BC), R_2 is free of invasion. See Figure 4.36d.
- In cases B (up) and AB , B is vertical (and goes up) and C is not. In cases C (down) and CD , C is vertical (and goes down), and B is not. So, in all these cases, R_3 is free of invasion. See Figure 4.36e.

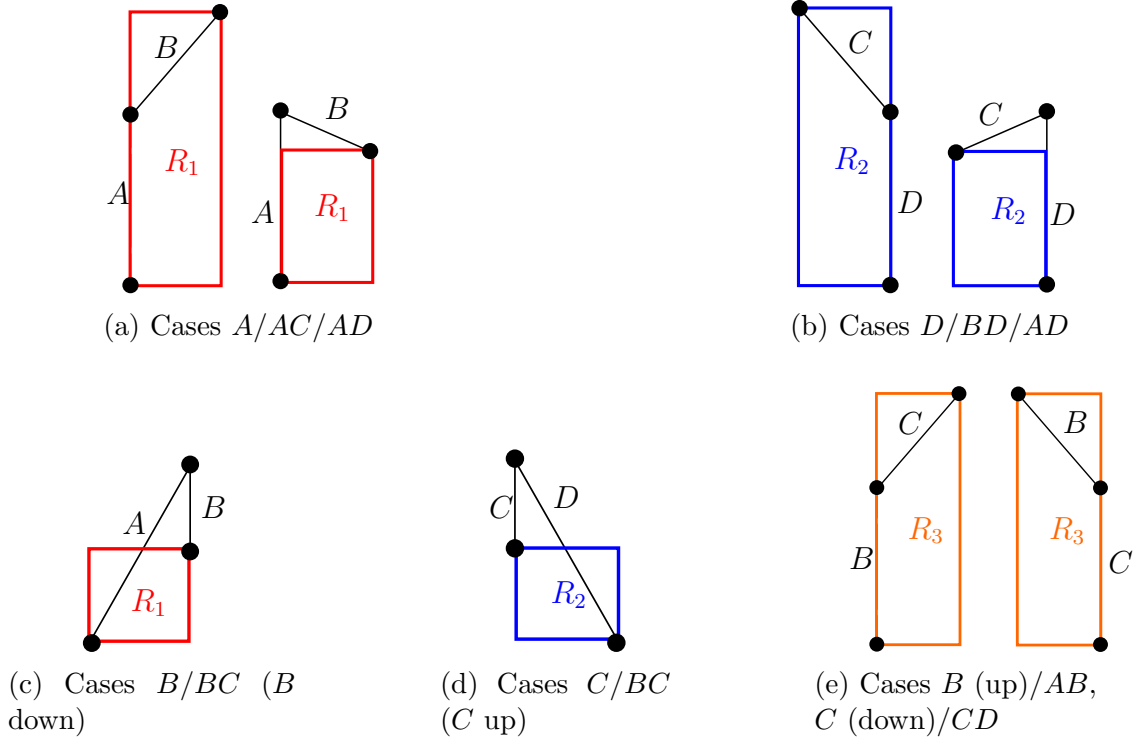


Figure 4.36: Representative diagrams for $\mathcal{H}_{\text{VM}}^*(2)$.

$$\underline{M \in \mathcal{H}_{\text{VM}}^*(1,1)}$$

Finally, consider $M \in \mathcal{H}_{\text{VM}}^*(1,1)$. The vertical flow on such M has one, two, or three vertical saddle connections which are fixed by η , corresponding to hyperelliptic measured ribbon graphs M1a, M2a/b, and M1b, respectively. Thus, we consider a decagon $\mathcal{D}$ which has combinatorial datum $\begin{bmatrix} ABCDE \\ EDCBA \end{bmatrix}$ and one to three pairs of vertical edges. There are 25 such cases to consider: 5 with one pair of vertical edges, 10 with two, and 10 with three. To be explicit, they are $A, B, C, D, E, AB, AC, AD, AE, BC, BD, BE, CD, CE, DE, ABC, ABD, ABE, ACD, ACE, ADE, BCD, BCE, BDE$, and CDE .

The same five candidate rectangles from the $\mathcal{H}^\circ(1,1)$ case will be considered: R_1 , with a diagonal homotopic to $D + E + A$, R_2 , with a diagonal homotopic to $E + A + B$, R_3 , with a diagonal homotopic to $A + B + C$, R_4 , with a diagonal homotopic to $C + D + E$, and R_5 , with a diagonal homotopic to $B + C + D$. Consult again Figure 4.33 on page 99, keeping in mind that the decagon $\mathcal{D}$ in those depictions has no vertical edges.

Let us begin by handling a few special cases, after which we can restore some simplifying assumptions from the $\mathcal{H}^\circ(1,1)$ argument.

- In cases AE, ABE, ACE , and ADE , both A and E are vertical. For AE, ABE , and ACE , no vertex can invade R_1 . For ADE (and also AE and ACE), no vertex can invade R_2 . See Figure 4.37a.
- In case BCD , if these edges all go down, R_3 is free of invasion, and if these

edges all go up, R_4 is free of invasion. See Figure 4.37b.

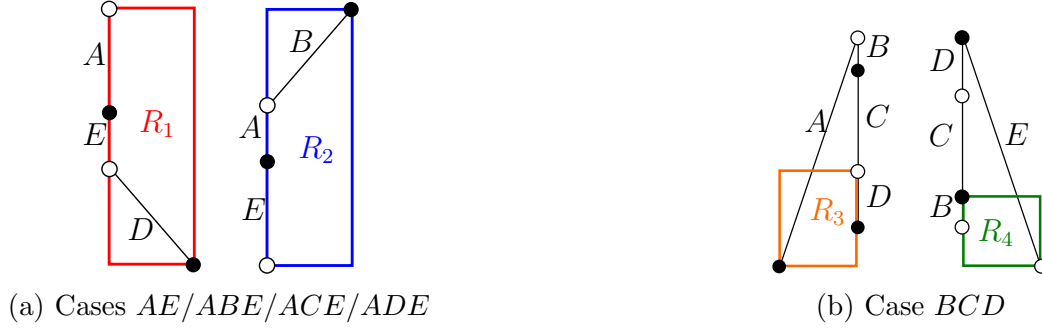


Figure 4.37: Representative diagrams for $\mathcal{H}_{\text{VM}}^*(1, 1)$, special cases.

Now, having handled all cases with $\lambda_A = \lambda_E = 0$, for the remaining cases we can assume as in the $\mathcal{H}^\circ(1, 1)$ argument that $\lambda_E < \lambda_A$, so if exactly one of A and E is vertical, it can be taken to be E . With this assumption in place, we eliminate cases A, AB, AC, AD, ABC, ABD , and ACD , reducing the remaining number of cases from 20 to 13.

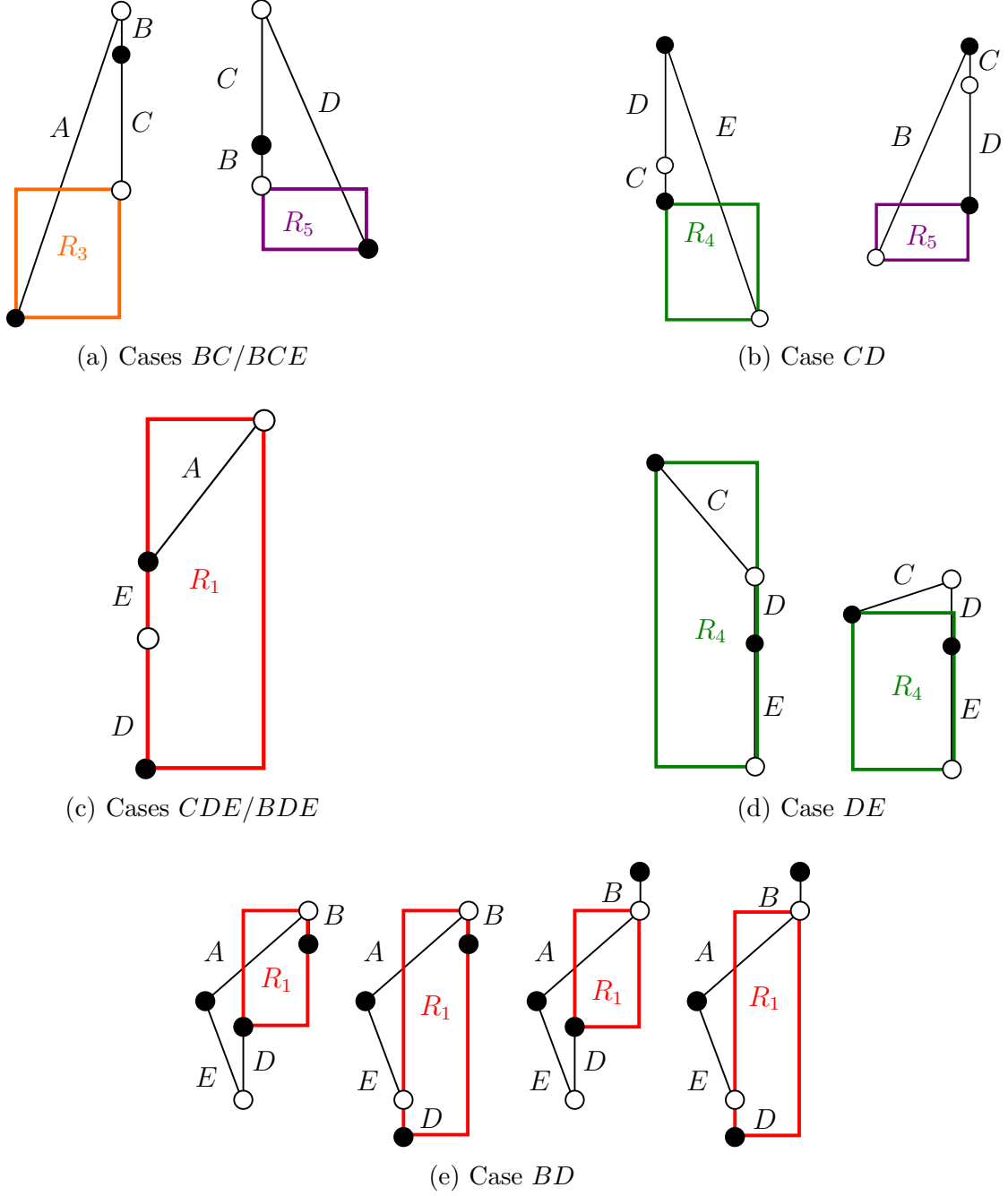
Additionally, we can do the appropriate cycle of Rauzy operations if necessary in order to keep the assumption from the $\mathcal{H}^\circ(1, 1)$ case that $\lambda_A < \lambda_E + \lambda_D + \lambda_C + \lambda_B$. (Note that this does not work for any of the $A\dots E$ cases as Rauzy operations are not defined on those polygons, and for case BCD , the assumption would be incompatible with the earlier assumption that $\lambda_E < \lambda_A$, as $\lambda_B = \lambda_C = \lambda_D = 0$.)

So, for all remaining cases we assume that $\lambda_E < \lambda_A < \lambda_E + \lambda_D + \lambda_C + \lambda_B$.

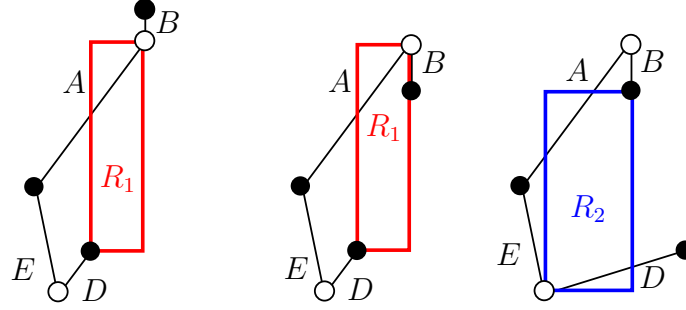
- In cases BC and BCE , if B and C go down, R_3 is free of invasion, and if B and C go up, R_5 is free of invasion. See Figure 4.38a.
- In case CD , if C and D go up, R_4 is free of invasion, and if C and D go down,

R_5 is free of invasion. See Figure 4.38b.

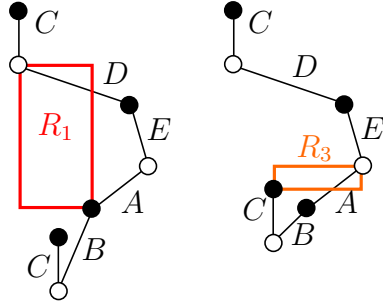
- In cases BDE and CDE , since $\lambda_D = \lambda_E = 0$, the assumption that $\lambda_A < \lambda_B + \lambda_C + \lambda_D + \lambda_E$ requires only that $\lambda_A < \lambda_B + \lambda_C$. For cases BDE and CDE , one of λ_B and λ_C is zero, and λ_A is shorter than the other. So, R_1 is free of invasion. See Figure 4.38c.
- In case DE , R_4 is free of invasion. See Figure 4.38d.
- In case BD , R_1 is always free of invasion. (Note that $\lambda_A < \lambda_E + \lambda_C$ since $\lambda_B = \lambda_D = 0$.) See Figure 4.38e.
- In cases B and BE , if B goes up, then R_1 must be free of invasion. (Again, note that $\lambda_A < \lambda_E + \lambda_D + \lambda_C$ since $\lambda_B = 0$.) If B goes down, then $\lambda_E + \lambda_D$ must be shorter than λ_A for this to be the case. However, if $\lambda_E + \lambda_D$ is longer than λ_A , R_2 is clear. See Figure 4.39a. (Case B is depicted in the figures, but E being vertical does not change anything except making the L-cut in the rightmost picture singular.)
- In cases C and CE :
 - If C goes down, then R_1 is free of invasion except possibly by the vertex at the top of C^- . If this happens, R_3 is free of invasion. See Figure 4.39b.
 - If C goes up, then R_4 is free of invasion except possibly by the vertex at the left end of E^+ . If this happens, one of R_2 and R_5 is free of invasion, according to whether B has negative or positive slope, respectively. See Figure 4.39c.

Figure 4.38: Representative diagrams for $\mathcal{H}_{\text{VM}}^*(1, 1)$, second batch.

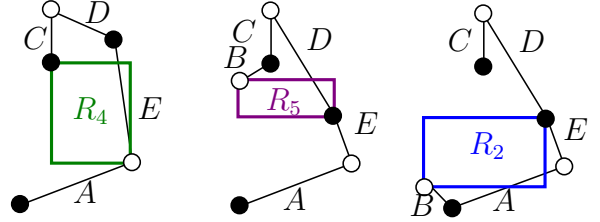
(Again, case C is depicted in the figures, but E being vertical only affects whether L-cuts are singular.)



(a) Cases B/BE



(b) Cases C/CE (C down)



(c) Cases C/CE (C up)

Figure 4.39: Representative diagrams for $\mathcal{H}_{\text{VM}}^*(1,1)$, third batch.

- In case D , R_1 may be free of invasion except possibly by the vertex at the left end of C^+ . If this happens, R_4 may be free of invasion, which is always the case when D goes up. When D goes down, R_4 may be invaded by the vertex at the bottom of D . If this invasion happens, one of R_2 and R_5 is free of invasion, according to whether B has negative or positive slope, respectively. See Figure 4.40.
- Finally, in case E , the argument presented for $M \in \mathcal{H}^\circ(1,1)$ can be reproduced

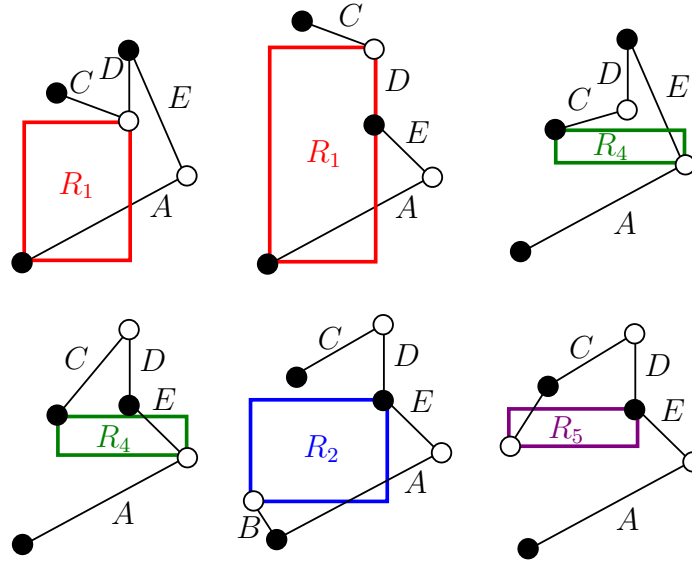


Figure 4.40: Representative diagrams for $\mathcal{H}_{\text{VM}}^*(1, 1)$, case D .

with E being vertical.

It has thus been established that if M is any genus two surface with no vertical saddle connections, it will exhibit a rectangle R satisfying the specified properties. The desired parallel pair of L-cuts (K, K') is then taken from two corresponding sides of R and $\eta(R)$, respectively. This completes the proof of Theorem 4.7. $\square$

CHAPTER 5

SMALE'S QUESTION AND THE FRANKS MAP

Smale's Question

We have at last arrived at one of the primary applications of Theorem 4.7. Smale posed a question about whether a hyperbolic toral automorphism can have a compact invariant set of dimension one. Smale's question was originally published by Hirsch in [4] in 1970. By the end of the decade, a few significant results had been established. Franks had proved that if a compact set invariant under a hyperbolic toral automorphism f_A contains a C^2 arc, then it must contain a subtorus (of dimension ≥ 2) which is invariant under some power of f_A [33]. Mañé extended this result to Lipschitz arcs [34]. Further, Mañé showed in [7] that an invariant C^1 submanifold must be a union of subtori.

More recently [8, 10, 12], the topic has taken the shape of studying whether a hyperbolic toral automorphism can have an invariant submanifold which is not a finite union of subtori. Before proceeding, we provide more background conducive to studying recent progress on Smale's question.

Hyperbolic Toral Automorphisms

If A is an $N \times N$ matrix with integer entries and $\det(A) = \pm 1$, then A is an automorphism of the integer lattice $\mathbb{Z}^N$. As the N -dimensional torus $\mathbb{T}^N$ can be

constructed by taking the quotient of $\mathbb{R}^N$ by $\mathbb{Z}^N$, A also induces an automorphism f_A on $\mathbb{T}^N$.

Definition 5.1. *If A has no eigenvalues with modulus one, the toral automorphism $f_A : \mathbb{T}^N \rightarrow \mathbb{T}^N$ given by A is called hyperbolic.*

In this case, $\mathbb{R}^N$ can be written as the product of spaces $E^u \times E^s$, where each of $E^{u/s}$ is invariant under A , and $A|_{E^u}$ is *expanding*, i.e. all its eigenvalues have modulus greater than one, and $A|_{E^s}$ is *contracting*, i.e. all its eigenvalues have modulus less than one. This decomposition leads to a natural comparison with *pseudo-Anosov* maps, described in the following section.

Pseudo-Anosov Maps

Definition 5.2. *A diffeomorphism f of a compact surface M is called pseudo-Anosov if and only if there are a transverse pair of measured foliations W_f^s and W_f^u on M and a real number $\lambda > 1$ (called the dilation constant of f) such that the foliations are preserved by f and their transverse measures μ_s and μ_u are multiplied by $1/\lambda$ and λ , respectively.*

Full contextual details of this definition can be seen in [13] and [35].

Here, we will point out that if M has genus ≥ 2 , the foliations must have (finitely many) singular points. This is a consequence of the Euler-Poincaré formula (see [35]), itself an immediate result of the Poincaré-Hopf formula.

Any pseudo-Anosov map $f : M \rightarrow M$ with orientable foliations $W^{s/u}f$ can be conjugated to a map of a translation surface such that W_f^s and W_f^u correspond to horizontal and vertical foliations, respectively. Specifically, one can use the transverse measures μ_s and μ_u to construct an atlas of charts on $M \setminus Z(\omega)$ so that the transition maps are translations and M becomes a translation surface. (Indeed, [17] lists this a definition of a translation surface which is equivalent to our earlier geometric and complex-analytic definitions seen in Chapter 2.) The chart on a neighborhood of a nonsingular point $p_0 \in M$ is given on a p near p_0 by $\phi(p) = \left(\int_{\gamma_s} d\mu_s, \int_{\gamma_u} d\mu_u \right) \in \mathbb{R}^2$, where γ_s and γ_u are transverse paths connecting the local $W^{s/u}$ leaves of p and p_0 . Further details may be found in [35].

For f with orientable foliations, the condition in Definition 5.2 on the invariance of foliations and behavior of transverse measure amounts to $f_*(\omega^s) = (1/\lambda)\omega^s$ and $f_*(\omega^u) = \lambda\omega^u$, where $\omega^s = dx$ and $\omega^u = dy$. This simply means that f contracts the leaves of W_f^s by a factor of λ and expands the leaves of W_f^u by λ (see [36]).

Singular points of the foliations correspond to the cone singularities of M (which are also the zeroes of ω). For $M \in \mathcal{H}(2)$, the sole singular point z_0 appears in the foliations as a saddle with six prongs: 3 stable and 3 unstable. For $M \in \mathcal{H}(1,1)$ there are two singular points z_0 and z_1 . Each of these is a four-pronged saddle in the foliations, having 2 stable and 2 unstable prongs. In either case, the leaves of the foliations $W_f^{s/u}$ are the stable/unstable sets for f .

From this point on, we assume that M is identified with a translation surface

$\mathcal{P}/\sim$ in such a way. In particular, we have a holomorphic one-form $\omega = \omega^s + i\omega^u$ on M that corresponds to $dz = dx + idy$ on $\mathcal{P}$. The following additional facts about pseudo-Anosov f are found in [36].

1. The topological entropy of f , $h_{top}(f)$ is given by $h_{top}(f) = \ln \lambda$.
2. For f with orientable foliations, the spectral radius of the induced automorphism f_* on $H_1(M, \mathbb{Z})$ is also $\rho(f_*) = \ln \lambda$.
3. Every leaf of $W_f^{s/u}$ is dense.

We also mention a related theorem, which was presented without proof in [37].

Theorem 5.3 (Zorich). *An orientable measured foliation on a closed Riemann surface is a horizontal foliation of a holomorphic differential in some complex structure if and only if any cycle obtained as a union of closed paths following in the positive direction a sequence of saddle connections is not homologous to zero.*

By item 3 above, neither W_f^s nor W_f^u have saddle connections (and thus satisfy the hypotheses of Theorem 5.3).

Application to Smale's Question

Fathi observed in [8] that a pseudo-Anosov map f with orientable foliations and dilation constant λ which has no conjugates over $\mathbb{Q}$ in the unit circle continuously factors onto an invariant subset of a hyperbolic toral automorphism. He did this by instantiating a map constructed by Franks in [9], which has the property that the

induced homomorphism on homology is surjective. Fathi showed that this map is locally injective on the complement of a finite set of points (the singularities of the foliations).

Barge and Kwapisz further proved in [10] that the Franks map is globally injective on a full-measure set unless f is a ramified covering of another (smaller) pseudo-Anosov map, which then has this property. Moreover, if the action of the original map f is hyperbolic on homology, then it almost everywhere embeds into a hyperbolic toral automorphism. We will return to the results of Fathi and Franks, and Barge and Kwapisz, on page 127, after which we will show that any hypothetical embedding of a pseudo-Anosov map with orientable foliations into a hyperbolic toral automorphism is equivalent to the Franks map.

More recently, Band considered a class of pseudo-Anosov maps with orientable foliations on surfaces of arbitrarily high genus but with only one singular point and specific combinatorics (akin to our central node) [12]. In particular, the Franks map cannot be locally injective at the singular point. (Note that the foliations have one singular point precisely when T is an interval exchange on an even number of intervals.)

In this dissertation, we show that for M of genus two, there is no embedding of a pseudo-Anosov map on M with orientable foliations into a hyperbolic toral automorphism.

ShadowingThe Homology Cover $\hat{M}$

Given a Riemann surface M of genus g , let $\tilde{M}$ be the universal cover. The fundamental group $\pi_1(M)$ naturally acts on $\tilde{M}$ by deck transformations. Let $K = [\pi_1(M), \pi_1(M)]$ be the commutator subgroup of $\pi_1(M)$.

Definition 5.4. *The homology cover of M is given by $\hat{M} = \tilde{M}/K$.*

The universal covering map $\tilde{\pi} : \tilde{M} \rightarrow M$ induces $\pi : \hat{M} \rightarrow M$, which is a regular covering with commutative deck group $\pi_1(M)/K$ isomorphic to the first homology group $H_1(M, \mathbb{Z})$ with integer coefficients. (See [38].) For the action of H_1 on $\hat{M}$, we will use additive notation, i.e., the result of $w \in H_1(M, \mathbb{Z}) \cong \mathbb{Z}^{2g}$ applied to $\hat{x} \in \hat{M}$ is denoted $\hat{x} + w$. Recall that $\hat{x} + w$ is, by definition, $\hat{\gamma}(1)$, where $\hat{\gamma} : [0, 1] \rightarrow \hat{M}$ is a lift of a loop γ in M passing through $\pi(\hat{x})$ and representing the homology class w , with $\hat{\gamma}(0) = \hat{x}$.

$\hat{M}$ is an unbounded surface with $\pi_1(\hat{M}) = K$. To better grasp it we note that it can be realized as a $\mathbb{Z}^{2g}$ -periodic surface embedded in $\mathbb{R}^{2g}$ via the Abel-Jacobi map (see, e.g., [21]). Additionally, a key property which we shall use is that a closed smooth form ω on M lifts to an exact form $\hat{\omega}$ on $\hat{M}$, i.e., there is a smooth function $\hat{\varphi} : \hat{M} \rightarrow \mathbb{R}$ such that $d\hat{\varphi} = \hat{\omega}$.

Global Shadowing

Given a homeomorphism h on a space X with metric d_X and point a , local stable

and unstable sets for a with respect to h are defined by fixing a sufficiently small $\varepsilon > 0$ and setting

$$W_h^s(a, \varepsilon) := \{x \in X : d_X(h^n(a), h^n(x)) < \varepsilon \ \forall n \geq 0\} \quad (5.1)$$

and

$$W_h^u(a, \varepsilon) := \{x \in X : d_X(h^n(a), h^n(x)) < \varepsilon \ \forall n \leq 0\}. \quad (5.2)$$

Then, global stable and unstable sets are defined in terms of local stable and unstable manifolds, i.e.

$$W_h^s(a) := \bigcup_{n \geq 0} h^n(W_h^s(a, \varepsilon)) \quad (5.3)$$

and

$$W_h^u(a) := \bigcup_{n \leq 0} h^n(W_h^u(a, \varepsilon)). \quad (5.4)$$

Generally, these depend on ε but they do not (once $\varepsilon > 0$ is small) for the h we discuss. Now, recall the general definition of a heteroclinic point:

Definition 5.5. *A point $p \in X$ is called heteroclinic with respect to a homeomorphism $h : X \rightarrow X$ and points a and b of h (possibly with $a = b$, in which case p is typically called homoclinic) if and only if $p \in W_h^u(a) \cap W_h^s(b)$.*

For such p , $h^n(p)$ will approach $h^n(a)$ as $n \rightarrow -\infty$ and approach $h^n(b)$ as $n \rightarrow \infty$.

We now make a definition modeled after this one to describe a similar situation for a pair of points in the homology cover $\hat{M}$, acted on by a lift $\hat{f}$ of a pseudo-Anosov map f with orientable foliations.

Definition 5.6. Two points $x, y \in M$ are called homologically co-heteroclinic with respect to the map $f : M \rightarrow M$ and its fixed points z_0 and z_1 (possibly with $z_0 = z_1$) if and only if x, y, z_0 and z_1 can be lifted to $\hat{x}, \hat{y}, \hat{z}_0, \hat{z}_1 \in \hat{M}$ (so $\pi(\hat{x}) = x$, $\pi(\hat{y}) = y$, $\pi(\hat{z}_0) = z_0$, and $\pi(\hat{z}_1) = z_1$) in such a way that both $\hat{x}$ and $\hat{y}$ lie in $W_{\hat{f}}^u(\hat{z}_0) \cap W_{\hat{f}}^s(\hat{z}_1)$.

We note that even though z_0 and z_1 are fixed by f , the lifts $\hat{z}_0$ and $\hat{z}_1$ may fail to be fixed points of $\hat{f}$.

If x and y lie at the turning points of two parallel L-cuts K and K' , respectively (as in Figure 5.1), they will be homologically co-heteroclinic with respect to f . Indeed, lifting K , starting from some fixed lift $\hat{z}_0$ of z , yields the union of a segment of $W_{\hat{f}}^u$ followed by a segment of $W_{\hat{f}}^s$, the latter ending at a lift of z_1 , say $\hat{z}_1$. Because K and K' are homologous, lifting K' starting at $\hat{z}_0$ yields an analogous union of segments of $W_{\hat{f}}^u$ and $W_{\hat{f}}^s$ that ends at the same $\hat{z}_1$. In Figure 5.1, the points x and y lying at the turning points of the L-cuts K and K' lift to points $\hat{x}$ and $\hat{y}$ in $\hat{M}$, both of which lie in $W_{\hat{f}}^u(\hat{z}_0) \cap W_{\hat{f}}^s(\hat{z}_1)$.

When $x, y \in M$ are homologically co-heteroclinic, the lifts $\hat{x}, \hat{y} \in \hat{M}$ can be described as $\hat{f}$ -shadowing each other, in the sense that

$$\sup_{n \in \mathbb{Z}} d_{\hat{M}}(\hat{f}^n(\hat{x}), \hat{f}^n(\hat{y})) < +\infty. \quad (5.5)$$

Indeed, observe in Figure 5.2 that since $\hat{x}, \hat{y} \in W^u(\hat{z}_0)$,

$$d_{\hat{M}}(\hat{f}^n(\hat{x}), \hat{f}^n(\hat{y})) \leq d_{\hat{M}}(\hat{f}^n(\hat{x}), \hat{f}^n(\hat{z}_0)) + d_{\hat{M}}(\hat{f}^n(\hat{z}_0), \hat{f}^n(\hat{y})) \rightarrow 0 \text{ as } n \rightarrow -\infty, \quad (5.6)$$

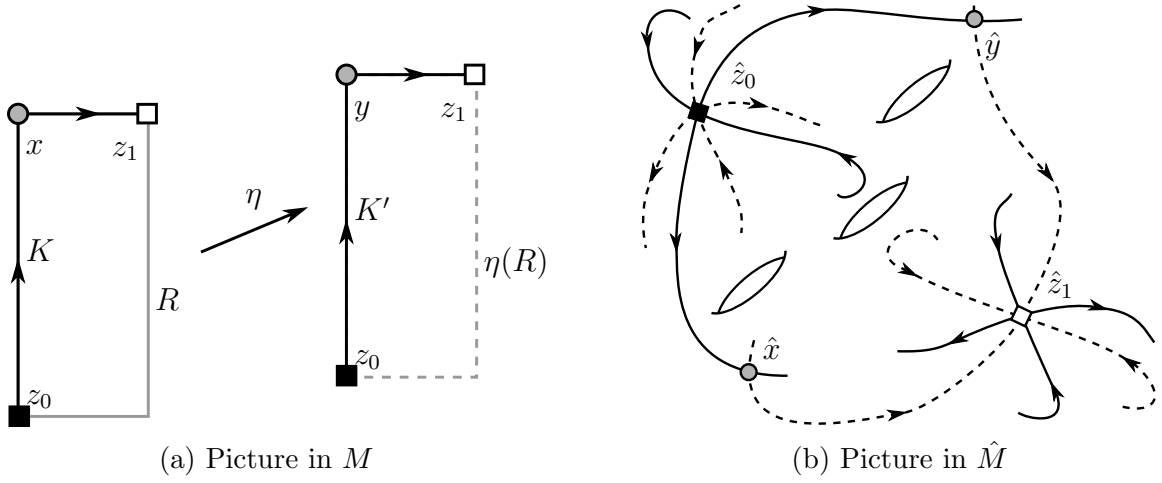


Figure 5.1: Homologically co-heteroclinic points $x, y \in M$ lift to heteroclinic points $\hat{x}, \hat{y} \in \hat{M}$. On the right, a portion of $\hat{M}$ including neighborhoods of conical singularities $\hat{z}_0, \hat{z}_1$ each with angle 4π (here, vertical is represented with solid lines and horizontal with dashed lines) is rendered schematically (meaning, the neighborhoods of $\hat{z}_0$ and $\hat{z}_1$ may be far apart, with much of the “structure” of $\hat{M}$ lying between them).

and since $\hat{x}, \hat{y} \in W^s(\hat{z}_1)$,

$$d_{\hat{M}}(\hat{f}^n(\hat{x}), \hat{f}^n(\hat{y})) \leq d_{\hat{M}}(\hat{f}^n(\hat{x}), \hat{f}^n(\hat{z}_1)) + d_{\hat{M}}(\hat{f}^n(\hat{z}_1), \hat{f}^n(\hat{y})) \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (5.7)$$

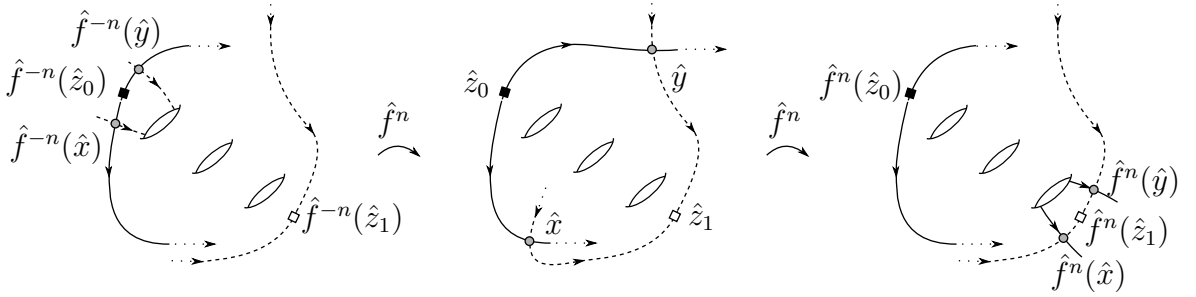


Figure 5.2: Homologically co-heteroclinic points x, y are such that $\hat{x}, \hat{y}$ shadow each other in $\hat{M}$.

Definition 5.7. Given a homeomorphism $f : M \rightarrow M$, a continuous map $\psi : M \rightarrow \mathbb{T}^N$ is said to be homologically f -shadowing adapted if, for all $x, y \in M$, $\psi(x) = \psi(y)$

if and only if there are lifts $\hat{x}, \hat{y} \in \hat{M}$ of x, y (i.e. $\pi(\hat{x}) = x$ and $\pi(\hat{y}) = y$) such that

$$\sup_{n \in \mathbb{Z}} |\hat{\psi}(\hat{f}^n(\hat{x})) - \hat{\psi}(\hat{f}^n(\hat{y}))| < +\infty, \quad (5.8)$$

where $\hat{\psi} : \hat{M} \rightarrow \mathbb{R}^N$ and $\hat{f} : \hat{M} \rightarrow \hat{M}$ are lifts of ψ and f , respectively.

We verify that the supremum in Definition 5.7 does not depend on the choice of lifts of ψ and f . First, note that, being continuous maps, ψ and f induce homomorphisms $\psi_* : H_1(M, \mathbb{Z}) \rightarrow H_1(\mathbb{T}^N, \mathbb{Z}) \cong \mathbb{Z}^N$ and $f_* : H_1(M, \mathbb{Z}) \rightarrow H_1(M, \mathbb{Z})$ satisfying $\hat{\psi}(\hat{x} + v) = \hat{\psi}(\hat{x}) + \psi_*(v)$ and $\hat{f}(\hat{x} + v) = \hat{f}(\hat{x}) + f_*(v)$, for all $\hat{x} \in \hat{M}, v \in H_1(M, \mathbb{Z})$. Let $\hat{f}_1 = \hat{f} + v$, for some $v \in H_1(M, \mathbb{Z})$, so $\hat{f}_1^n = (\hat{f} + v)^n = \hat{f}^n + v + f_*(v) + \dots + f_*^{n-1}(v) =: \hat{f}^n + w$. Also, let $\hat{\psi}_1 = \hat{\psi} + u$ for some $u \in \mathbb{Z}^N$. Then $\hat{\psi}_1 \circ \hat{f}_1^n = (\hat{\psi} + u)(\hat{f}^n + w) = \hat{\psi} \circ \hat{f}^n + \psi_*(w) + u$. So, $|\hat{\psi}_1(\hat{f}_1^n(\hat{x})) - \hat{\psi}_1(\hat{f}_1^n(\hat{y}))| = |\hat{\psi}(\hat{f}^n(\hat{x})) - \hat{\psi}(\hat{f}^n(\hat{y}))|$.

Note also that Definition 5.7 defines an equivalence relation on the points of M . Symmetry and reflexivity are self-evident, and transitivity follows from the triangle inequality. Indeed, suppose for points $\hat{x}, \hat{y}$, and $\hat{z}$ in $\hat{M}$ and $v \in H_1(M, \mathbb{Z})$, that for all $n \in \mathbb{Z}$ there are $C_1, C_2 \in \mathbb{R}$ such that $|\hat{\psi}(\hat{f}^n(\hat{x})) - \hat{\psi}(\hat{f}^n(\hat{y}))| < C_1$ and $|\hat{\psi}(\hat{f}^n(\hat{y} + v)) - \hat{\psi}(\hat{f}^n(\hat{z}))| < C_2$. Then

$$\begin{aligned} |\hat{\psi}(\hat{f}^n(\hat{x})) - \hat{\psi}(\hat{f}^n(\hat{z} - v))| &\leq |\hat{\psi}(\hat{f}^n(\hat{x})) - \hat{\psi}(\hat{f}^n(\hat{y}))| + |\hat{\psi}(\hat{f}^n(\hat{y})) - \hat{\psi}(\hat{f}^n(\hat{z} - v))| \\ &\leq |\hat{\psi}(\hat{f}^n(\hat{x})) - \hat{\psi}(\hat{f}^n(\hat{y}))| + |\hat{\psi}(\hat{f}^n(\hat{y} + v)) - \hat{\psi}(\hat{f}^n(\hat{z}))| \\ &< C_1 + C_2, \end{aligned}$$

where we used translation invariance of the metric $|\cdot|$ in $\mathbb{R}^d$ and the fact that

$$\hat{\psi} \circ \hat{f}^n(\cdot + v) = \hat{\psi} \circ \hat{f}^n(\cdot) + \psi_* \circ f_*^n. \quad (5.9)$$

Finally, observe that uniform continuity of ψ means that for a pair of homologically co-heteroclinic points x and y , the boundedness given by (5.5) guarantees that (5.8) is also satisfied.

Non-Injectivity of Pseudo-Anosov Embeddings

Passing to $\tilde{M}$

Suppose $\psi : M \rightarrow \mathbb{T}^N$ is a continuous map satisfying

$$\psi \circ f = f_A \circ \psi, \tag{5.10}$$

where $f : M \rightarrow M$ is a homeomorphism and $f_A : \mathbb{T}^N \rightarrow \mathbb{T}^N$ is a hyperbolic toral automorphism.

Recall that $\pi : \hat{M} \rightarrow M$ is the covering map from the homology cover $\hat{M}$. Also, viewing $\mathbb{T}^N$ as $\mathbb{R}^N/\mathbb{Z}^N$, denote by π_N the associated covering map, $\pi_N : \mathbb{R}^N \rightarrow \mathbb{T}^N$.

Lift $\psi : M \rightarrow \mathbb{T}^N$ to $\hat{\psi} : \hat{M} \rightarrow \mathbb{R}^N$ and $f : M \rightarrow M$ to $\hat{f} : \hat{M} \rightarrow \hat{M}$. Note that these lifts are unique only up to deck transformations. Thus, by establishing the following claim, we will take measures to ensure that the commutation given by (5.10) persists on the level of these lifts.

Lemma 5.8. *After replacing f and f_A by their iterates f^n and f_A^n , if necessary, we can ensure that, for a suitable choice of the lifts $\hat{\psi}$ and $\hat{f}$,*

$$A \circ \hat{\psi} = \hat{\psi} \circ \hat{f}. \tag{5.11}$$

Proof. Certainly from (5.10), we can say that, for any $\hat{x} \in \hat{M}$, $A \circ \hat{\psi}(\hat{x}) = \hat{\psi} \circ \hat{f}(\hat{x}) + u(\hat{x})$

for some $u(\hat{x}) \in H_1(M, \mathbb{Z})$. But, since $\hat{x} \mapsto u(\hat{x})$ is continuous, it must be constant, owing to the discreteness of $H_1(M, \mathbb{Z})$. Let us see that by adjusting $\hat{\psi}$ and $\hat{f}$, we can pick $u = 0$. Begin by identifying a fixed point p_0 of f , and let $\hat{p}_0$ be a lift of p_0 .

First, let us arrange that $\hat{\psi}(\hat{p}_0) = 0$. Indeed, $\hat{\psi}(\hat{p}_0) =: \hat{x}_0$ is a π_N -preimage of a fixed point of f_A , so $A\hat{x}_0 = \hat{x}_0 + v$ for some $v \in \mathbb{Z}^N$. Conjugating A by the translation $T_s := \hat{x} \mapsto \hat{x} + s$, we get $A_{(new)} := T_s^{-1} \circ A \circ T_s : \hat{x}_0 \mapsto A(\hat{x}_0 + s) - s = \hat{x}_0 + v + As - s$. Picking $s = (I - A)^{-1}v$ so that $v + As - s = 0$, ensures $A_{(new)}\hat{x}_0 = \hat{x}_0$. (We used here that A is hyperbolic so $I - A$ is invertible.) By translating the origin of $\mathbb{R}^N$ to this $\hat{x}_0$, we get $\hat{\psi}(\hat{p}_0) = 0$.

We are ready to adjust $\hat{f}$. Since $f(p_0) = p_0$, we have $\hat{f}(\hat{p}_0) = \hat{p}_0 + w$ for some $w \in H_1(M, \mathbb{Z})$. Upon replacing $\hat{f}$ by $T_{-w} \circ \hat{f} = \hat{f} - w$, we get $\hat{f}(\hat{p}_0) = \hat{p}_0$. With these choices we are assured that $u(\hat{p}_0) = 0$ and thus $u(\hat{x}) = 0$ for all $\hat{x} \in \hat{M}$. $\square$

From now on, we assume that (5.11) holds.

Now, $\hat{\psi}$ will not necessarily be an embedding, even if ψ is. Recall the induced homomorphism $\psi_* : H_1(M, \mathbb{Z}) \rightarrow H_1(\mathbb{T}^N, \mathbb{Z}) \cong \mathbb{Z}^N$, which satisfies $\hat{\psi}(\hat{x} + v) = \hat{\psi}(\hat{x}) + \psi_*(v)$ for all $\hat{x} \in \hat{M}, v \in H_1(M, \mathbb{Z})$.

Let $\check{M} := \hat{M} / \ker \psi_*$, where $\psi_* : H_1(M, \mathbb{Z}) \rightarrow H_1(\mathbb{T}^N, \mathbb{Z}) \cong \mathbb{Z}^N$ is the induced homomorphism. For $v \in \ker \psi_*$, we get $\hat{\psi}(\hat{x} + v) = \hat{\psi}(\hat{x})$. Thus, we have a well defined quotient map $\check{\psi} : \check{M} \rightarrow \mathbb{R}^N$, $\check{\psi} := \hat{\psi} \bmod \psi_*$. Similarly, $\check{f} : \check{M} \rightarrow \check{M}$ is given by $\hat{f} \bmod \ker \psi_*$.

The following corollary of Lemma 5.8 pushes the commutation from $\hat{M}$ to $\check{M}$.

Corollary 5.9. $A \circ \check{\psi} = \check{\psi} \circ \check{f}$

Proof. This is a routine verification. Let $\bar{\pi} : \hat{M} \rightarrow \check{M}$ be the canonical projection (a regular covering with deck group $\ker \psi_*$) and for $\hat{x} \in \hat{M}$ set $\check{x} := \bar{\pi}(\hat{x})$. Projecting $(A \circ \hat{\psi})(\hat{x}) = (\hat{\psi} \circ \hat{f})(\hat{x})$ via $\bar{\pi}$, with $\hat{\psi} = \check{\psi} \circ \bar{\pi}$, we get $(A \circ \check{\psi} \circ \bar{\pi})(\hat{x}) = (\check{\psi} \circ \bar{\pi} \circ \hat{f})(\hat{x})$. Using $\bar{\pi} \circ \hat{f} = \check{f} \circ \bar{\pi}$, this becomes $(A \circ \check{\psi} \circ \bar{\pi})(\hat{x}) = (\check{\psi} \circ \check{f} \circ \bar{\pi})(\hat{x})$, that is, $(A \circ \check{\psi})(\check{x}) = (\check{\psi} \circ \check{f})(\check{x})$. $\square$

Let G be the group $H_1(M, \mathbb{Z}) / \ker \psi_*$, so $\check{M}$ is a G -covering of M . The covering map $\check{\pi} : \check{M} \rightarrow M$ satisfies the commuting diagram seen in Figure 5.3.

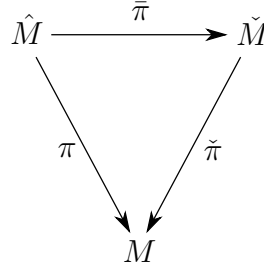


Figure 5.3: Relationship of M , $\hat{M}$, and $\check{M}$; π has deck group $H_1(M, \mathbb{Z})$, $\bar{\pi}$ has deck group $\ker \psi_*$, and $\check{\pi}$ has deck group G

By the isomorphism theorem, $\check{\psi}_* : G \rightarrow H_1(\mathbb{T}^N, \mathbb{Z})$ induced by ψ_* is injective. Thus, for all $\check{x} \in \check{M}, g \in G$, we have that $\check{\psi}(\check{x} + g) = \check{\psi}(\check{x}) + \check{\psi}_*(g)$, and $g \neq 0 \Rightarrow \check{\psi}_*(g) \neq 0$.

Lemma 5.10. *Let $\check{p}, \check{q} \in \check{M}$, and set $p := \check{\pi}(\check{p}), q := \check{\pi}(\check{q})$. If $\check{\psi}(\check{p}) = \check{\psi}(\check{q})$ and $\check{p} \neq \check{q}$, then $\psi(p) = \psi(q)$ and $p \neq q$.*

Proof. Suppose $\check{\psi}(\check{p}) = \check{\psi}(\check{q})$ and $\check{p} \neq \check{q}$. Applying π_N to both sides, via $\pi_N \circ \check{\psi} = \psi \circ \check{\pi}$,

we get $\psi(p) = \psi(\tilde{\pi}(\check{p})) = \psi(\tilde{\pi}(\check{q})) = \psi(q)$. Now, $\tilde{\pi} : \check{M} \rightarrow M$ is a covering with deck group G , so if $p = q$, then $\check{q} = \check{p} + g$ for some $g \in G$. So $\check{\psi}(\check{q}) = \check{\psi}(\check{p} + g) = \check{\psi}(\check{p}) + \bar{\psi}_*(g)$ and thus $\bar{\psi}_*(g) = 0$. Since $\bar{\psi}_*$ is injective on G , $g = 0$, contradicting $\check{p} \neq \check{q}$. $\square$

The following corollary is immediate from the above.

Corollary 5.11. *$\check{\psi} : \check{M} \rightarrow \mathbb{R}^N$ is an embedding provided ψ is an embedding.*

Pseudo-Anosov Maps with Orientable Foliations in Genus Two

We will now show that if f is an orientable pseudo-Anosov map of a genus two surface M , then f cannot be embedded into a hyperbolic toral automorphism. Note that if such an f exists, its foliations $W_f^{s/u}$ have dense leaves (fact 3 in the list on page 114, which comes from [36]). Thus, in particular, M has no vertical or horizontal saddle connections and Theorem 4.7 applies.

Theorem 5.12 (Non-Embedding). *Let M be a surface of genus two. If $\psi : M \rightarrow \mathbb{T}^N$ is a continuous map satisfying $\psi \circ f = f_A \circ \psi$, where $f : M \rightarrow M$ is a pseudo-Anosov map with orientable foliations and $f_A : \mathbb{T}^N \rightarrow \mathbb{T}^N$ is a hyperbolic toral automorphism, then ψ is not injective.*

The proof will proceed by showing that $\check{\psi} : \check{M} \rightarrow \mathbb{R}^N$ (as discussed in the previous section) cannot be injective, and so, by Corollary 5.11, neither is ψ .

Proof. By Theorem 4.7, there is a parallel pair of L-cuts K and K' on M . Label their turning points as x and y , respectively, and observe that x, y are homologically coheteroclinic. So, $\hat{x}, \hat{y}$ shadow each other (as in Figure 5.2 on page 119), and we have

that for some n_0 large, $\hat{f}^{n_0}(\hat{x})$ and $\hat{f}^{n_0}(\hat{y})$ are close enough that they are within some ball $B(f^{n_0}(\hat{z}_1))$ on which the covering map $\bar{\pi} : \hat{M} \rightarrow \check{M} := \hat{M} / \ker \psi_*$ is injective.

Consider the situation in $\check{M}$. Stable and unstable manifolds for f of z_0, z_1 lift to stable and unstable manifolds for $\check{f}$ of $\check{z}_0, \check{z}_1$. (Recall that in M these manifolds are simply level lines of the forms $\omega^s = \operatorname{Re} \omega$ and $\omega^u = \operatorname{Im} \omega$; in $\check{M}$, they are $\check{\omega}^s := \operatorname{Re} \check{\omega}$ and $\check{\omega}^u := \operatorname{Im} \check{\omega}$, where $\check{\omega}$ is the one-form on $\check{M}$ obtained from lifting ω via $\tilde{\pi} : \check{M} \rightarrow M$.)

Hence, in $\check{M}$, we get heteroclinic points $\check{x} := \bar{\pi}(\hat{x})$ and $\check{y} := \bar{\pi}(\hat{y})$ satisfying $\check{x}, \check{y} \in W^u(\check{z}_0) \cap W^s(\check{z}_1)$. As a result,

$$\sup_{n \in \mathbb{Z}} d_{\check{M}}(\check{f}^n(\check{x}), \check{f}^n(\check{y})) < +\infty, \quad (5.12)$$

where $d_{\check{M}}$ is some fixed, G -invariant metric on $\check{M}$.

Now, we see that, for the above selected n_0 , $\check{f}^{n_0}(\check{x}) \neq \check{f}^{n_0}(\check{y})$, by observing that $\check{f}^{n_0}(\check{x}) = \check{f}^{n_0} \circ \bar{\pi}(\hat{x}) = \bar{\pi} \circ \hat{f}^{n_0}(\hat{x}) \neq \bar{\pi} \circ \hat{f}^{n_0}(\hat{y}) = \check{f}^{n_0} \circ \bar{\pi}(\hat{y}) = \check{f}^{n_0}(\check{y})$, where the inequality comes from local injectivity of $\bar{\pi}$. Since $\check{f}$ is a homeomorphism, we get $\check{x} \neq \check{y}$.

Finally, applying uniform continuity of $\check{\psi}$ to (5.12), we get

$$\sup_{n \in \mathbb{Z}} |\check{\psi} \circ \check{f}^n(\check{x}) - \check{\psi} \circ \check{f}^n(\check{y})| < +\infty. \quad (5.13)$$

Applying the commutation given in Corollary 5.9 yields

$$\sup_{n \in \mathbb{Z}} |A^n \check{\psi}(\check{x}) - A^n \check{\psi}(\check{y})| < +\infty. \quad (5.14)$$

Since A is hyperbolic, $\check{\psi}(\check{x}) = \check{\psi}(\check{y})$. On the other hand, $\check{x} \neq \check{y}$. This contradicts injectivity of $\check{\psi}$, and Corollary 5.11 implies Theorem 5.12. $\square$

Characterization of Embeddings

Suppose $\psi : M \rightarrow \mathbb{T}^N$ is an embedding of a compact orientable surface M such that $\Lambda := \psi(M)$ is a compact subset invariant under a hyperbolic toral automorphism f_A on $\mathbb{T}^N$. We start with a definition and a simple observation.

Definition 5.13. *A homeomorphism $h : X \rightarrow X$ on a compact metric space X with metric d_X is called expansive if there exists a constant $\alpha > 0$ such that for $x, y \in X$, with $x \neq y$, $d_X(h^n(x), h^n(y)) > \alpha$ for some $n \in \mathbb{Z}$. The number α is called the expansiveness constant for h .*

Lemma 5.14. *The homeomorphism $g := \psi^{-1} \circ f_A \circ \psi : M \rightarrow M$ is expansive.*

Proof. It is well-known that f_A is expansive (see, e.g., [39]). If Δ is the expansiveness constant for f_A , and $\delta > 0$ is such that $d(x, y) < \delta$ implies $d(\psi(x), \psi(y)) < \Delta$, then g is δ -expansive. Indeed, if $d(g^n(p), g^n(q)) < \delta$ for all $n \in \mathbb{Z}$, then $d(f_A^n \circ \psi(p), f_A^n \circ \psi(q)) = d(\psi \circ g^n(p), \psi \circ g^n(q)) < \Delta$ for all $n \in \mathbb{Z}$. So, $\psi(p) = \psi(q)$ and thus also $p = q$ (since ψ is injective). $\square$

The following result, given independently by Hiraide in [13] and Lewowicz in [14], implies that such g must be pseudo-Anosov.

Theorem 5.15 (Hiraide, Lewowicz). *Every expansive homeomorphism of a compact surface is conjugate to a pseudo-Anosov map.*

We note that there is no guarantee that the $W^{s/u}$ foliations of g are orientable. However, as a direct consequence of Theorem 5.15, together with Theorem 5.12, we get the following:

Corollary 5.16. *Any embedding ψ of a genus two surface onto an invariant set of a hyperbolic toral automorphism f_A must be such that $g := \psi^{-1} \circ f_A \circ \psi$ is a pseudo-Anosov map with non-orientable foliations.*

It is believed that no such ψ actually exist, but we can't prove this using the same methods as for Theorem 5.12. The proof would require development analogous to our theory of parallel pairs of L-cuts and their existence for non-orientable genus two pseudo-Anosov maps. By the device of orientation covering this leads to translation surfaces of genus $g > 2$, for which combinatorial complexity is considerably greater; even for genus three, the polygons have twelve or more sides. Additionally, and more devastatingly to our approach, surfaces of genus $g \geq 3$ are no longer guaranteed to be hyperelliptic.

Franks Map

Now let M be a surface of any genus $g \geq 2$. The following theorem is the result found in [8] which was introduced on page 114 of this dissertation. It states that a pseudo-Anosov map f on M semi-conjugates onto an invariant subset of a hyperbolic toral automorphism f_A via a continuous map $h : M \rightarrow \mathbb{T}^N$ constructed by Franks.

Theorem 5.17 (Fathi-Franks). *Let M be a closed surface of genus g . Suppose that $f : M \rightarrow M$ is a pseudo-Anosov map with orientable foliations W_f^s and W_f^u , whose dilation coefficient λ has no conjugate over $\mathbb{Q}$ in the unit circle. Let Λ be any family of eigenvalues of the action $f_* : H_1(M, \mathbb{Z}) \rightarrow H_1(M, \mathbb{Z})$ induced on the first homology such that Λ contains λ and $1/\lambda$, is closed under conjugation over $\mathbb{Q}$, and avoids the unit circle. If N is the number of eigenvalues in Λ (counted with multiplicity), then there is a continuous $h : M \rightarrow \mathbb{T}^N$ that is locally injective at every non-singular point (of W_f^s and W_f^u) and such that*

$$h \circ f = f_A \circ h, \quad (5.15)$$

where $f_A : \mathbb{T}^N \rightarrow \mathbb{T}^N$ is a hyperbolic toral automorphism associated to a matrix A with eigenvalues Λ . The map $h_ : H_1(M, \mathbb{Z}) \rightarrow H_1(\mathbb{T}^N, \mathbb{Z})$ is surjective.*

Fathi showed that this map is locally one-to-one on the complement of the singularities and globally finite-to-one. Barge and Kwapisz showed that either this map is almost everywhere one-to-one, or the situation can be reduced to one in which this is the case [10].

Theorem 5.18 (Barge-Kwapisz). *In the context of Theorem 5.17, there is an $m \geq 1$ such that f factors via an m -to-one branched covering $\delta : M \rightarrow M_1$ to $f_1 : M_1 \rightarrow M_1$ that is pseudo-Anosov with orientable foliations (or Anosov if $M_1 = \mathbb{T}^2$). Moreover, h factors via δ to $h_1 : M_1 \rightarrow \mathbb{T}^N$ such that $h_1 \circ f_1 = f_A \circ h_1$ and h_1 is almost everywhere one-to-one on M_1 .*

The proof involves showing h is homologically f -shadowing adapted, as in Definition 5.7 on page 119. Barge and Kwapisz go on to point out that it follows immediately from Theorem 5.18 that, by passing to f_1 , M_1 , and h_1 if necessary, one may assume that $N = 2g$, so that h is almost everywhere one-to-one and f is almost conjugated to f_A restricted to an invariant subset.

Now, suppose $\psi : M \rightarrow \mathbb{T}^N$ is some injective map which satisfies $\psi \circ f = f_A \circ \psi$. First, note that $V := \psi_*(H_1(M, \mathbb{R}))$ is an A -invariant subspace of $\mathbb{R}^N$. It contains $\Gamma := \psi_*(H_1(M, \mathbb{Z}))$ as a lattice (in fact, $V = \text{span}_{\mathbb{R}} \Gamma$). Hence V/Γ is an f_A -invariant subtorus of $\mathbb{T}^N$.

$f_A|_{V/\Gamma}$ is again a hyperbolic toral automorphism, i.e., it can be conjugated to $f_{\tilde{A}} : \mathbb{T}^{\tilde{N}} \rightarrow \mathbb{T}^{\tilde{N}}$, where $\tilde{N} := \dim V$ and $\tilde{A}$ is an $\tilde{N} \times \tilde{N}$ integer matrix representing $A|_V$ in some integral basis of V . The following lemma will allow us to replace f_A by $f_{\tilde{A}}$ and $\mathbb{T}^N$ by $\mathbb{T}^{\tilde{N}} \cong V/\Gamma$ in subsequent consideration:

Lemma 5.19. *If $\psi : M \rightarrow \mathbb{T}^N$ is some embedding such that $\psi \circ f = f_A \circ \psi$, where f is a homeomorphism and $f_A : \mathbb{T}^N \rightarrow \mathbb{T}^N$ is a hyperbolic toral automorphism, then $\psi(M) \subseteq V/\Gamma$.*

Proof. We first establish that $\hat{\psi}(\hat{M})/V \subset \mathbb{R}^N/V$ is bounded. To do this, fix $\hat{x}_0 \in \hat{M}$ and let $\hat{x} \in \hat{M}$ be arbitrary. Let $\hat{\gamma}$ be any smooth curve in $\hat{M}$ from $\hat{x}_0$ to $\hat{x}$, with $\gamma := \pi \circ \hat{\gamma}$. Set $x := \pi(\hat{x})$ and take α to be a smooth curve in M connecting $x_0 := \pi(\hat{x}_0)$ to x with length $|\alpha| \leq C_1$, a constant independent of x . Let $\hat{\alpha}$ be the lift of α with initial point $\hat{x}$. The other end of $\hat{\alpha}$ is $\hat{x}_0 + [\alpha \circ \gamma]$ because it is the end of $\hat{\alpha} \circ \hat{\gamma}$ which

is a lift of $\alpha \circ \gamma$, a curve starting at $\hat{x}_0$. (Here $[\alpha \circ \gamma]$ is the homology class of the loop $\alpha \circ \gamma$.)

Apply $\hat{\psi}$ to get

$$|\hat{\psi}(\hat{x}) - \hat{\psi}(\hat{x}_0 + [\alpha \circ \gamma])| \leq |\hat{\psi} \circ \hat{\alpha}| \leq C_2, \quad (5.16)$$

where C_2 is a constant independent of x since $\hat{\psi}$ is uniformly continuous. Using the definition of ψ_* and the identification of $\mathbb{R}^N \cong H^1(\mathbb{T}^N, \mathbb{R})$, we get

$$|\hat{\psi}(\hat{x}) - \hat{\psi}(\hat{x}_0) - \psi_*[\alpha \circ \gamma]| \leq C_2. \quad (5.17)$$

Thus

$$|\hat{\psi}(\hat{x}) - \psi_*[\alpha \circ \gamma]| \leq C_3 := C_2 + |\hat{\psi}(\hat{x}_0)|, \quad (5.18)$$

that is, $\hat{\psi}(\hat{M})$ is contained within the C_3 -neighborhood of V , making $\hat{\psi}(\hat{M})/V$ bounded.

Since $\hat{\psi}(\hat{M})/\text{im } \psi_*$ is invariant under the A -induced automorphism of $\mathbb{R}^N/V$, which is hyperbolic, it must equal $\{0\}$. Thus $\hat{\psi}(\hat{M}) \subseteq V$, which means that $\psi(M) \subseteq V/\Gamma$. $\square$

Hence, in considering embeddings ψ of M onto invariant subsets of a hyperbolic toral automorphism, by passing to an f_A -invariant subtorus of $\mathbb{T}^N$ if necessary, we may restrict to the case where $\psi_* : H_1(M, \mathbb{R}) \rightarrow H_1(\mathbb{T}^N, \mathbb{R}) \cong \mathbb{R}^N$ is surjective.

Theorem 5.20 (Essential Uniqueness of Embedding). *If $\psi : M \rightarrow \mathbb{T}^N$ is some embedding such that*

$$\psi \circ f = f_A \circ \psi, \quad (5.19)$$

where f is a pseudo-Anosov map with orientable foliations and $f_A : \mathbb{T}^N \rightarrow \mathbb{T}^N$ is a hyperbolic toral automorphism, and $\psi_* : H_1(M, \mathbb{R}) \rightarrow H_1(\mathbb{T}^N, \mathbb{R}) \cong \mathbb{R}^N$ is surjective, then ψ coincides with a Franks map $h : M \rightarrow \mathbb{T}^N$ constructed from a suitable set of cohomology classes.

To prove the theorem, we start with the choice of appropriate cohomology classes. For $i = 1, 2, \dots, N$, let ω_i be a smooth form on M representing the singular cohomology class $\psi^*[dx_i]$ and define $\Omega := \text{im } \psi^* = \text{span}_{\mathbb{R}}(\omega_1, \dots, \omega_N)$. To see that the $(\omega_i)_{i=1}^N$ have hyperbolic behavior, note that (5.19) yields $f^* \circ \psi^* = \psi^* \circ f_A^*$, on the level of cohomology. Thus Ω is f^* -invariant and the action of f^* on Ω is linearly conjugated to that of f_A^* . Let A_Ω be the matrix of the linear transformation $f^*|_\Omega$ with respect to the basis $([\omega_1], \dots, [\omega_N])$. The spectrum of A_Ω is a subset of that of A , making A_Ω hyperbolic. Thus we can speak of the Franks map $h : M \rightarrow \mathbb{T}^N$ as constructed in [10].

Take $\hat{\omega}_i = \pi^*\omega_i$, the lifted version of ω_i on $\hat{M}$. Now, by the key property of the homology cover $\hat{M}$, there are smooth functions $\hat{\varphi}_i : \hat{M} \rightarrow \mathbb{R}, i = 1, \dots, d$, such that $d\hat{\varphi}_i = \hat{\omega}_i$.

Lemma 5.21. $\hat{\psi}_i : \hat{M} \rightarrow \mathbb{R}$, for $i = 1, \dots, N$, obtained as the i -th components of $\hat{\psi}$, satisfy $\sup_{\hat{x} \in \hat{M}} |\hat{\varphi}_i(\hat{x}) - \hat{\psi}_i(\hat{x})| < +\infty$.

Proof. To begin, let $\xi : M \rightarrow \mathbb{T}^N$ be a smooth C^0 ε -approximation of $\psi : M \rightarrow \mathbb{T}^N$ so that ψ and ξ are homotopic and $\hat{\xi} : \hat{M} \rightarrow \mathbb{R}^N$ be a lift of ξ satisfying $|\hat{\xi}(\hat{x}) - \hat{\psi}(\hat{x})| < \varepsilon$

for all $\hat{x} \in \hat{M}$. We have:

$$\omega_i = \psi^*[dx_i] = \xi^*[dx_i] = [\xi^*(dx_i)], \quad (5.20)$$

where the middle equality comes from homotopic equivalence of ψ and ξ and the last equality uses the identification of $H^1(M, \mathbb{R})$ with deRham cohomology and the fact that pullback of forms by ξ descends to the map on the cohomology induced by ξ (which is why we are overloading the notation ξ^*). From this, $\omega_i = \xi^*(dx_i) + d\eta_i$ for some smooth $\eta_i : M \rightarrow \mathbb{R}$. Fix a base point $\hat{x}_0 \in \hat{M}$; for an arbitrary $\hat{x} \in \hat{M}$ we can integrate along a smooth path from $\hat{x}_0$ to $\hat{x}$ to get:

$$\begin{aligned} \hat{\varphi}_i(\hat{x}) - \hat{\varphi}_i(\hat{x}_0) &= \int_{\hat{x}_0}^{\hat{x}} d\hat{\varphi}_i = \int_{\hat{x}_0}^{\hat{x}} \hat{\omega}_i = \int_{\hat{x}_0}^{\hat{x}} \pi^* \omega_i \\ &= \int_{\hat{x}_0}^{\hat{x}} \pi^*(\xi^*(dx_i) + d\eta_i) \\ &= \int_{\hat{x}_0}^{\hat{x}} \hat{\xi}^* \circ \pi_N^*(dx_i) + d(\eta_i \circ \pi) \quad (\text{from } \xi \circ \pi = \pi_N \circ \hat{\xi}) \\ &= \int_{\hat{x}_0}^{\hat{x}} d(\hat{\xi}^* \hat{x}_i) + d(\eta_i \circ \pi) = \int_{\hat{x}_0}^{\hat{x}} d\hat{\xi}_i + d(\eta_i \circ \pi) \\ &= \hat{\xi}_i(\hat{x}) - \hat{\xi}_i(\hat{x}_0) + (\eta_i \circ \pi)(\hat{x}) - (\eta_i \circ \pi)(\hat{x}_0) \\ &= \hat{\xi}_i(\hat{x}) - \hat{\xi}_i(\hat{x}_0) + \eta_i(x) - \eta_i(x_0) \end{aligned}$$

Since $|\eta_i(x) - \eta_i(x_0)| \leq C_4 := 2 \sup_{x \in M} |\eta_i(x)| < +\infty$, the above yields

$$|\hat{\varphi}_i(\hat{x}) - \hat{\psi}_i(\hat{x})| \leq |\hat{\varphi}_i(\hat{x}) - \hat{\xi}_i(\hat{x})| + \varepsilon \leq |\varphi_i(\hat{x}_0) + |\xi_i(\hat{x}_0)| + C_4 + \varepsilon =: C_5 < +\infty \quad (5.21)$$

for all $\hat{x} \in \hat{M}$. □

To finish the proof of Theorem 5.20, define $\hat{\Phi} = (\hat{\varphi}_1, \dots, \hat{\varphi}_N) : \hat{M} \rightarrow \mathbb{R}^N$ and let $\hat{h} : \hat{M} \rightarrow \mathbb{R}^N$ be the lift of h constructed in [10], characterized by the global shadowing

property

$$\sup_{n \in \mathbb{Z}} \text{dist} \left(A^n \circ \hat{h}(\hat{x}), \hat{\Phi} \circ \hat{f}^n(\hat{x}) \right) < +\infty \quad (5.22)$$

for $\hat{x} \in \hat{M}$.

Thus, for any $\hat{x} \in \hat{M}$, Lemma 5.21 gives that

$$\sup_{n \in \mathbb{Z}} |\hat{\Phi} \circ \hat{f}^n(\hat{x}) - \hat{\psi} \circ \hat{f}^n(\hat{x})| < +\infty, \quad (5.23)$$

and the triangle inequality applied to (5.22) and (5.23) yields

$$\sup_{n \in \mathbb{Z}} |A^n \circ \hat{h}(\hat{x}) - \hat{\psi} \circ \hat{f}^n(\hat{x})| < +\infty. \quad (5.24)$$

Since $\hat{\psi} \circ \hat{f}^n = A^n \circ \hat{\psi}$ by Lemma 5.8 (page 121), this is the same as

$$\sup_{n \in \mathbb{Z}} |A^n \circ \hat{h}(\hat{x}) - A^n \circ \hat{\psi}(\hat{x})| < +\infty. \quad (5.25)$$

Now, since A is hyperbolic, $\hat{h}(\hat{x}) = \hat{\psi}(\hat{x})$. By arbitrariness of $\hat{x}$, $\hat{h} = \hat{\psi}$. Thus also

$h = \psi$. □

REFERENCES CITED

- [1] Curtis T. McMullen. Dynamics of $\mathrm{SL}_2(\mathbb{R})$ over moduli space in genus two. *Ann. of Math. (2)*, 165(2):397–456, 2007.
- [2] Pascal Hubert, Erwan Lanneau, and Martin Möller. $\mathrm{GL}_2^+(\mathbb{R})$ -orbit closures via topological splittings. In *Surveys in differential geometry. Vol. XIV. Geometry of Riemann surfaces and their moduli spaces*, volume 14 of *Surv. Differ. Geom.*, pages 145–169. Int. Press, Somerville, MA, 2009.
- [3] William A. Veech. Gauss measures for transformations on the space of interval exchange maps. *Ann. of Math. (2)*, 115(1):201–242, 1982.
- [4] Morris W. Hirsch. On invariant subsets of hyperbolic sets. In *Essays on Topology and Related Topics (Mémoires dédiés à Georges de Rham)*, pages 126–135. Springer, New York, 1970.
- [5] S. G. Hancock. Construction of invariant sets for Anosov diffeomorphisms. *J. London Math. Soc. (2)*, 18(2):339–348, 1978.
- [6] Feliks Przytycki. Construction of invariant sets for Anosov diffeomorphisms and hyperbolic attractors. *Studia Math.*, 68(2):199–213, 1980.
- [7] Ricardo Mañé. Invariant sets of Anosov’s diffeomorphisms. *Invent. Math.*, 46(2):147–152, 1978.
- [8] Albert Fathi. Some compact invariant sets for hyperbolic linear automorphisms of torii [tori]. *Ergodic Theory Dynam. Systems*, 8(2):191–204, 1988.
- [9] John Franks. Anosov diffeomorphisms. In *Global Analysis (Proc. Sympos. Pure Math., Vol. XIV, Berkeley, Calif., 1968)*, pages 61–93. Amer. Math. Soc., Providence, R.I., 1970.
- [10] Marcy Barge and Jaroslaw Kwapisz. Hyperbolic pseudo-Anosov maps almost everywhere embed into a toral automorphism. *Ergodic Theory Dynam. Systems*, 26(4):961–972, 2006.
- [11] M. Gromov. Super stable Kählerian horseshoe? In *Essays in mathematics and its applications*, pages 151–229. Springer, Heidelberg, 2012.

- [12] Gavin Band. Identifying points of a pseudo-Anosov homeomorphism. *Fund. Math.*, 180(2):185–198, 2003.
- [13] Koichi Hiraide. Expansive homeomorphisms of compact surfaces are pseudo-Anosov. *Osaka J. Math.*, 27(1):117–162, 1990.
- [14] Jorge Lewowicz. Expansive homeomorphisms of surfaces. *Bol. Soc. Brasil. Mat. (N.S.)*, 20(1):113–133, 1989.
- [15] Anton Zorich. Explicit Jenkins-Strebel representatives of all strata of abelian and quadratic differentials. *J. Mod. Dyn.*, 2(1):139–185, 2008.
- [16] Maxim Kontsevich and Anton Zorich. Connected components of the moduli spaces of Abelian differentials with prescribed singularities. *Invent. Math.*, 153(3):631–678, 2003.
- [17] Howard Masur. Ergodic theory of translation surfaces. In *Handbook of dynamical systems. Vol. 1B*, pages 527–547. Elsevier B. V., Amsterdam, 2006.
- [18] Steven Kerckhoff, Howard Masur, and John Smillie. Ergodicity of billiard flows and quadratic differentials. *Ann. of Math. (2)*, 124(2):293–311, 1986.
- [19] Howard Masur and Serge Tabachnikov. Rational billiards and flat structures. In *Handbook of dynamical systems, Vol. 1A*, pages 1015–1089. North-Holland, Amsterdam, 2002.
- [20] Anton Zorich. Flat surfaces. In *Frontiers in number theory, physics, and geometry. I*, pages 437–583. Springer, Berlin, 2006.
- [21] H. M. Farkas and I. Kra. *Riemann surfaces*, volume 71 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1992.
- [22] Adriano M. Garsia. An imbedding of closed Riemann surfaces in Euclidean space. *Comment. Math. Helv.*, 35:93–110, 1961.
- [23] Reto A. Rüedy. Embeddings of open Riemann surfaces. *Comment. Math. Helv.*, 46:214–225, 1971.
- [24] H. Masur. Closed trajectories for quadratic differentials with an application to billiards. *Duke Math. J.*, 53(2):307–314, 1986.

- [25] J. Smillie. The dynamics of billiard flows in rational polygons. In *Dynamical systems, ergodic theory and applications*, volume 100 of *Encyclopaedia of Mathematical Sciences*, pages 360–382. Springer, Berlin, 2000.
- [26] Yaroslav Vorobets. Periodic geodesics on generic translation surfaces. In *Algebraic and topological dynamics*, volume 385 of *Contemp. Math.*, pages 205–258. Amer. Math. Soc., Providence, RI, 2005.
- [27] Jean-Christophe Yoccoz. Continued fraction algorithms for interval exchange maps: an introduction. In *Frontiers in number theory, physics, and geometry. I*, pages 401–435. Springer, Berlin, 2006.
- [28] Marcelo Viana. Ergodic theory of interval exchange maps. *Rev. Mat. Complut.*, 19(1):7–100, 2006.
- [29] Michael Keane. Interval exchange transformations. *Math. Z.*, 141:25–31, 1975.
- [30] Anatole Katok and Boris Hasselblatt. *Introduction to the modern theory of dynamical systems*, volume 54 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1995. With a supplementary chapter by Katok and Leonardo Mendoza.
- [31] M. Mulase and M. Penkava. Ribbon graphs, quadratic differentials on Riemann surfaces, and algebraic curves defined over $\overline{\mathbf{Q}}$. *Asian J. Math.*, 2(4):875–919, 1998. Mikio Sato: a great Japanese mathematician of the twentieth century.
- [32] Anton Zorich. Square tiled surfaces and Teichmüller volumes of the moduli spaces of abelian differentials. In *Rigidity in dynamics and geometry (Cambridge, 2000)*, pages 459–471. Springer, Berlin, 2002.
- [33] John M. Franks. Invariant sets of hyperbolic toral automorphisms. *Amer. J. Math.*, 99(5):1089–1095, 1977.
- [34] Ricardo Mañé. Orbits of paths under hyperbolic toral automorphisms. *Proc. Amer. Math. Soc.*, 73(1):121–125, 1979.
- [35] Benson Farb and Dan Margalit. *A primer on mapping class groups*, volume 49 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 2012.
- [36] *Travaux de Thurston sur les surfaces*. Société Mathématique de France, Paris, 1991. Séminaire Orsay, Reprint of it *Travaux de Thurston sur les surfaces*,

Soc. Math. France, Paris, 1979 [MR0568308 (82m:57003)], Astérisque No. 66-67 (1991).

- [37] Anton Zorich. How do the leaves of a closed 1-form wind around a surface?
In *Pseudoperiodic topology*, volume 197 of *Amer. Math. Soc. Transl. Ser. 2*,
pages 135–178. Amer. Math. Soc., Providence, RI, 1999.
- [38] James R. Munkres. *Topology: a first course*. Prentice-Hall Inc., Englewood Cliffs,
N.J., 1975.
- [39] Clark Robinson. *Dynamical systems*. Studies in Advanced Mathematics. CRC
Press, Boca Raton, FL, second edition, 1999. Stability, symbolic dynamics,
and chaos.